Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: the assessment of the value of a candidate.*

Didier Dubois
Institut de Recherche en Informatique de Toulouse (IRIT) – CNRS
Université Paul Sabatier,
118 route de Narbonne
31062 Toulouse Cedex 4, France

Michel Grabisch
Thomson-CSF,
Corporate Research Laboratory
Domaine de Corbeville
91404 Orsay, France

Henri Prade
Institut de Recherche en Informatique de Toulouse (IRIT) – CNRS
Université Paul Sabatier,
118 route de Narbonne
31062 Toulouse Cedex 4, France

Philippe Smets
IRIDIA
- Université Libre de Bruxelles
50 av. Roosevelt, CP 194-6
1050 Bruxelles, Belgium

June 24, 2000

*This paper is a revised, unified and extended version of two conferences papers [Dubois et *al.*, 1999a] and [Dubois et *al.*, 1999b], the second being only devoted to the possibility approach.

Abstract

The problem of assessing the value of a candidate is viewed here as a multiple combination problem. On the one hand a candidate can be evaluated according to different criteria, and on the other hand several experts are supposed to assess the value of candidates according to each criterion. Criteria are not equally important, experts are not equally competent or reliable. Moreover levels of satisfaction of criteria, or levels of confidence are only assumed to take their values in linearly ordered scales, whose nature is rather qualitative. The problem is discussed within two frameworks, the transferable belief model and the qualitative possibility theory. They respectively offer a quantitative and a qualitative setting for handling the problem, providing thus a way to emphasize what are the underlying assumptions in each approach.

1 Introduction

1.1 Preamble

The purpose of this paper is essentially to show how to apply two uncertainty modeling frameworks, the transferable belief model (TBM) and the qualitative possibility theory (QPT) approaches, to a complex problem that combines, voluntarily, many difficulties. The problem concerns the assessment of the value of a candidate evaluated by several experts using several criteria. Thus it exhibits both a multiple criteria decision making and a data fusion facet. Moreover it is assumed that for each criteria, there exists a unique value, usually pervaded with uncertainty, which represents the true value of the candidate w.r.t. the criterion. The example is only illustrative, and we do not claim the techniques exemplified here are the best methods to solve the problem. Other approaches more dedicated to multiple criteria or group decision could be defended as well. In fact the example is envisaged mainly as a multiple sensor problem in the TBM approach, while the QPT-based method distinguishes the fusion of expert opinions from the multiple criteria aggregation. The paper shows how the two representation frameworks can be used in practice, rather than discussing the appropriateness of the solutions proposed w.r.t. the value assessment problem.

1.2 Introducing the example

The problem of assessing the value of a candidate (it may be a person, an object, or any abstract entity for instance) is often encountered in practice. It is usually a preliminary step before making a choice. Such a problem can be handled in different manners depending on what kind of information the evaluation is based. One may have expert generic rules which aim at classifying candidates in different categories (e.g., 'excellent','good', ... ,'very bad'). One may have a base of examples made of previous evaluations from which a similarity-based evaluation of the new candidate can be performed. This corresponds to two popular approaches in Artificial Intelligence (namely, expert systems and case-based reasoning), that one might also like to combine.

In the following, the value assessment problem is rather posed in terms of multiple criteria evaluation, whose value for a given candidate can be more or less precisely assessed with some level of confidence by various experts (whose opinions are also to be fused). This problem is sometimes referred to as a subjective evaluation process [Club CRIN Flou, 1997]. For a discussion of the relation between the rule-based, case-based and criteria-based approaches, the reader is referred to [Dubois and Prade, 1994b] and [Dubois and Prade, 1997]. Note that the problem considered in the following is not simply to rank-order candidates on the basis of a set of criteria. Then pairwise comparisons could lead to a simple outranking solution. The problem here is to design a procedure where multiple experts assessments are faithfully represented and fused. As we shall see different families of solutions are possible according to the way the criteria are interacting.

An important issue in such a problem is that the assessment of the values of the criteria for a candidate by an expert and their relevance for the selection itself are often pervaded with uncertainty and imprecision. Numerical or qualitative modeling of the data can be considered. In this paper, we show how the transferable belief model (TBM, [Smets, 1998]) and qualitative possibility theory (QPT) can be applied to the value assessment problem, thus illustrating the two types of approaches, and showing how the methods can be applied. Moreover their underlying assumptions are laid bare.

The paper is organized in four main parts. The problem is first precisely stated and the questions that it raises are pointed out. Then the two proposed approaches are presented and illustrated on the same example. Lastly, a comparative discussion of the two approaches is given.

