Evaluation of Group Decision Making Methods

Dissertation excerpt from Kirti Peniwati Srisoepardani, Ph.D.

Chapter 6

Evaluation of Group Decision Making Methods

6.1 Introduction

In this chapter we are concerned with the evaluation of different methods of group decision-making on a wide set of criteria which range from the strictly technical, to the psychophysical and social, and finally, to the logical and scientifically valid. Our purpose is to identify similarities and differences with the aim of showing from such wide consideration which method is more reliable, and is likely to gain greater attention both in academia and in practice. This writer is not an inexperienced (and completely unbiased) observer as I have familiarized myself with, have applied several of these methods, and have developed opinions and preference about them. It will be seen below that the AHP is the most comprehensive of these methods in structure, in analysis, in mathematical validity, and in its producing truthful results. Its outcome would survive outside influences because it is able to incorporate such influences in its hierarchies, and assess their relative impact on the outcome in a way that does not tax one's intelligence to accept its procedures, nor do these procedures alienate the user. The Analytic Network Process (ANP) is a natural generalization and extension of the AHP that allows feedback and dependence among decision elements and clusters of elements [Saaty, 1996]. The fact that numerous examples of applications of the ANP exist shows that the AHP can be generalized, and is thus a validation of the AHP itself, as generalizability is a necessary condition for a good decision theory.

There are criteria proposed in the literature to judge group decision making. We use these criteria as a basis to establish new criteria for judging group decision support methods. We identify and briefly describe most of the known methods in decision-making. We define various intensity measures on each of the new criteria. We evaluate each of these methods by assigning it the intensity that describes it best on each criterion, explaining briefly why that intensity is appropriate for the method. In this manner, the methods are compared and contrasted with respect to each criterion, and an AHP model is developed to obtain an overall rating of the methods.

6.2 Criteria for Group Decision Making Methods

Rubin [Swap and associates, 1984] proposed six quality indicators for group decision making that address both achievement and maintenance goals [Brightman; 1980, 1988): *efficiency, careful development and analysis of alternatives, fairness, member satisfaction and morale, leadership effectiveness,* and *growth over time*. These indicators are developed from a group process point of view, and need to be translated into another set of indicators before they can be used as criteria for evaluating the methodologies that facilitate for a group to excell on those quality indicators. We exclude efficiency from our analysis because it is highly dependent on the way the group is organized and lead. However, we will later

outline how the AHP facilitates efficient group decision making without making an attempt to compare or contrast it with the other methods. We perceive *growth over time* as *learning*. We assume that a method which addresses group maintenance (leadership effectiveness and learning) will also ensure member satisfaction and morale, hence we do not consider the latter as a criterion explicitly.

First of all, a general method for group decision making must provide a facilitator with the means to lead the group toward its achievement and maintenance goals. The method must also assist the facilitator in enhancing individual and group learning, both single loop or small "l" learning and double loop or big "L" learning [Argyris, 1977, 1994; Pascale, 1991]. It addresses the first if it enables the group to solve problems of implementing organizational policies and achieving the goals of the organization through incrementally, based on past performance and knowledge. It addresses the second if it facilitates questioning the underlying assumptions of those policies and goals through breakthrough shift of knowledge. Careful development of alternatives means that the group must not view a problem from a scope too narrow to ensure a meaningful solution or too broad to ensure controllable actions. It also means that the group must be able to identify a set of distinct alternatives from a level of abstraction that is adequate for the group. For example, a group of top executives would view a problem from a higher level of abstraction than would a group of operational managers because they have a much wider choice space from which to draw controllable alternative courses of actions. Careful analysis of alternatives requires the group to work with a model/structure [Reagan-Cirincione, 1994] with the appropriate breadth (for relevance) and depth (for precision). A successful analysis depends on faithfulness of judgment elicitation, psychophysical applicability, and the depth of the analysis. One must first accept the premise that eliciting judgment by comparing two objects with respect to a certain property would produce the most faithful representation of one's tacit preference relations. Faithful judgment can be obtained if: (1) it is expressed directly by the decision maker, rather than derived from some other form of judgments, (2) it is not clear to the decision maker as to how that particular judgment would ultimately affect the outcome and hence would not play games with it to influence the outcome, thus preventing *strategic judgment* [Dummett, 1984], and (3) the decision maker has the choice to express preference relations numerically (as a minimum for representing objective measurement) or verbally (for representing perception or feeling), or even graphically. Depth of analysis means how well an analytical method provides the means to guide decision makers' thinking to ensure the validity of the outcome. It includes, for example, having a feedback mechanism for making changes and adjustments or directing the decision-maker to an expert or looking for specific information.

Fairness is addressed both during group interaction, and when the diverse information or judgments from individuals must be mathematically aggregated into one judgment for the group. On this criterion, we are only concerned with the method of aggregation, since group discussion would be controlled by the facilitator. A strong condition for a successful decision theory is that it needs to result in the alternatives being cardinally separated rather than simply ordered. The group members themselves may need to be prioritized as to the reliability of their opinions. Other actors or stakeholders who may be affected by the implementation of the decision often need to be considered, and a successful method needs to have a way to include their judgments.

Most significantly, a method must be applicable, valid, and reflect the truth. One would be concerned with such issues as: (1) is the method applicable to conflict resolution, (2) does it apply to intangibles in the same way it does to tangibles, (3) does it have mathematical validity and generality, and is it supported with axioms and theorems, (4) can the method

be applied to psychophysical measurement, and (5) is the outcome valid, ensuring, for example, reliability in prediction.

