Methods for Evaluating Multi-Level Decision Trees in Imprecise Domains

Jim Johansson

Dept. of Information Technology and Media
Mid Sweden University
SE-851 70 Sundsvall, Sweden
jim.johansson@miun.se

Abstract

Over the years numerous decision analytical models based on interval estimates of probabilities and utilities have been developed, and a few of these models have also been implemented into software. However, only one software, the Delta approach, are capable of handling probabilities, values and weights simultaneously, and also allow for comparative relations, which are very useful where the information quantity is limited. A major disadvantage with this method is that it only allows for single-level decision trees and cannot non-trivially be extended to handle multi-level decision trees. This paper generalizes the Delta approach into a method for handling multi-level decision trees. The straight-forward way of doing this is by using a multi-linear solver; however, this is very demanding from a computational point of view. The proposed solutions are instead to either recursively collapse the multi-level decision tree into a single-level tree or, preferably, use backward induction, thus mapping it to a bilinear problem. This can be solved by LP-based algorithms, which facilitate reasonable computational effort.

Introduction

Matrix, tree, and influence diagram are extensively used decision models, but since precise numeric data are normally the only input in the models, and since this type of data rarely can be obtained, they are less suitable for real-life decision-making. Various types of sensitivity analysis might be a partial solution for these tools, but even in small decision structures, such analysis is difficult to manage. (Danielson and Ekenberg 1998)

A number of models, which include representations allowing for imprecise statements, have been developed over the past 50 years. However, the majority of the models focus more on representation, and less on evaluation and implementation. (Danielson and Ekenberg 1998) Some approaches concerning evaluation have been suggested by, e.g., (Levi 1974) and (Gärdenfors and Sahlin 1982), but do not address computational and implementational issues.

A systematic approach for interval multi-criteria decision analysis, addressing computational issues, is PRIME presented by (Salo and Hämäläinen 2001). PRIME is, however, primarily developed for multi-criteria decisions under certainty, thus there is no support for the construction and evaluation of decision trees involving several uncertain outcomes. There are also other software, see e.g. (Olson 1996) for a survey, but very few is capable of handling both probabilities and attributes simultaneously and does not allow for comparative relations, which is quite useful in situations where the information quantity is very limited. Most of them also provide very little information when the interval values of the alternatives are overlapping.

The theories by (Park and Kim 1997) is one of the more interesting approaches capable of handling uncertain events in multi-criteria decisions, and also address some computational issues. The result is unfortunately only based on ordinal ranking with no support of sensitivity analysis and can not handle multi-level trees.

The Delta method suggested by (Danielson 1997) is by far the most interesting approach of solving real-life decision situations. This approach can handle weights, probabilities, values and comparative relations simultaneously and is inspired by earlier work on handling decision problems involving a finite number of alternatives and consequences, see e.g., (Malmnäs 1994).

However, the main disadvantage with the Delta method is that it only allows for single-level trees and cannot non-trivially be extended to a multi-level approach. Since multi-level decision trees appear naturally in many real-life situations (e.g. events with dependent outcomes), it is of great importance being able to evaluate such a representation.

This paper combines the Delta approach with a multi-level decision tree structure, making the approach much more user friendly. For the approach to be interactively useful, it is less suitable to use a standard solver for bi- or multi-linear optimization problems. Instead we use a solver based on reductions to linear programming problems, solvable with the Simplex method (Danielson and Ekenberg 1998).

The Delta Method

The Delta method has been developed for real-life decision situations where imprecise information, in the form of interval statements and comparative relations, are provided. This does not force the decision maker to precise numerical numbers in cases where this is unrealistic. (Danielson and Ekenberg 1998) The method allows for two kinds of user statements. *Interval statements* of the form: "the probability of c_{ij} is between the numbers a and b" are translated into $p(c_{ij}) \in [a,b]$. *Comparative relations* of the form: "the value of c_{ij} is greater than the value of c_{ik}" is translated into an inequality $v_{ij} > v_{ik}$.

The conjunction of probability constraints of the types above, together with the normalization constraint $\sum p(c_{ij}) = 1$ for each alternative a_i involved, is called the *probability base* (P). The *value base* (V) consists of value constraints of the types above, but without the normalization constraint. A collection of interval constraints concerning the same set of variables is called a *constraint set*. For such a set of constraints to be meaningful, there must exist some vector of variable assignments that simultaneously satisfies each inequality, i.e., the system must be *consistent*.