2 The multiple expert multiple criteria assessment problem

The value of a candidate for a given position has to be assessed by a decision maker (DM). For this evaluation, m criteria are used. For each criterion, the value of the candidate is assessed (maybe imprecisely) by n experts (or sources). Criteria can be rank-ordered according to their importance for the position. Their relation to the global "goodness" score of the candidate for the position may be also available. Each expert provides a precise or imprecise evaluation for each criterion and a level of confidence is attached to each evaluation by the expert who is responsible for it. The general reliability of each expert is also qualitatively assessed by DM.

Some general comments about the problem which is so far informally stated, have to be made. With a set of candidates, one may want to i) choose the best one(s); ii) rank all candidates from best to worst; iii) give a partial order between them, with possibly some incomparabilities; iv) cluster the candidates in several groups (e.g., good ones, bad ones, those with a weak point w.r.t. one criterion etc.). In fact, many methods in decision analysis proceed by a pairwise comparison of candidates which supposes all of them are known from the beginning. In this paper, we look for the global evaluation of a candidate which may be unique. If the problem were to rank-order candidates only, a lexicographic approach (based on the comparison of vectors made of scores of each candidate w.r.t. all criteria ordered according to their importance) might be sufficient. However such an approach assumes that a complete order exists between scores (which is not in agreement with the fact that the scores may be imprecise and pervaded with uncertainty).

The overall assumption of the model underlying our example is that there exist true scores for each individual criterion and for a global goodness, all attached to the candidate, the last being a function of the first ones. The problem is that these true values are unknown to the DM and must be estimated from the expert opinions.

This approach is thus an estimation problem where the experts are considered as measurement tools or sensors that determine some kind of 'objective' parameters. Other approaches based on classical multiple criteria decision making could of course have been defended. As already said, comparing the merits of such other methods is not the purpose of this paper. We focus on how to apply two particular methods, not on deciding which is good or bad.

2.1 Notations

We now introduce some notation. The candidate is denoted K, when necessary. Criteria are numbered by i: $i = 1, \dots, m$. The true (unknown) value of the level of satisfaction of criterion i for the candidate K is denoted $c_i(K)$, with $c_i(K) \in L_s$, where L_s is the ordinal scale of levels of satisfaction, e.g. $L_s = \{1, 2, 3, 4, 5\}$ with 1=very bad, ... , 5=very good. An element of L_s is denoted by s.

Experts are numbered by j: $j = 1, \dots, n$. The function π_i^j is a mapping from L_s to L_π. In the example we shall use $L_\pi = \{\emptyset, a, b, \mathbb{1}\}$ which is an ordinal scale, where $\emptyset$ corresponds to impossibility, and $\mathbb{1}$ to total possibility. The true (unknown) global score of K is denoted $c(K) \in L_s$. The confidence of the decision maker in expert j when judging criterion i is denoted γ_{ij}, and these levels are defined on an ordinal scale L_γ which can be related to L_π as we shall see. The confidence of the decision maker in expert j's opinions is denoted α_j, $j = 1, \dots, n$, and the α_j's are defined on an ordinal scale L_α. For instance, $L_\alpha = \{\emptyset, u, v, \mathbb{1}\}$ with $\emptyset$ = not confident at all, u = not very confident, ... , $\mathbb{1}$ = very confident. The levels of importance of criteria are denoted β_i, $i = 1, \dots, m$, and the β_i also belong to an ordinal scale L_β. For instance $L_\beta = \{\emptyset, e, f, g, \mathbb{1}\}$ with $\emptyset$ = not important at all, e = not very important, ... , $\mathbb{1}$ = very important.

2.2 Data for the example

For illustrative purpose we use the following example with $m = 6$ criteria and $n = 4$ experts. Imagine that in some company, a new collaborator K has to be hired for the marketing department. Six criteria are used for assessing his qualifications: analysis capacities (Ana), learning capacities (Lear), past experience (Exp), communication skills (Com), decision-making capacities (Dec) and creativity (Crea). The four experts are the directors of Marketing (Mkt), Financial (Fin), Production (Prod), and Human Resources (HR) departments. The data are summarized in Table 1. Blanks correspond to missing values, i.e., the director did not provide an evaluation for the considered criteria. In fact all missing values correspond here to the criteria i for which the expert j is considered by DM as not competent ($\gamma_{ij} = \emptyset$). They may be assimilated to a $[0, 5]$ assessment as we equate 'no opinion' to 'opinion not provided', a natural assumption here.