Applicability to conflict resolution means the method must provide a way for each conflicting party to evaluate the costs and the benefits of giving up some of what it has, in return for getting what it wants from the other party. Applicability to intangibles involves inclusion, and measurement of, the multidimensionality of the factors involved. Mathematical validity and generality calls for formal mathematical representation of the logic and reasoning behind a theory and the economy of additional assumptions required for its generalization. Psychophysical applicability means that an analytical method must deal with the measurement of relationships between the physical attributes of stimuli and the resulting sensation reflecting diminishing response to increasing stimulus such as, for example, that described by the Weber-Fechner law. Validity of the outcome involves the accuracy of the outcome in predicting situations. One should be careful, however, to define what constitutes a prediction situation. In an experimental study, Schoemaker and Waid [1982] showed that guesswork direct estimation of the rank of multicriteria objects produces a very different ordinal ranking from that of the AHP.

The following criteria are used to compare and contrast the various methods:

1. Group maintenance: leadership effectiveness.

We assume that all group methods enhance leadership effectiveness. We use a democratic leader's characteristics as criteria for leadership effectiveness, assuming that the group mostly works in moderate situational control in terms of leader-member relations, task clarification, and position power [Lewin, Lippit, and White, 1939; Fiedler, 1973]. A method is rated *low* if it is highly technical or does not involve much interaction and where leadership is of a little concern, *medium* if it provides no more than structure to facilitate group leadership, and *high* if it also provides other collaborative tools and the necessary control mechanism to guide the facilitator's leadership actions in pursuing the group's achievement and maintenance goals.

2. Group maintenance: learning

It is assumed that objective knowledge, widely accepted and agreed upon knowledge, is considered less important by the people involved in the group than what they know from their experience relevant to the issues and what they learn by problem solving within the group. A method is rated *low* if it advances technical learning that has little to do with the group member's subjective values, *medium* if it improves understanding with regard to cause-effect relations in a problem (but actions may not be clear, single loop or small 'l' learning only, or, it does not provide clear evaluation of alternatives), *high* if it facilitates both single and double loop learning, or small "l" and big "L" learning (leading to action), and *very high* if it also enables one to produce the necessary material to facilitate learning beyond the membership of the group.

3. Problem abstraction: scope

The need for problem abstraction or definition is inherent in any decision making, therefore this indicator is assumed to be addressed by all methods. The question is whether a method explicitly addresses this issue or not. Voting is an exception for which alternatives are always given, hence problem abstraction is not applicable and this method is rated *NA*. A method is rated *low* if it does not propose a specific technique and does not involve problem analysis that enhances the scope of abstraction, *medium* if its technique creates boundaries that limit group thinking, or, if it does not propose a specific technique but involves problem analysis that serves as feedback to broaden problem abstraction, and *high* if double loop learning is explicitly addressed.

4. Problem abstraction: development of alternatives.

It is generally assumed that the alternatives are not given to the group, hence any method involving problem structuring must go through a process of identifying alternatives. It is assumed that multicriteria methods require a process of generating alternatives that allows a certain degree of interaction among group members. It is also assumed that a method for enhancing problem abstraction leads to a set of alternatives. Again, voting is an exception because a set of alternatives is always given. A method is rated *NA* if the alternatives must be given, *low* if it does not provide a specific technique for identifying alternatives, *medium* if it ensures a free wheeling environment without group interaction, or, if it generates incremental alternatives (it is assumed that innovative change is more preferred to incremental change), *high* if it ensures a free wheeling environment as well as group interaction but no requirement that the alternatives selected satisfy certain properties or requirements (e.g., distinct or independent), *very high* if it is also based on challenged assumptions, if it systematically generates alternatives, or, if it requires the alternatives to satisfy certain properties to ensure the validity of the outcome.

5. Structure: breadth

A structure is said to be broad if it has many distinct elements (criteria) that are assumed to be independent of each other. A problem that is modelled by more than one such structure is considered to be even broader. A method is ranked *NA* (not applicable) if it does not involve problem structuring, *low* if the method allows only one element (direct comparison), *medium* if the method creates a constraint with respect to the number of elements, and *high* if there is no such constraint.

6. Structure: depth

A structure is said to be deep if each element is broken down into sub-elements, each sub-element into sub-sub-elements and so on down to the most detailed elements. A method is ranked *NA* if it does not involve structuring, *low* if it allows only one element, *medium* if it creates a constraint with respect to the number of elements, and *high* if there is no such

constraint.

7. Analysis: faithfulness of judgments

A method is rated *NA* if it does not involve problem analysis, *low* if it does not include intensity of preferences, *medium* if it involves direct assignment of numbers to represent intensity on an unvalidated scale, *high* if it is derived from some other judgments carefully elicited, *very high* if it is elicited in the most elementary way (pairwise comparison with respect to a property), expressed in a way that fits the decision maker best (numerically, verbally, or graphically), or, if it is by design an objective method, or, if it is continuously improved.

8. Analysis: breadth and depth of analysis (what if)

A method is rated *NA* if it does not involve problem analysis, *low* if it allows judgment, but not analysis, *medium* if the depth of analysis is constrained by the method's structure, *high* if it provides the means for careful thinking (but it is difficult to review previous analysis), and *very high* if it facilitates careful thinking and review.