Evaluation of the alternatives is necessarily made pair-wise when the alternatives are dependent, otherwise the result will be incorrect. (Danielson 1997) The main principle used for evaluation is the strength concept, a generalization of $PMEU$.

Definition: The *strength* δ_{ij} of alternative a_i compared to a_j denotes the expression $EV(a_i) - EV(a_j) = \sum_k p(c_{ik}) \cdot v(c_{ik}) - \sum_l p(c_{jl}) \cdot v(c_{jl})$.

To analyze the strength of the alternatives, $\max(\delta_{ij})$ is calculated, which means that the feasible solutions to the constraints in P and V that are most favorable to $EV(a_i)$ and demeaning to $EV(a_j)$ are chosen. In similar manners $\min(\delta_{ij})$ is calculated. Thus, the concept of strength expresses the maximum differences between the alternatives under consideration. It is however used in a comparative way so that the maximum and minimum is calculated. The strength evaluation requires bilinear optimization which is computationally demanding using standard approaches, but this can be reduced to linear programming problems, solvable with the Simplex method. (Danielson and Ekenberg 1998)

Often the probability, value and weight distributions are not uniform; rather some points are more likely than others. The Delta approach is taking this into account using triangle shaped distributions in the form of an interval and a contraction point (the most likely value), see figure 1.

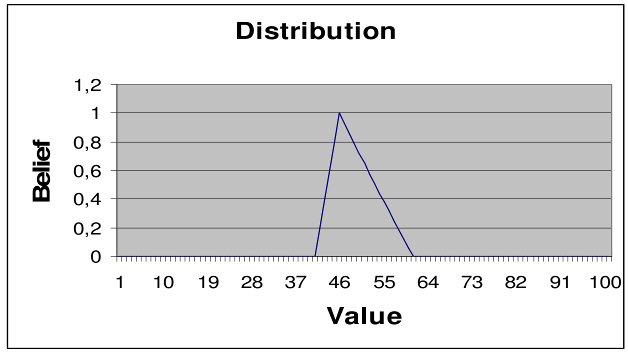

Figure 1: Distribution of probabilities, values and weights, using interval and contraction point.

A problem with evaluating interval statements is that the results could be overlapping, i.e., an alternative might not be dominating[1] for all instances of the feasible values in the probability and value bases. A suggested solution is to further investigate in which regions of the bases the respective alternatives are dominating. For this purpose, the *hull cut* is introduced in the framework.

The hull cut can be seen as generalized sensitivity analyses to be carried out to determine the stability of the relation between the alternatives under consideration. The hull cut avoids the complexity in combinatorial analyses, but it is still possible to study the stability of a result by gaining a better understanding of how important the interval boundary points are. This is taken into account by cutting off the dominated regions indirectly using the hull cut operation. This is denoted cutting the bases, and the amount of cutting is indicated as a percentage p, which can range from 0 % to 100 %. For a 100 % cut, the bases are transformed into single points (contraction points), and the evaluation becomes the calculation of the ordinary expected value.

Multi-Level Decision Trees

As has been described, the Delta approach is a single-level approach, not able to construct multi-level decision trees. However, single-level trees and decision tables are only two different ways of describing the situation containing the same amount of data. In situations involving multiple choices in a certain order, or where the outcome of one event affects the next, the multi-level decision tree is much more appropriate and contains more information of the decision situation. Multi-level decision trees are also very useful in complex decision situations, where the decision tree provides a graphical representation of the decision and shows all the relations between choices and uncertain factors (Hammond et al. 2002).

Decisions under risk is often given a tree representation, described in, e.g., (Raiffa 1968), and consists of a base,

[1]Alternative i dominates alternative j iff $\min(\delta_{ij}) > 0$.

representing a decision, a set of intermediary (event) nodes, representing some kind of uncertainty and consequence nodes, representing possible final outcomes, se figure 2.

