Barriers to the Implementation of Decision Tools
(or why the best laid plans go awry)

Annie R. Pearce, Ph.D. Candidate
School of Civil & Environmental Engineering
Georgia Institute of Technology, Atlanta, GA USA

Introduction

While a substantial body of literature exists on the theory of human judgment and decision making with respect to issues such as the underlying cognitive structures which support such behavior, there is a relative dearth of information on issues surrounding the *implementation* of theory to aid humans in making judgments and decisions. Within the literature of many specific domains in engineering and applied science, on the other hand, many examples can be found of tools which have been developed to either aid humans in making decisions or replace them altogether (e.g., citations). For example, the primary focus of research in the field of artificial intelligence is to develop machines which can imitate or improve upon the (presumably) intelligent behavior of humans.

Yet very few connections can be found between the domains of theoretical and applied decision making. Even within the domain of theoretical work on human judgment and decision making, a dichotomy exists with respect to how decision making should be studied. On one side of the theoretical playing field, normative researchers in human judgment and decision making develop Olympian models of how humans *should* make decisions, and scoff at how seemingly irrational the behavior of humans really is. Meanwhile, researchers on the descriptive side of the field spend their time looking at how humans actually behave, and justify their work by telling themselves that "people live in a world more complex than can be handled by even sophisticated normative models, and most get by. This is a remarkable feat. It deserves considered study." (Lopes 1986).

A need exists for facilitating the transfer of results developed in the theoretical domain of decision making research to assist the efforts of researchers in applied domains who are developing decision aids or tools. This need is exacerbated by the many barriers which must be surmounted in the implementation of decision aids or tools developed by applied researchers.

Motivation for this study

The initial objective of this study was to conduct a meta-level analysis of the literature to assess the reasons for user acceptance or rejection of the results of artificial decision processes (ADPs), and to develop a set of guidelines for developing and implementing ADPs which could lead to increased user acceptability of their output. The study was motivated by a need on the part of the researcher for guidelines to aid in the development of a decision support system for the problem of construction materials selection. The primary question to be answered by this work was the following: what can designers of decision support tools do to ensure that those tools are useful and that they generate results which are acceptable to their users?

Due to time constraints and lack of documentation of specific cases of *applied* ADPs in the literature, the focus of the work has been shifted toward examination of a particular case study from the author's experience, which exemplifies the many difficulties encountered in implementation of decision aids in real life. Based on a review of the general literature available in the domain of human judgment and decision making, three general classes of barriers to implementation have been identified, and several strategies for overcoming these barriers are proposed.

Overview of the paper

Following this introduction, the report consists of three additional sections. In the next section, three categories of barriers to the successful implementation of decision tools are identified and discussed: weaknesses in model development, problems with user acceptability, and lack of model robustness. This section also contains a discussion of some potential strategies which may be helpful in overcoming barriers encountered in implementing decision tools. The subsequent section of the report contains a case study from the author's research experience which serves to illustrate some of the implementation barriers discussed in previous sections. The report concludes with a discussion of conclusions and a list of references.

Barriers to Implementation of Decision Aids

In this section of the paper, three categories of potential barriers to the implementation of decision aids are identified based on an overview of the literature on human decision making. Barriers relating to the development of supporting judgment and evaluation models, model robustness, and user acceptability are presented, and potential strategies for overcoming these barriers are proposed.

Model development

The potential for barriers to impede the implementation of decision tools begins at the earliest phases of model development. From the very beginning, flaws in the underlying assumptions of applying theoretical models can invalidate the resulting decision aid and render its implementation faulty. Failure to adhere to the axioms of rationality (transitivity, continuity, etc.), for example, can render any tool based on utility theory hopelessly unpredictable and inaccurate. Similar faults can occur from assuming independence between attributes when those attribute variables are in fact interdependent. Thus, careful attention must be paid to make certain that the fundamental theories and assumptions on which the tools are founded are accurately and appropriately applied (Hogarth 1986).

Additional barriers can arise without earlier warning when decision tools are completely developed without ensuring representative design in the experiments used to develop the models on which those tools are based (e.g., Dawes 1979, Hammond 1986). For example, generalization across experimental tests or cases is at least as important as generalization across subjects in order to create a model which will be generalizable to any domain of application (Hammond 1986, Hogarth 1986, Brunswik Lens Model stuff). If the model which underlies the decision aid is based on some set of cases which are unrepresentative of the world, then the predictive capability of that model is practically worthless in any situation other than the simplistic laboratory setting in which it was developed (Goldberg 1968, Roy 1988).