9. Fairness: cardinal separation of alternatives

This indicator is applicable only to methods that involve aggregation of judgments of individual members. Alternatives can only be treated either fairly (*high*) or not fairly (*low*). A method is evaluated according to its consistency with the impossibility problem intrinsic in ordinal group aggregation. An aggregation method is rated *low* if it uses an ordinal scale of measurement and *high* if it uses an interval or a ratio scale. A method is rated *NA* if it does not involve judgment aggregation.

10. Fairness: prioritizing of group members

This indicator is also applicable only to methods that involve aggregation of individual judgments. Voting theories usually operationalize fairness as equal treatment of the voters. In group decision making, there may be circumstances in which the group may want to apply the concept of fairness with unequal treatment of the individuals involved. For example, weights may need to be assigned to the members according to the relevance of their expertise or to their known previous contribution to the goal. A method is rated *NA* if it does not involve judgment aggregation, *low* if individual preferences are represented on an ordinal scale, *medium* if the preferences are represented on an interval or ratio scale (but the individuals must carry the same weight), *high* if it also provides a group with an option to treat group members unequally (but the weights are assigned arbitrarily), and *very high*

if it provides a method to determine the weights as appropriately as the group wishes.

11. Fairness: consideration of other actors and stakeholders

This criterion is applicable only to methods that involve problem analysis. A method is rated *NA* if it does not involve problem analysis, *low* if addressing fairness to other actors which might be possible (but it is not yet made explicit in the method), *medium* if it addresses the issue explicitly but qualitatively, and *high* if it addresses the issue both explicitly and quantitatively.

12. Scientific and mathematical generality

A method is rated *NA* if it does not involve problem analysis, *low* if it does not involve any mathematics, *medium* if it involves mathematics that is not axiomatized, or, it involves multidimensional concepts that may be axiomatized differently by different researchers leading to a diversity of theorems, *high* if it is axiomatized with more or less unified conceptualization (but its generalization has considerable mathematical rigor), and *very high* if its theorems are axiomatized and generalizable in a natural and less taxing way by not requiring many new assumptions.

13. Applicability to tangibles

A method is rated *NA* if it does not involve problem analysis, *low* if it does not involve quantification of intangibles, or, simply assigns arbitrary ordinal numbers to intangibles, *medium* if it involves measuring intangibles on an interval or a ratio scale (but must be represented by tangibles or intensities in absolute terms with no assigned priority, *high* if it involves measuring intangibles on an interval or a ratio scale, but must be represented by tangibles or intensities in absolute terms with assigned priority), and *very high* if its measurement is applicable to intangibles and gives an assessment of their relative importance, both absolutely or relatively, as the user wishes.

14. Psychophysical applicability

A method is rated *NA* if it does not involve problem analysis, *low* if it does not address the Weber-Fechner law, *medium* if it could but requires a complex model that may not be practical to develop or to apply, and *high* if it is psychophysically applicable.

15. Applicability to conflict resolution

A method must have a theory and perhaps also normative standards for best solution of a

group conflict that is understandable, acceptable, practical, flexible, and has been demonstrated to work well in practice. Such a method would be rated *high*. However, secrecy makes it hard to use such an approach in a clear step by step fashion, and hence people often resort to less structured and less explicit methods. For this reason, an analytical method for dealing with conflict resolution is rated *medium*.

16. Validity of the outcome (prediction)

A method is rated *NA* if it does not involve problem analysis, *low* if it uses ordinal measurement with no structural representation of a problem, *medium* if it uses cardinal measurement, but its main concern is computation, or, if it uses ordinal measurement with some problem representation, or, if it provides a rigorous model without measurement, *high* if it uses cardinal measurement, but mathematical validity sets limits on the structural representation of a problem, and *very high* if it uses cardinal measurement and no theoretical limit with respect to the structural representation of the problem.

6.3 Group Decision Making Methods

The methods, other than the AHP described in earlier chapters, are outlined to highlight the points of interest for our evaluation. Couger [1995] provides a summary of most of the methods.

6.3.1 Structuring

Analogy and *attribute association* are methods for gaining fresh prespectives on a problem to create an alternative space from which meaningful and controllable distinct alternatives are likely to be identified. They involve the use of key words from the original formulation of a problem as the means to identify relations between the otherwise unrelated analogy/association and the original problem.

Boundary examination is a conscious effort to openly challenge and restructure the underlying assumptions that prevent one from seeing a problem from a broad perspective. The *progressive abstraction method* increases problem abstraction implied in the goal step by step. This, along with the first, differ in technique but their purpose is so similar that we do not consider them as different methods.

Brainstorming [Osborne, 1957] is based on the premise that deferred judgments enhance creativity and that oral communication diminishes it. Its modification includes, e.g., brainwriting (generating ideas in writing), bug list and negative brainstorming (generating complaints to identify weaknesses), the Crawford blue slip method (independently brainstorms in response to a number of questions that are related to a problem), and discussion among group participants as long as it is not judgmental.

Morphological connection is an attempt to broaden the space of alternatives not through problem abstraction, but from different combinations of problem attributes as in a

hierarchy. Despite what the term may imply, this method is not designed for connecting or structuring different ideas related to a problem to make a decision.

Why-What's Stopping is proposed for formulating ill-structured problems [Basadur, Ellspermann, and Evans; 1994]. It consists of a series of diverging and converging ideas by seeking responses to the questions: "How might we..." (to elicit ideas on alternative solutions) "What's stopping us ..." (to provide narrower subproblems for each response to the "How might we..." query), and "Why would we need to ..." (to ensure that we work on the right problem as stated in the "How might we..." query). The outcome of this process is a big picture of a problem, indicating relationships among problems and sub-problems, to help decision makers select the most meaningful problem area to work on.