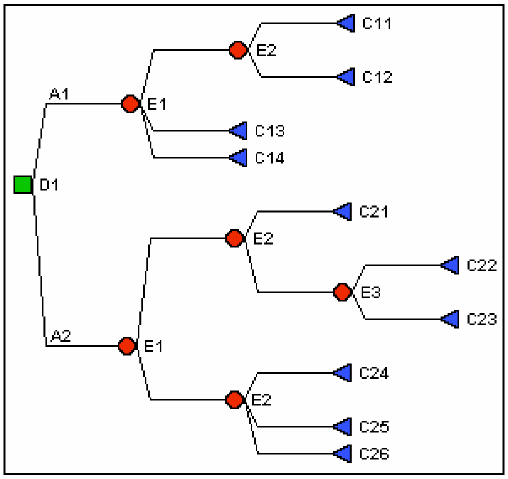

Figure 2: A multi-level decision tree.

Usually the maximization of the expected value is used as the evaluation rule. The expected value of alternative a_i is calculated according to the following formula:

$$EV(a_i) = \sum_{j=1}^{n_i} (p_{e_1}(c_{ij}) \cdot p_{e_2}(c_{ij}) \cdot \ldots \cdot$$

$p_{e_{m_i-1}}(c_{ij}) \cdot p_{e_{m_i}}(c_{ij}) \cdot v(c_{ij}))$, where $p_{e_k}(c_{ij})$ denote the probability of event e_k (towards c_{ij}) and $v(c_{ij})$ the value of the consequence c_{ij}. $k \in [1,\ldots,m_i]$ where m_i is the level, in terms of depth, where consequence c_{ij} is located. n_i is the number of consequences in the alternative a_i.

Since neither the probabilities nor the values are fixed numbers, the evaluation of the expected value yields multi-linear objective functions.

Tree Collapse

A multi-linear function is very problematical from a computational viewpoint, so to maintain all Delta features, including relations between arbitrary consequences; one solution could be to recursively collapse the multi-level tree into a single-level, thus mapping it to a bilinear problem. The collapse is straight-forward in that each path in the tree is replaced by a consequence representing the joint event chain leading up to the final consequence.

The probability of a consequence in a collapsed tree is defined as a *joint probability* (φ), and the probability of an event on the path from the decision node to the consequence is defined as a *local probability* (γ). The upper and lower joint probability is calculated through multiplying the upper and lower local probabilities respectively. This is the purpose of algorithm 1.

Algorithm 1

A *unique path* in a decision tree is a set of edges $E(c_i) = \{e_1(c_i),\ldots,e_n(c_i)\}$ leading from the root node to a consequence c_i, where n is determined by the level of depth of c_i. Given a unique path:

Let $p_{c_i}^{\varphi\min}$ denote the lower joint probability for consequence c_i occurring, such that $p_{c_i}^{\varphi\min} = p_{e_1}^{\gamma\min}(c_i) \cdot \ldots \cdot p_{e_n}^{\gamma\min}(c_i)$. $p_{e_j}^{\gamma\min}(c_i)$ denotes the lower local probability of the uncertain event represented by edge e_j (on the unique path towards c_i, i.e. not necessarily independent), and where j is the level, in terms of depth, from the alternative.

Let $p_{c_i}^{\varphi\max}$ denote the upper joint probability for consequence c_i occurring, such that $p_{c_i}^{\varphi\max} = p_{e_1}^{\gamma\max}(c_i) \cdot \ldots \cdot p_{e_n}^{\gamma\max}(c_i)$. $p_{e_j}^{\gamma\max}(c_i)$ denotes the upper probability of the uncertain event represented by edge e_j (on the unique path towards c_i, i.e. not necessarily independent), and where j is the level, in terms of depth, from the alternative.

Three major issues have arisen during the collapse; distribution of contraction points, probability propagation and incongruence of intermediary values. The remainder of the chapter will discuss those properties as they are calculated in a collapsed tree.

Distribution of Contraction Points

Given a multi-level asymmetric decision tree with no probabilities explicitly set by the decision maker, a tree collapse distributes the probability evenly, since the collapse equals the probability of all consequences. However, with no probabilities set the most intuitive implication, when no other information is available, is that the probability distribution is between 0 and 1 and the most likely probability (the contraction point) is dependent on the level of depth where the consequence is located.