In situations like these, reevaluation of the underlying assumptions and experimental designs which support the model and decision tool must occur in order for implementation to be successful. If careful consideration is given to ensuring representative design and the validity of assumptions from the beginning of model development by those who develop decision tool applications, then problems with model generalizability and decision aid implementation stemming from faulty models are much less likely to occur (Hogarth 1986). In addition, researchers who develop theoretical models of decision making, both normative and descriptive, should give some consideration to the manner in which their models may be applied, and should use this consideration to shape the nature in which their results are presented in the literature and in practice.

Model robustness

The robustness of a model, or its ability to withstand minor fluctuations in input without dramatic changes in output, is an important contributor to the successful implementation of a decision tool based on that model. Lack of model robustness often stems from assumptions inherent in how preferences and indifferences are represented in models. For example, even if the difference in preference between two outcomes is very small, models often represent that difference in such a way as to imply that a well-defined preference exists for one outcome over the other (Roy 1988). This representation convention tends to make decision models, and the tools which are based on them, extremely susceptible to relatively small variations in the arbitrary way in which the model is defined. Such lack of robustness makes implementation of decision tools difficult not only because the reliability of the tools can be questioned, but also because the confidence of the user in the tool is correspondingly lower.

Roy provides several suggestions for increasing the robustness of decision aids by changing the traditional mono-criterion approach to establishing preferences for modeling decisions (ibid.). He suggests that developers of decision tools and models need to remember that the models of the alternatives for selection are not the same as the alternatives themselves, and as such are necessarily impoverished and somewhat arbitrary. Consequently, the developer should not forget that the representations chosen for a given model are not the only, or necessarily the best, representations of reality. Roy proposes that a multi-criterion approach be taken to establishing preferences and measuring the validity of assumptions used to construct decision models and aids. By verifying the reasonableness of components used to construct models, the robustness of those models should improve. As a result, the reliability of the models should increase, reducing the potential number of barriers associated with implementing the decision tool.

User acceptability

From the standpoint of users, decision support technologies and aids often suffer the same acceptability problems as most new technologies. Users who were experts using the former methods and tools are suddenly reduced to novices when the new tools are put into place. Although their domain expertise remains intact, these users often experience difficulty in understanding the new tools and learning how to use them. Especially if users are not informed about how and why the new tools were developed, they may feel resistant to learning to use the tools. In some cases, users may feel threatened by the tools due to a fear that the tools were intended to replace them or perform tasks that the user him- or herself previously performed. Users may (correctly or incorrectly) interpret the implementation of new tools to mean that management thinks they could not adequately meet their task requirements. Thus, these users will naturally experience some resentment and resistance about switching to new tools.

One solution to this problem with user acceptability is including potential users in the development process for new tools and technology. When users are allowed to contribute to the development process and their input is taken seriously in making development decisions, they will most likely feel a sense of confidence with respect to the product and develop a certain degree of vested interest in the success of the tool. On the other hand, if the tool will incorporate some judgment mechanism which does not have face validity with the judgment process that the users feel they themselves use, knowledge of that fact may engender additional resistance to the resulting tool (Hoffman et al. 1968).

In many cases approaches such as linear regression models of cue usage may yield superior accuracy in judgment tasks, compared to the more complex configural or curvilinear models which experts often claim to use in their judgment processes. However, if expert users are aware that the decision tool is based on such non-representational model, they may offer additional resistance to its implementation (e.g., Hoffman et al. 1968). In cases such as these, smooth implementation of the tool may prove to be impossible as long as users are uncomfortable in relying on its output.

Case Study: Subjective group decision making for evaluation of alternatives

As an example of some barriers which can hamper the implementation of decision tools, a case study of the attempted implementation of a tool for subjective group decision making is presented below. First, the implementation is discussed as it was expected to occur, followed by a description of what actually happened when implementation was attempted. Then, some explanations for the failure of implementation are presented, and some lessons that were learned as a result of the failure are discussed† .

The process as it was expected to occur

An ongoing research project with which the author is associated is dedicated to identifying unique or innovative project management practices used in commercial facility construction, which promote project and operational cost effectiveness. The approach taken in this research involves interviews with companies which are currently undertaking new facility construction, selected on the basis of several performance criteria. The objective of the research is to compile an analysis of the successful and cost effective management practices used by these companies, based on the interviews.

The company selection process involved two separate groups of decision-makers: a task force of professional managers from industry (TF) who are members of the organization which funded the research, and an academic research team (ART) who is actually carrying out the contracted research, consisting of several professors and graduate students. The first phase of the process, initial selection of a pool of potential companies to be contacted for interviews, occurred in four steps:

1) Based on the objectives of the research project agreed upon by ART and the TF, a brainstorming and evaluation session was held by the TF to establish a set of selection criteria and mutually agreed upon weightings for company selection.