6.3.2 Ordering and Ranking

Voting, as has been discussed at length in the previous chapters, elicits ordinal judgments and mathematically aggregates them into a group judgment. It is considered as a single criterion analysis since the individuals compare alternatives directly. For our purpose, interaction among members is considered irrelevant.

The Nominal Group Technique (NGT) [Delbecq, Van de Ven, and Gustafson; 1975] takes advantage of the positive aspects of brainstorming and brainwriting and structured communication that improves alignment of group members' perception of the problem without working towards concensus.

The Delphi method [Turoff, 1970; Linstone and Turoff, 1975; Gustafson, Shukla, Delbecq, and Walster, 1973] is similar to NGT except that the group members do not meet face to face. A great deal of preparation is required due to the nature of written communication.

Disjointed incrementalism is a method to select the best policy based on its incremental consequences. This method was proposed to deal with complex policy decisions, typically in the government, in which a holistic approach for policy decisions is either impossible or impractical. It has been argued that muddling through is a science.

Matrix evaluations refer to methods for presenting information to facilitate the evaluation of alternatives. It may describe factors and subfactors involved in a problem with their ranking scores, or by providing the relative overall positions of alternatives in a multidimensional space. For example, various company products may be evaluated with respect to their market share and growth (BCG matrix) or various organizational improvement with respect to their importance and imminence [Camillus and Datta, 1991]. These methods, however, do not provide a methodological way to arrive at a decision.

Goal programming is an approach to optimize a set of objective functions subject to constraints. However, it does not necessarily suggest decisions that optimize the objective functions [Ching and Ming; 1987]. It only yields decisions that "satisfice" [Simon, 1957]. The outcome is perceived as indicating trade offs that need to be made in terms of reducing a certain objective in return for an increase in some other objectives.

Conjoint measurement is concerned with predicting the values of a dependent variable by combining a set of independent variables in some functional form. The coefficients of the

function are usually estimated by regression techniques. A conjoint analysis measure has been suggested for use as a numerical basis for estimating the priorities of a goal programming problem [O'Leary and O'Leary, 1985].

The concept of *outranking* was developed by B. Roy based on Multiattribute Utility Theory (MAUT) principles with the motivation to improve efficiency without affecting the outcome while considering less information. The idea is to see whether there are enough arguments to decide that an alternative A_i is at least as good as A_j, while there is no *essential reason* to refute that statement. Researchers in this area have worked toward the satisfactory axiomatization of the concept, in which criterion prioritization has been their major preoccupation [Roy and Bouyssau, 1985; Vincke, 1989]. In the meantime, ten different methods have been developed to apply the concept. They differ in how the reason is formalized that leads to refuting the statement that A_i is at least as good as A_j, the type of decision problem (choice, scoring, or ranking) they address, the preference model they adopt (whether or not Weber-Fechner's psychophysical law is to be embraced), whether or not the concept of probability is used, and the way criteria weights are determined. A main weakness of the method is the ordinal way used to combine concordance and discordance that leaves one in doubt about the accuracy of its outcome.

6.3.3 Structuring and Measuring

Bayesian analysis is a popular statistical decision making process which provides a paradigm for updating information in the form of probabilities. It is based on the premise that decisions involving uncertainty can only be made with the aid of information about the uncertain environment in which the decision is made. Bayesian theory updates information by using Bayes theorem, a statement in conditional probabilities relating causes (states of nature) to outcomes. Outcomes are results of experiments used to uncover the causes. Bayesian theory revises initial or prior probabilities of causes, known from a large sample of a population, into posterior probabilities by using the outcome of an experiment or test with a certain probability of success. Prior probabilities are obtained either subjectively or empirically by sampling the frequency of occurence of a cause in a population. Posterior probabilities are those based on the prior probabilities and on both the outcome of the experiment and on the observed reliability of that experiment. Bayesian analysis often makes heavy use of probability trees and that is why we have included in this section.

Multiattribute Utility (Value) Theory (MAUT/MAVT) [Luce and Raiffa, 1957] attempts to maximize a decision maker's utility (under uncertainty) or value (preference) which is represented by a function that maps an object measured on an absolute scale into the decision maker's utility or value relations. The function is constructed by, for example in the case of MAUT, asking lottery questions involving probability to articulate decision makers' value trade-offs among the conflicting attributes. Preferences are used in MAVT. The functional representation of a multicriteria problem is obtained by aggregating the different single attribute functions, each representing a different attribute, by taking into consideration the relative weights of the attributes. The use of objective measurement leads to a complex functional representation if the Weber-Fechner law is to be embraced. The law suggests that the relation between a stimulus and an individual's response is not as smooth as may be indicated by a continuous utility function. Maintaining that "it is now firmly established that expected utility (EU) theory and subjective expected utility (SEU) theory are descriptively invalid," Miyamoto [1992] proposes a *generic utility theory*, designed as a general framework for descriptive multiattribute utility modelling. A group utility or value

function that takes the diversified evaluations of its individual members into consideration, can be obtained either by aggregating individual functions or by partial identification of the group function [Seo, 1985]. *Game theory*, which is based on utility theory, has been used to study conflict resolution. Recent versions of *MAUT/MAVT* have tended to look at the broad complexity of a problem within a structured framework and not simply as criteria and alternatives.