Given $p_{e_1}^{\gamma}(c_1,c_2), p_{e_1}^{\gamma}(c_3), p_{e_2}^{\gamma}(c_1), p_{e_2}^{\gamma}(c_2) \in [0,1]$, where $p_{e_1}^{\gamma}(c_1,c_2)$ is the probability interval preceding $p_{e_2}^{\gamma}(c_1), p_{e_2}^{\gamma}(c_2)$; the most likely probability, i.e. the joint contraction point $\overline{k}_{c_i}^{\varphi}$, should, in an intuitive interpretation, be $\overline{k}_{c_3}^{\varphi} = \frac{1}{2}$ and $\overline{k}_{c_1}^{\varphi}, \overline{k}_{c_2}^{\varphi} = \frac{1}{4}$, if no explicit contraction points are set. However, this does not hold when the decision tree is collapsed to a single-level. Since no probability is explicitly set, the tree collapse assumes that the three consequences occur with the same probability, thus $\bar{k}^{\varphi}_{c_1}, \bar{k}^{\varphi}_{c_2}, \bar{k}^{\varphi}_{c_3} = \frac{1}{3}$.

The approach for solving this matter is to use the most likely probability for each edge between the root and the consequence, i.e. the local contraction point, and multiply them, according to algorithm 2.

Algorithm 2
Given a unique path:

- Let $\bar{k}^{\varphi}_{c_i}$ denote the joint contraction point for consequence c_i, such that $\bar{k}^{\varphi}_{c_i} = \bar{k}^{\gamma}_{e_i}(c_i) \cdot \ldots \cdot \bar{k}^{\gamma}_{e_n}(c_i)$, where $\bar{k}^{\gamma}_{e_j}(c_i)$ denotes the local contraction point of edge e_j, (towards c_i) where j is the level, in terms of depth, from the alternative.

Probability Propagation

The problem with probability propagation becomes evident in a multi-level decision tree having at least two levels, where an outcome at level one has two (or more) children nodes. The joint probability of the two nodes are $p^{\varphi}_{c_1} = p^{\gamma}_{e_1}(c_1, c_2) \cdot p^{\gamma}_{e_2}(c_1)$ and $p^{\varphi}_{c_2} = p^{\gamma}_{e_1}(c_1, c_2) \cdot p^{\gamma}_{e_2}(c_2)$, where $p^{\gamma}_{e_x}(c_i)$ denotes the local probability of the event in level x (towards c_i), thus $p^{\varphi}_{c_i}$ depends on the probability of its predecessors.
In a tree collapse with the probability explicitly set to e.g., $p^{\gamma}_{e_2}(c_1), p^{\gamma}_{e_2}(c_2) = x$, and no probability set in e_1, thus $p^{\gamma}_{e_1}(c_1, c_2) \in [0,1]$, results in $p^{\varphi}_{c_1} \in [0, x]$ and $p^{\varphi}_{c_2} \in [0, x]$. However, since c_1 and c_2 share the same predecessor e_1, the probability should be equal at all times, thus $p^{\varphi}_{c_1} = p^{\varphi}_{c_2}$, e.g. if $p^{\varphi}_{c_1} = b$, then $p^{\varphi}_{c_2} = b$. Without some kind of comparative relation like $p^{\varphi}_{c_1} = p^{\varphi}_{c_2}$ explicitly set, the tree collapse assumes $p^{\varphi}_{c_1}$ and $p^{\varphi}_{c_2}$ as being independent of each other, thus not necessarily $p^{\varphi}_{c_1} = p^{\varphi}_{c_2}$.

Example
As can be seen in Figure 3, the probability propagation of c_2 and c_3 are $p^{\varphi}_{c_2}, p^{\varphi}_{c_3} \in [0, 0.5]$, which is correct given the input.

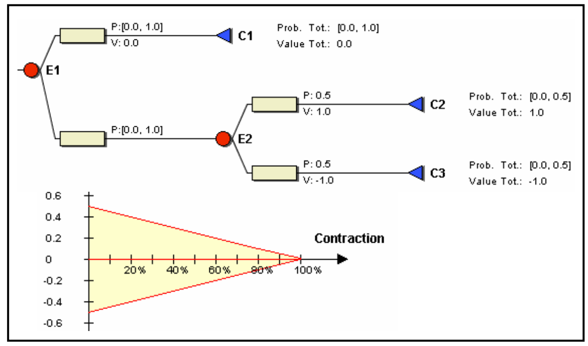

Figure 3: Incorrect evaluation result due to the problem with probability propagation.