† Please note that this case study has been included without permission from the principal investigators of the research project, and without any formal documentation of the attempted implementation having been published. The author has made every attempt to protect the dignity and maintain the confidentiality of those involved, but entreats the reader to abstain from discussing the case study with anyone. The case study has been presented from the perspective of the author who was a research assistant assisting in implementation, and the author takes no credit or blame for design of the decision tool or implementation plan. There, have I covered my butt sufficiently?

2) Using industry-accepted financial indicators and capital project investment data which reflected the selection criteria and weightings established by the TF, members of the ART identified a pool of "successful" companies in each of fourteen industries.

3) A semi-subjective aggregation procedure was used by the ART to select the top five successful companies within each of the fourteen industries, based on the financial indicators and capital project data. A compensatory selection strategy was used.

4) The list of selected companies from the fourteen industries was given to members of the TF, who were supposed to use a subjective group decision making aggregation tool to collectively rank the companies in each industry and select the company in each category with the highest ranking.

The fourth step, implementation of the subjective group decision making aggregation tool, was where the whole process crashed and burned. The tool itself was a rating matrix (see Attachment 1) which was distributed to each TF member and would have been compiled using a linear additive compensatory technique based on a numerical valuation of the possible answers for each criterion, the TF's predesignated weightings of the importance of each criterion, and the user's subjective weighting of his or her confidence in his or her answer. The criteria along the left side of the matrix were exactly the same criteria selected by the TF in the first step of the process.

Respondents were asked to complete a rating for each of the five companies in all fourteen industries across all of the evaluation criteria, based on any subjective or objective information available to them. Confidence ratings for each answer could range from 1 to 10, so that the respondent would feel free to include purely speculative answers and assign low confidence weightings accordingly. Evaluation matrices were sent to all members of the TF (N=19).

Each respondent was a voluntary participant in the task force, so it was assumed that the cooperation of the respondents would extend to participation in the group decision making process. In addition, all members of the TF were briefed in advance about the process, and agreed collectively to complete the matrix. Instructions were provided on how to complete the forms (daunting as they might be), and a contact person from the ART was designated to field any questions which might arise during the response period. A deadline for replies was set, and provisions for numerically aggregating the replies and generating ratings for each company were made.

It was expected that not all respondents would provide answers for every company and for every criterion; however, we *did* expect that each TF member would respond in some fashion. The aggregation procedure was designed to eliminate biases resulting from variations in the frequency of response. The confidence weightings for each answer on each response were designed to accommodate subjectivity on the part of the respondents; however, subjectivity was encouraged in order to capture qualitative or undocumentable information which industry respondents might have about the companies under evaluation.

What really happened

Only two out of nineteen subjects returned responses to the ART. Neither of those two respondents provided answers for more than five total companies, and neither provided information for all criteria for any of the companies. As it turned out, the PI requested that the three research assistants who collected the financial data for step 2 complete the matrix based on their subjective ideas about the companies for which they were able to obtain indicator information. Ranking of companies as reported by the ART to the TF was based solely on these three relatively complete responses. The TF subsequently decided to abandon the entire selection strategy and its results (much to the disgust and dismay of the ART), and adopted a new company selection strategy based solely on a subjective "whoever yells loudest and is most convincing" process at the next TF meeting.

What we did wrong

Although the ART had no explicit reason to expect that the TF members could supply information for each of the criteria they collectively specified in step 1, we assumed that they could supply at least a decent number of subjective data points for each company and each criteria, since the TF picked out the criteria themselves and designated them to be important for the selection process. We were wrong. Given that the TF members were being cooperative as they promised, and assuming that they agreed with the indicators selected by the ART for preliminary selection of a data pool (they had been briefed about our approach, and no one dissented), the most promising explanation for the disappointing response lies with the numerical selection process we chose to use.

Since the TF approved of the approach we took at various stages along the way and actively contributed to selecting evaluation criteria, we have no reason to believe that they found the approach to be unacceptable due to the user acceptability barrier described earlier. Although the selection process eventually adopted by the TF was different in execution than the numerical process we attempted to use, the outcomes of the two approaches could have been reasonably similar, since the confidence weightings self-assigned to each subject's "votes" in our response matrix should roughly correspond to how strongly the subject would argue in favor of each company in the "whoever yells loudest" approach.

The model development barrier of representativeness discussed previously should not have been a problem in this case, since the set of alternatives came straight from the real world and included all the American companies for which we could find data from any source. In addition, the selection criteria were not established in some sterile laboratory setting by researchers intent on simplification. Instead, criteria were chosen by industrial practitioners in a brainstorming session, with every opportunity taken to consider all factors which might be relevant.

The biggest problem in this case turned out to be that of data availability. This huge constraint prevented the effective use of a linear additive compensatory model, where most or all of the ramifications of all relevant alternatives need to be known in order to make comparisons. Instead, a non-compensatory strategy had to be adopted (a la Einhorn et al. 1979), most likely a conjunctive one (or possibly disjunctive - I wasn't at the TF meeting, and we didn't bother to go back and reexamine the selected companies in terms of our indicators). A simple feedback exercise at the TF meeting where brainstorming of the evaluation criteria occurred might have made the data availability problem obvious, although it's hard to say in retrospect.