6.4 Evaluation of the Methods on the Criteria

1. Group maintenance: leadership effectiveness

Analogy/association, *brainstorming*, *morphological connection*, *voting*, *goal programming*, and *conjoint analysis* are rated *low* because the methods are highly technical. *Boundary examination, why-what's stopping, NGT, Delphi, disjointed incrementalism, matrix evaluation, outranking, Bayesian analysis,* and *MAUT/MAVT* are rated *medium* because they provide nothing more than simple structures to assist a facilitator. *AHP* is rated *high* because it provides collaborative tools to enhance communication effectiveness, inconsistency and incompatibility measures that provide feedback to the group members to ensure validity of the outcome, structure to facilitate task division, and the means to balance consensus and voting to obtain group judgments.

2. Group maintenance: learning

Brainstorming, voting, goal programming, and *conjoint analysis* are rated *low* because they involve highly technical knowledge. *Brainstorming* excludes member interaction because of its requirement that there be no discussion or criticism of ideas proposed. *Analogy/association, boundary examination, morphological connection, why-what's stopping, NGT, Delphi,* and *matrix evaluation* are rated *medium* because they improve understanding of the problem, but actions to take from them may not be readily clear. *Disjointed incrementalism, outranking, Bayes analysis, MAUT/MAVT* are rated *high* because it is assumed that their outcomes provide learning that leads to action. Research indicates, however, that despite group satisfaction, study participants rated the combination of *NGT* and *MAUT* as *low* in improving knowledge about the content of the issue [Thomas, McDaniel, and Dooris; 1989]. *AHP* is rated *very high* because it provides a highly summarized description of the problem that facilitates learning beyond membership of the group. Participants in an experimental study ranked the AHP as the least difficult and the most trustworthy method among those studied [Schoemaker and Waid, 1982]. It is assumed that the easier to apply and the more trustworthy a method is, the more one learns from its application.

3. Problem Abstraction: Scope

Voting is rated *NA* because a group does not generally generate alternatives, and thus

broaden the scope, but is somehow given a set of alternatives. *Brainstorming* does not involve a specific technique to enhance problem abstraction and does not involve problem analysis, and so it is rated *low*. The use of key words from the original formulation of a problem in *analogy* and *attribute association*, which ensures some relations between the analogy or association problem with the original problem, at the same time sets perceptual boundaries. For example, an analogy to a difficulty is usually another difficulty (as opposed to an opportunity) and a spatial problem is likely to generate attributes that direct thinking to increasing the productivity of the space given the same demand, rather than reducing the demand itself. For this reason, these methods are ranked *medium*. *Nominal Group Technique* and *Delphi* are also rated *medium* because they include careful preparation of a questionnaire for the group to respond to which implies the development of problem abstraction. *Disjointed incrementalism, matrix evaluation, goal programming, conjoint analysis, outranking, Bayes, MAUT/MAVT, and AHP* do not involve a technique to broaden problem abstraction, but since analysis enhances problem abstraction, they are rated *medium.* Also *outranking, Bayes, MAUT/MAVT, and AHP* are rated *medium* because they are assumed to apply techniques such as *NGT* or *Delphi* which are rated *medium*. *Morphological analysis* is rated *high* because of its systematic search for combinations of attributes that produce candidates for alternatives. *Why-what's stopping* is also rated *high* because its *why's* questions uncover the assumptions underlying the difficulties in implementing the suggested solutions identified by the *what's (how's)* questions. Structuring the responses to the repeated questions provides highly comprehensive relationships among problems, subproblems, and alternative courses of action. *Boundary examination* systematically challenges the underlying assumptions regarding the problem, hence it is also rated *high*.

4. <u>Problem abstraction: development of alternatives</u>

Analogy and *attribute association*, *boundary examination*, *matrix evaluation, goal programming, conjoint analysis,* and *Bayes analysis* are rated *low* because identifying alternatives is not an explicit part of the method. *Brainstorming/brainwriting* is rated *medium* because it ensures a free-wheeling environment but does nothing to take advantage of the positive aspects of interaction among group members. This method assumes that an alternative ranked high by the group is the most relevant solution to the problem, which may not be generally true because the group does not get to bring out possible disadvantages to the suggested alternatives. This weakness is inherent in brainstorming as well as in its modifications, e.g., brainwriting (generating ideas in writing), bug list and negative brainstorming (generating complaints to identify weaknesses). Crawford's blue slip method (independently brainstorming in response to a number of questions that are related to a problem) does not tell one how to organize the information. Brainstorming addresses the negative aspect of communication by removing interaction from the decision process, at the cost of taking advantage of its positive aspects. This may be the reason why this popular method is observed as the least effective technique [Couger, 1995]. *Disjointed incrementalism* is also rated *medium*, but because it generates incremental alternatives rather than distinct ones. *Nominal Group Technique (NGT)* and *Delphi* are rated *high* because a certain degree of alignment of group member's perceptions takes place. *Outranking* and *MAUT/MAVT* are also rated *high*, the same as NGT and Delphi, because the complexity of the problem approached using these methods is assumed to require an application of either one of the two alternative generating methods. *Morphological connection* is mostly useful for new product or new system development, and is rated *very high* under development of alternatives. *Why-what's stopping* is also rated *very high* because the outcome of this method is a highly comprehensive view of the problem and its subproblems, with alternative courses of action included. One potential problem may be that presenting such a broad and detailed analysis may be quite a

challenge. *AHP* is rated *very high* because, although it may begin with brainstorming as to what alternatives should be located at the bottom of the hierarchy, the level of problem abstraction represented by its hierarchy of criteria provides the opportunity to question whether or not the alternatives that are known indicate appropriate breadth for that level of abstraction.