Since they share the same predecessor and $p^{\gamma}_{e_2}(c_1), p^{\gamma}_{e_2}(c_2) = 0.5$, and $U^{\varphi}_{c_1} = 0$, $U^{\varphi}_{c_2} = 1$, $U^{\varphi}_{c_3} = -1$, the expected utility should be $EU^{\max}_{a_1}, EU^{\min}_{a_1} = 0$, but the result is $EU^{\max}_{a_1} = 0.5$ and $EU^{\min}_{a_1} = -0.5$, which can be seen at 0% contraction in the evaluation graph.

Incongruence of Intermediary Values

Intermediary value statements lead to similar problems as the probability propagation. The problem is that when performing the tree collapse, the final value of a consequence $u^{\varphi}_{c_{ij}}$ becomes independent of shared predecessors (intermediary nodes). Given $u^{\gamma}_{e_2}(c_1) = a$, $u^{\gamma}_{e_2}(c_2) = a + b$, where $b > 0$, and the common predecessor $u^{\gamma}_{e_1}(c_1, c_2) \in [c, d]$, this will result in $u^{\varphi}_{c_1} \in [a + c, a + d]$ and $u^{\varphi}_{c_2} \in [a + b + c, a + b + d]$ Since $u^{\varphi}_{c_1}$ and $u^{\varphi}_{c_2}$ may be overlapping, i.e., $(a + d) > (a + b + c)$, and the information about the predecessor will be lost in the tree collapse, a final comparative relation (relations between final values) saying $u^{\varphi}_{c_1} > u^{\varphi}_{c_2}$ would be accepted, despite that $u^{\gamma}_{e_2}(c_1) = a$, $u^{\gamma}_{e_2}(c_2) = a + b$.

Example
Figure 4 shows a decision situation with the intermediate values $u^{\gamma}_{e_1}(c_1, c_2) = [0,1000]$, $u^{\gamma}_{e_2}(c_1) = -100$ and $u^{\gamma}_{e_2}(c_2) = 0$.

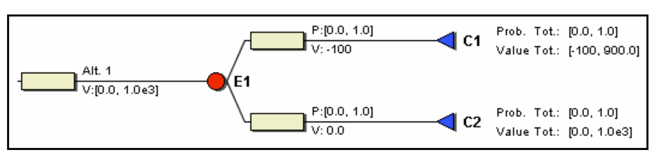

Figure 4: Intermediary value statements

For these consequences, the minimum and maximum joint values are calculated:

$$u_{c_1}^{\varphi \min} = u_{e_1}^{\gamma \min}(c_1, c_2) + u_{e_2}^{\gamma}(c_1) = 0 + (-100) = -100$$

$$u_{c_1}^{\varphi \max} = u_{e_1}^{\gamma \max}(c_1, c_2) + u_{e_2}^{\gamma}(c_1) = = 1000 + (-100) = 900$$

$$u_{c_1}^{\varphi} \in [-100, 900]$$

$$u_{c_2}^{\varphi \min} = u_{e_1}^{\gamma \min}(c_1, c_2) + u_{e_2}^{\gamma}(c_2) = 0 + 0 = 0$$

$$u_{c_2}^{\varphi \max} = u_{e_1}^{\gamma \max}(c_1, c_2) + u_{e_2}^{\gamma}(c_2) = 1000 + 0 = 1000$$

$$u_{c_2}^{\varphi} \in [0, 1000]$$

When looking at $u_{c_1}^{\varphi} \in [-100, 900]$ and $u_{c_2}^{\varphi} \in [0, 1000]$, the comparative relation $u_{c_1}^{\varphi} > u_{c_2}^{\varphi}$ could be accepted. However, the information of the intermediate relation has disappeared in the tree collapse and the comparative relation is unfortunately inconsistent since $u_{c_1}^{\varphi} \in [u_{e_1}^{\gamma \min} - 100, u_{e_1}^{\gamma \max} - 100] <$ $u_{c_2}^{\varphi} \in [u_{e_1}^{\gamma \min} + 0, u_{e_1}^{\gamma \max} + 0]$.

Backward Induction

The abovementioned problems could be solved using a modified version of the backward induction (also known as the rollback method) instead of a tree collapse. Backward induction is an iterative process for solving finite extensive form or sequential games. A drawback with the backward induction is that statements expressing relations between nodes with different direct predecessors cannot be set. However, relations between nodes having the same direct predecessor are still possible with the method of backward induction. See algorithm 3.