What we learned

This researcher learned to ALWAYS do prototype studies on a small scale before expending a tremendous amount of effort attempting to design and implement a process which may not work at all. Even when the TF blessed the approach at every step and supplied the evaluation criteria themselves, there was no guarantee that the process could be implemented successfully. And even though the TF probably intellectually wanted the selection process to be objective and defensible (as suggested in Tversky & Kahneman 1981), the result of this experience shows that sometimes users decide to abandon the rigor of formal decision tools in favor of purely subjective, undocumented decision making, despite the most carefully laid plans and convincing arguments to the contrary.

In future research, the wisdom of using non-compensatory or satisficing strategies will be one the first things considered by this author. The vast reduction in information requirements and increase in processing efficiency, along with the high probability that no one will appreciate all the effort put into a compensatory process anyway, makes using a non-compensatory strategy appealing under all but the most rigorous or delicate conditions. In cases such as this one, the added robustness of the process resulting from using a compensatory strategy adds little value to the overall accuracy of the inherently subjective evaluation process anyway.

Conclusions

Three general classes of barriers to implementing decision aids or tools have been identified: weaknesses in model development, lack of model robustness, and problems with user acceptability. Several strategies have been suggested for surmounting these barriers, including careful questioning of underlying assumptions, adopting a multi-criterion approach to establishing thresholds or values for variables in decision models, and including potential users in the development process for decision tools. The case study presented here is an illustration of how difficult implementation of decision tools in the real world can be. In designing and implementation of decision tools, thoughtful consideration should be given not only the data requirements of particular decision strategies, but also the degree of value added by using a more rigorous compensatory decision strategy as opposed to a non-compensatory one.

References

Chapanis, A. (1967). "The Relevance of Laboratory Studies to Practical Situations," *Ergonomics*, v. 10, n. 5, pp. 557-577.

Dawes, R.M. (1979). "The robust beauty of improper linear models in decision making," *American Psychologist*, v. 34, pp.

Edwards, W. (19). "Use of Multiattribute Utility Measurement for Social Decision Making," in *Conflicting Objectives in Decisions*, D.E. Bell, R.L. Keeney, & H. Raiffa, eds. John Wiley & Sons, Chichester, pp. 247-276.

Edwards, W. (1973). "Divide and Conquer: How to Use Likelihood and Value Judgments in Decision Making," in *Systems Concepts: Lectures on Contemporary Approaches to Systems*, R.F. Miles, Jr., ed. pp. 87-110.

Einhorn, H.J., Kleinmuntz, D.N., and Kleinmuntz, B. (1979). "Linear regression and process-tracing models of judgement," *Psychological Review*, v. 86, n. 5.

Goldberg, L.R. (1968). "Simple models or simple processes? Some research on clinical judgments," *American Psychologist*, v. 23, pp. .

Hammond, K.R. (1986). "Generalization in operational contexts: What does it mean? Can it be done?" *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-16, n. 3.

Hoffman, P.J., Slovic, P., and Rorer, L.G. (1968). "An analysis-of-variance model for the assessment of configural cue usage in clinical judgment," *Psychological Bulletin*, v. 69, n. 5, pp. .

Hogarth, R.M. (1980). *Judgment and Choice: The Psychology of Decision*. John Wiley & Sons, Chichester.

Hogarth, R.M. (1986). "Generalization in decision research: the role of formal models," *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-16, n. 3.

Klein, G.A. (Preprint). "Recognition-primed decisions," in *Advances in Man-Machine Systems Research*, W.B. Rouse, ed. v. 5, JAI Press, Greenwich, CT.

Larichev, O.I. (19). "A Practical Methodology of Solving Multicriterion Problems with Subjective Criteria," in *Conflicting Objectives in Decisions*, D.E. Bell, R.L. Keeney, & H. Raiffa, eds. John Wiley & Sons, Chichester, pp. 197-208.

Lopes, L.L. (1986). "Aesthetics and the decision sciences," *IEEE Transactions on Systems, Man, and Cybernetics*, v. SMC-16, n. 3.

Payne, J.W. (1982). "Contingent decision behavior," *Psychological Bulletin*, v. 92, n. 2, pp. .

Roy, B. (1988). "Main sources of inaccurate determination, uncertainty and imprecision in decision models," in *Compromise, Negotiation and Group Decision*, B.R. Munier and M.F. Shakun, eds. pp. 43-62.

Tversky, A., and Kahneman, D. (1981). "The framing of decisions and the psychology of choice," *Science*, v. 211, pp. .