5. Structure: breadth

This indicator does not apply to *analogy/association, boundary examination, brainstorming/brainwriting, and morphological connection, voting, conjoint analysis,* and *Bayesian analysis. NGT* and *Delphi* are rated *low* because they are direct comparison methods. *Why-what's stopping, disjointed incrementalism, matrix evaluation, outranking, MAUT/MAVT,* and *AHP* are rated *high* because they do not limit the number of criteria or factors considered in the analysis.

6. Structure: depth

This indicator does not apply to *analogy/association, boundary examination, brainstorming/brainwriting, morphological connection, voting, conjoint analysis*, and *Bayesian analysis*. *NGT* and *Delphi* are rated *low* because they are direct comparison methods. Lack of measurement and of theoretical foundation for *disjointed incrementalism* and *matrix evaluation* prevent them from constructing a deep structure, hence they are rated *low*. *Goal programming, outranking,* and older *MAUT* are rated *low* because they have no provision for subcriteria. *Why-what's stopping* and *AHP* are rated high because they do not limit the level of detail of the analysis with respect to breaking down criteria into subcriteria, sub-subcriteria and so on.

7. Analysis: faithfulness of judgments

This indicator, and all others here, do not apply to *analogy/association, boundary examination, brainstorming/brainwriting, morphological connection,* and *why-what's stopping. NGT* and *Delphi* include a voting process to determine which alternative is preferred by the majority of the group members. However, there is an opportunity to use them together with a ratio scale evaluation method like the AHP. *Voting* is rated *low* because it uses an ordinal scale. *Disjointed incrementalism, matrix evaluation,* and *outranking* are rated *medium* because they involve assigning numbers which can be assumed to represent intensity of importance better than the ordinal rating of voting, for example. *MAUT/MAVT* is rated *high* because intensity of preference is derived from lottery judgments which are once removed from direct elicitation of preferences, and *AHP* is rated *very high* because it elicits elementary judgments.

8. Breadth and depth of analysis (analysis)

Voting is rated *low* because it involves judgment, but not analysis. *Disjointed incrementalism, matrix evaluation, goal programming, conjoint analysis,* and *Bayesian analysis* are rated *medium* because they are structurally constrained. *MAUT/MAVT* is rated *high* because it provides more structural flexibility but it is difficult to go back and review previous analysis. The AHP is rated *very high* because its structural flexibility facilitates in-depth analysis of a problem. It also provides inconsistency and incompatibility measures to indicate if some improvement in judgments, and some effort to align perceptions among group members are required. Its supporting software provides the information as to where the sources of inconsistency and incompatibility are.

9. Fairness: cardinal separation of alternatives

This indicator is applicable only to *voting, outranking, MAUT/MAVT*, and *AHP*. *Voting* is rated *low* because it uses an ordinal scale, and the others are rated *high* because they use cardinal scales. Arrow's theorem indicates that any ordinal preference relation, be it expressed as a set of pairwise comparisons or point allocations, does not treat the alternatives fairly.

10. Fairness: prioritizing group members

This indicator is also applicable only to *voting, outranking, MAUT/MAVT*, and *AHP*. *Voting* is rated *low* because fairness is operationalized using head counting with no regard to intensity of preference, which has been argued as unsatisfactory [Dummett, 1984]. *Outranking* and *MAUT/MAVT* treat individual members of the group equally. They may in fact, implicitly give them unequal weights, as for example, by giving the boss's opinion greater accord than that of other members of the group in constructing their measures, but the lack of method requires that the relative weights can only be assigned rather arbitrarily. With the *AHP*, it is at the decision-maker's discretion to determine what concept of fairness is appropriate. A hierarchy can be structured, with the different individuals at the bottom of the hierarchy. The criteria levels may include area of responsibilities or expertise that can be used to prioritize the individuals.

11. Fairness: consideration of other actors and stakeholders

This indicator is not applicable to *analogy/association*, *boundary examination*, *brainstorming/brainwriting*, and *morphological connection* because they do not involve problem analysis. It is unlikely that this indicator would be applicable to *Bayesian analysis* because of its complex cause-effect relationship with the states of nature, hence it is rated *low*. *Conjoint analysis* is rated *low* because it may be possible for a creative user to represent other actor's concerns in its model. *NGT* and *Delphi* are rated *low* because they do not make explicit this concern, which might be made implicit by individual members of the group. *Matrix evaluation* is rated *low* because of its highly constrained structural representation and non-quantifiable analysis. *Outranking* is rated *low* because it obtains a decision with incomplete information, and its theoretical foundation is not yet settled even for the most fundamental issues, making it unlikely that this concern would be addressed

and settled once and for all. *MAUT/MAVT* is rated *low* because, although it may incorporate this concern as one of its criteria, its limited structural representation makes it difficult to address the possible diversity of actors. *Why-what's stopping* and *disjointed incrementalism* may address the issue explicitly, but qualitatively, and are rated *medium*. It appears that the *AHP* is the only method that facilitates for a group to explicitly include other actor's concerns in detail as parts of the problem structure, and quantify them, hence it is rated *high*.