Algorithm 3

Maximizing expected utility of strategy a_1, with intermediary values and also intermediary comparative relations present, using backward induction, gives the following expression:

1. $U_{e_i}^{\varphi \max}(c_{11},...,c_{1k}) = \sup(p_{e_i}^{\gamma}(c_{11}) \cdot u_{e_i}^{\gamma}(c_{11}) + ... + p_{e_i}^{\gamma}(c_{1k}) \cdot u_{e_i}^{\gamma}(c_{1k}))$, where $u_{e_x}^{\varphi \max}(c_{11},...,c_{1k})$ is the maximum expected utility of event e_x (towards $c_{11},\ldots,c_{1k}$), and x being the level. $p_{e_x}^{\gamma}(c_{11})$ denotes the local probability of the event in level x (towards c_{11}), $u_{e_x}^{\gamma}(c_{11})$ is the local utility of the event in level x (towards c_{11}).
2. Continuing on the next level:

$$U_{e_{(i-1)}}^{\varphi \max}(c_{11},...,c_{1m}) = \sup(p_{e_{(i-1)}}^{\gamma}(c_{11},...,c_{1k}) \cdot (u_{e_{(i-1)}}^{\gamma}(c_{11},...,c_{1k}) + U_{e_i}^{\varphi \max}(c_{11},...,c_{1k})) + ... + p_{e_{(i-1)}}^{\gamma}(c_{1l},...,c_{1m}) \cdot (u_{e_{(i-1)}}^{\varphi \max}(c_{1l},...,c_{1m}) + U_{e_i}^{\varphi \max}(c_{1l},...,c_{1m}))$$

The backward induction repeats, until finally the maximum expected utility is calculated through:

3. $EU_{a_1}^{\varphi \max} = \sup(p_{e_1}^{\gamma}(c_{11},...c_{1n}) \cdot U_{e_1}^{\varphi \max}(c_{11},...c_{1n}) + ... + p_{e_1}^{\gamma}(c_{1p},...c_{1r}) \cdot U_{e_1}^{\varphi \max}(c_{1p},...c_{1r})) + U_{a_1}^{\gamma \max}$

(no comparative relations between the alternatives)

Minimizing expected utility of strategy a_1, using backward induction, is performed in the same manner.

Concluding Remarks

Since multi-level decision trees appear naturally in many real-life situations, it is important to be able to evaluate such a representation. This paper generalizes a method for handling single-level decision trees, when vague and numerically imprecise information prevail, into a method for handling multi-level decision trees. Instrumental concepts required for the transformation has been presented as well as a discussion of some vital issues linked with the various transformations.

The proposed solutions are to either recursively collapse the multi-level decision tree into a single-level tree or, preferably, use backward induction, thus mapping it to a bilinear problem. This can be solved by the LP-based algorithms in (Danielson and Ekenberg 1998), which facilitate reasonable computational effort.

References

Danielson, M. 1997. *Computational Decision Analysis*, Ph.D. thesis, Royal Institute of Technology.

Danielson, M., and Ekenberg, L. 1998. A Framework for Analysing Decisions Under Risk, *European Journal of Operational Research*, vol. 104/3.

Gärdenfors, P., and Sahlin, N-E. 1982. Unreliable Probabilities, Risk Taking, and Decision Making, *Synthese,* Vol.53.

Hammond, J. S., Keeney, R. L., and Raiffa, H. 2002. *Smart Choices: A Practical Guide to Making Better Decisions*, Broadway Books.

Levi, I. 1974. On Indeterminate Probabilities, *The Journal of Philosophy*, Vol.71.

Malmnäs, P-E. 1994. *Towards a Mechanization of Real Life Decisions*, Logic and Philosophy of Science in Uppsala.

Olson, D.L. 1996. *Decision Aids for Selection Problems*, Springer-Verlag.

Park, K.S., and Kim, S.H. 1997. Tools for interactive multiattribute decisionmaking with incompletely identified information, *European Journal of Operational Research 98*.

Raiffa, H. 1968. *Decision Analysis*, Addison-Wesley, Reading, Massachusetts.

Salo, A.A, and Hämäläinen, R.P. 2001. Preference Ratios in Multiattribute Evaluation (PRIME) – Elicitation and Decision Procedures Under Incomplete Information, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 31/6.