12. Scientific and mathematical generality

This indicator is not applied to *analogy/association*, *boundary examination*, *brainstorming/brainwriting*, *morphological connection, why-what's stopping, NGT, Delphi*, because they do not involve problem analysis. *Disjointed incrementalism* and *matrix evaluation* are rated *low* because they do not involve mathematics. *Voting* is rated *medium* because there are many procedures proposed for aggregating ordinal votes, with or without axiomatization. The ones that are axiomatized are usually mathematically complex to deal with the impossibility inherent in ordinal group aggregation. *Goal programming*, and *conjoint analysis* are rated *medium* because they do not involve axiomatization. *Outranking* is rated *medium* because it is not yet axiomatized. *Bayesian Analysis,* and *MAUT/MAVT* are rated *high* because they are axiomatically solid but their generalization's have considerable mathematical difficulties. The AHP is rated *very high* because its mathematical foundation is generalizable without additional assumptions.

13. Applicability to intangibles

This indicator is not applied to *analogy/association*, *boundary examination*, *brainstorming/brainwriting*, *morphological connection, why-what's stopping, NGT, Delphi*, and *voting* because they do not involve problem analysis. *Disjointed incrementalism* and *matrix evaluation* are qualitative methods and are rated *low*. *Goal programming* and *conjoint analysis* may incorporate intangibles in their model, but they must be represented by tangibles with absolute measurement, hence they are rated *medium*. *Outranking* and *MAUT/MAVT* are rated *medium* because they must use absolute measurement. *Medium* is probably a generous judgment because *MAUT* is riddled with unresolved paradoxes and problems and "the standard theory is being challenged on several grounds from both within and outside economics [Machina, 1987]." *Bayesian analysis* deals with the probability of events, and is rated *medium* because it often contrives and guesses at its prior probabilities without adequate scientific justification. *AHP* is rated *very high* because its fundamental measurement ensures its applicability to intangibles naturally, that gives discretion to the user whether to use relative, ideal or absolute measurement [Saaty, 1990].

14. Psychophysical applicability

Psychophysical applicability does not apply to *voting, goal programming,* and *conjoint analysis*. *Disjointed incrementalism, matrix evaluation* and *Bayes analysis* are rated *low* because psychophysical law is irrelevant. *Outranking* and *MAUT/MAVT* are rated *medium*

because they generally do not incorporate the psychophysical phenomenon. If they do, it would complicate the mathematical representation of the theory considerably. *AHP* is rated *high* because in many examples, its priority scales approach has produced measurement of responses to physical stimuli that corresponded closely to the normalized values of physical measurement of those stimuli in the homogeneous ranges in which they were examined.

15. Applicability to conflict resolution

There are only two theories applicable to conflict resolution, game theory which is based on the utility theory, and the AHP. They are both rated *medium.*

16. Validity of the outcome (what if)

This indicator is not applied to *analogy/association*, *boundary examination*, *brainstorming/brainwriting*, *morphological connection, why-what's stopping, NGT,* and *Delphi*, because they do not involve problem analysis. *Voting* is rated *low* because it uses ordinal measurement with no problem representation, *Disjointed incrementalism* and *matrix evaluation* are rated *medium* because they are limited in terms of measurement and model representation, *Goal programming, conjoint analysis,* and *Bayes analysis* are rated *medium* because their main concern is with computation, not with problem representation. *Outranking* and *MAUT/MAVT* are rated *medium* because they use cardinal measurement with a relatively simplified model representation. *AHP* is rated *high* because its reliance on ratio scales derived from paired comparisons, enabling one to model a problem by ordering its elements and levels in a fine, structured way to legitimize the meaningfulness of the comparisons, and also because different ratio scales can be multiplied and divided to obtain an outcome from hierarchies of benefits, costs, risks, and opportunities.

Research indicates that sometimes a method does not perform as intended. For example, instead of directing decision makers to profitable investment, a series of experiments indicate that the use of the Boston Consulting Group (BCG) matrix increases the subject's likelihood of selecting less profitable investment [Armstrong and Brodie, 1994] due to misuse of the method [Wensley, 1994].

6.5 Discussion

An AHP model shown in Figure 6.1 was developed to obtain an overall rating of the methods, but no numbers are included in it. Our intention is to offer an opportunity for evaluating the methods to the reader. We must confess that we are not entirely unbiased in our preference and hence we have avoided trying to be too explicit. This exercise provides an example of the absolute rating approach of the AHP to do ranking. The reader would have to use numbers to represent his or her judgments according to the fundamental scale of the AHP. Table 6.2 gives the criteria and the intensities for each criterion together with their local priorities. We believe that we have thought through carefully about the relative importance of the intensities for each criterion that they may be of use to an individual who wants to establish priorities for the criteria, and then select the appropriate intensity for

each method on each criterion and then obtain the total for that method.

We would like to note that not every decision problem needs to be analyzed in a comprehensive and elaborate way as one does with the AHP. We see two kinds of problems. Simple ones that do not require a great deal of deliberation or the need to justify to someone else or to maintain a record. For such problems one often does not stop to consciously structure the decision and only uses brief flashes of thought to make trade-offs without a need for arithmetic. In such cases even if the decision is not optimal, one soon forgets the incident and goes on to make the same kind of decisions in the same or in a modified way without a feeling of inadequacy. One may refer to this kind of decision making as a single overall criterion decision, or decision making in bulk. Such decision making is often done with the aid of straight brainstorming.

The other kind of decision is the one we have discussed in this dissertation that requires detailed analysis with structure and judgments and a methodology to deal with tangibles and intangibles, benefits and costs and the like. The proponents of these methods will argue that their method is best for this or that reason, and the debate, if ever, will be eventually settled empirically and also on theoretical grounds when there is sufficient understanding of the field that mathematicians will be able to determine the correctness and accuracy of the various claims. The AHP is a tried and tested tool for that purpose. Whenever one has to make numerical trade-offs among the criteria and among the alternatives, the problem of the scale origin of the numbers becomes preeminent. In that case, proponents of different techniques must show that all the operations they perform on their numbers, preserve their integrity. With the exception of the AHP and its derived (not assumed) ratio scales from judgments of homogeneous elements based on an absolute Fundamental Scale, the question of what kind of number represents the very final decision has been the Achilles' heel of multicriteria decision making.

6.6 Efficiency

The efficiency of a process is very dependent on how it is organized and on its leader and could not be used as one of the criteria. Nevertheless, it is an important consideration. Here we discuss how a facilitator can ensure group efficiency with the AHP. The AHP makes group decision making intrinsically efficient for at least three reasons: (1) it provides a framework and tools for group collaboration that systematize the group process, (2) it enables the group to break its task into distinct sub-tasks, with each managed almost independently with respect to the manpower allocated and to the group techniques used, and (3) it provides feedback measures to facilitate judgment improvement while allowing a certain degree of inconsistency in the judgments. The fact that the AHP involves redundancy in judgments, does not make it an inefficient method, because one can always choose to do quick and dirty evaluation by eliminating redundancy, although one may get an inaccurate result. A facilitator can enhance group efficiency as follows:

(1) <u>Divide group task to minimize man-hour utilization</u>

Group decision making is said to be efficient if it consumes a reasonable amount of resources, mainly in terms of the time and number of people involved. Research indicates

that groups are relatively more costly when measured by the man-hours spent. The larger the group size, the more the hours spent by each person [Shaw, 1932; Taylor and Faust, 1952; Marquart, 1955], hence the smaller the group size, the more efficient the group process. The AHP enables a group to reduce the amount of time spent by its members by:

(a) Breaking the group task down into a set of sub-tasks: constructing a hierarchy or parts of a hierarchy, providing judgments for a given hierarchy or parts of a hierarchy.

(b) Determining the smallest number of people possible to complete each sub-task. If one person can do a job well, no need to assign a team to do the job. However, fair distribution of tasks needs to be considered.

(c) Allowing reduced number of judgments (see Chapter 5).

(d) Allowing an acceptable level of inconsistency.

(e) Improving judgments by looking for both the most inconsistent judgment and the individual that causes it.

However, for group decision-making, efficiency is also manifested in how well the group process moves, does not get bogged down, and gets to the point. Pursuing efficiency is an achievement orientation goal. Sometimes, the concern for the maintenance goal may require that the effort towards greater efficiency be somewhat relaxed: assigning more people than is necessary, allowing lengthy discussion on important matters, or organizing expert presentations rather than asking them to make judgments for the group.

(2) Use the underlying intentions of deferred judgment and structured communication.

The AHP generalizes the application of the concepts of deferred judgment in brainstorming, and structured communication in the NGT. It separates the process of problem structuring (which in turn consists of two distinct processes of identifying the elements of the problem and of defining dominance relations), judgment elicitation, and evaluation. The group can conduct brainstorming, NGT, or Delphi sessions to identify alternatives and criteria efficiently.

During problem structuring, differences with regard to structuring the hierarchy do not need to be resolved. This is a win-win kind of disagreement. The group effort should be focused on establishing dominance relations among elements rather than arguing too early whether or not an element should or should not be in a hierarchy. If some elements do not fit logically in a hierarchy, consider another hierarchy. For example, cost criteria cannot be logically located in the same hierarchy with benefit criteria. Facilitate efficiency in judgment elicitation by ensuring an optimal size of the hierarchy. If the structure is too deep, encourage the group to remove some of its levels and learn later whether or not the resulting ambiguity creates difficulties in judgment elicitation.

During judgment elicitation, obtaining a pairwise comparison judgment should be the end of a discussion. For example, if the group agrees that a criterion is strongly more important than another, the members do not have to agree on the reasoning behind the judgment.

(3) Balance between consensus and voting

What makes conventional group decision making demanding is that it is expected to reach consensus on which alternative(s) to select, although the group may generally agree on the criteria and their relative importance. It is reasonable to assume that it would be easier to reach consensus on comparing two things than on many things at once. The AHP makes it even easier by not requiring a group to reach consensus on a judgment. If voting is necessary for reasons of efficiency, its effect must be minimized. Voting indicates an individual's incompatibility with the group and may lead to judgment inconsistencies. A narrow range of judgments can be ensured by discussing the differences first and selecting the appropriate individuals to vote. In fact, disagreement and incompatibility between an individual and the group provides an opportunity for organizational learning. The majority are not always right.

6.7 Conclusion

This chapter has provided us with a challenging opportunity to bring many criteria and many methods under one umbrella, which again illustrates the simplicity and practicality of the AHP to deal with intangibles and with a wide-ranging set of alternatives. We believe that our schematization is a good start and may eventually be improved upon in subsequent revisions and extensions of the criteria used and in debating the importance of these criteria.

Download PDF of this page for printing
Continue to Table 6.1

Copyright © 2001 Expert Choice, Inc.