On the one hand this example is intentionally complex as we want to illustrate the power and flexibility of the two approaches. On the other hand, the way the criteria should be aggregated is unspecified. In real life, some of the subtleties introduced here will be irrelevant, simplifying the computation of the solution.

C_{ij}	Mkt D	Fin D	Prod D	HR D
Ana		4	2	
Lear	[2,3]	[1,5]	4	[2,4]
Exp	4			
Com	4			4
Dec	[1,5]	[1,5]	[1,2]	3
Crea	5			1

γ_{ij}	Mkt D	Fin D	Prod D	HR D	β_i
Ana	$\emptyset$	$\mathbb{1}$	$\mathbb{1}$	$\emptyset$	g
Lear	b	a	b	$\mathbb{1}$	e
Exp	$\mathbb{1}$	$\emptyset$	$\emptyset$	$\emptyset$	e
Com	$\mathbb{1}$	$\emptyset$	$\emptyset$	$\mathbb{1}$	$\mathbb{1}$
Dec	a	a	a	a	g
Crea	$\mathbb{1}$	$\emptyset$	$\emptyset$	$\mathbb{1}$	$\mathbb{1}$
α_j	$\mathbb{1}$	u	u	v	

Table 1: Upper table: Assessment by each director on each criteria of candidate K using L_s = 1, 2, 3, 4, 5. Lower table: Confidence of directors w.r.t. each criteria

3 The TBM approach

To represent belief functions, we use the next notational convention: $bel_Y^\Omega[PEV]$ $(\omega_0 \in A)$ denotes the strength of the weighted opinion (called belief) held by the agent Y about the fact that the actual world ω_0 belongs to A, a subset of the frame of discernment Ω, given the piece of evidence PEV. Thanks to the fact we use the same indexing for the basic belief assignments, the belief functions and the plausibility functions induced by each other, we can just use the notations $m_Y^\Omega[PEV]$, $bel_Y^\Omega[PEV]$ and $pl_Y^\Omega[PEV]$, the indexing indicating their mutual links.

Many beliefs are used in this example. They are analogous to subjective probabilities, except they do not satisfy the additivity rule of probability theory. Their operational meaning and their assessment are obtained by methods similar to those used by the Bayesians [Smets and Kennes, 1994].

3.1 DM's beliefs on the real value of each criteria

In this section, we consider only candidate K and all beliefs are held by one DM, so we can avoid mentioning them. Let c_i be the actual (but unknown) value of the level of satisfaction of criterion i. Let C_{ij} denote the data collected from director j on criteria i (for K), C_{ij} being the intervals given in Table 1. Missing data are equated to the whole interval $[1, 5]$.

Let $\Pi(C_{ij} = a|c_i = x)$ be the degree of (quantitative) possibility[1] that Director j states $C_{ij} = [a, a]$ with $a \in \{1, ..., 5\}$ given $c_i = x$ with $x \in \{1, ..., 5\}$. It represents the link between what the director will say and the actual value c_i. For simplicity's sake, we use the same possibility function for every criteria and every director. For each criteria, we assume that the possibility is 1 when $a = x$, .5 when $|a - x| = 1$, and 0 otherwise. More complex possibility functions could be used, the proposed one being only 'reasonable'. It reflects the idea that it is (fully) possible that the director states the actual value, 'quite possible' that he states a value at one level deviation from the actual value, and impossible at two level deviations (director can be wrong, but the error will be small).

A few classical relations are needed. We have:

$$\text{if } \Pi(A) = p \text{ then } pl(A) = p, \tag{1}$$

$$\Pi(X|Y)\Pi(Y) = \Pi(Y|X)\Pi(X), \tag{2}$$

$$\Pi(X) = \max_{x \in X} \Pi(x) \tag{3}$$

We also assume a total a priori ignorance on C_{ij} and c_i, i.e.,

$$\Pi(C_{ij}) = 1 \ \forall C_{ij} \text{ and } \Pi(c_i) = 1 \ \forall c_i. \tag{4}$$

Using these relations, the degree of plausibility held by DM (before considering DM's opinions about Director j's opinions) that the actual value of c_i is in $A \subseteq L_{s_i}$ where L_{s_i} is the domain of c_i (equal to L_s in this example), given what the Director j states about criteria i, is given by:

$$pl^{L_{s_i}}[C_{ij}](c_i \in A) = \Pi(c_i \in A|C_{ij}) \quad \text{by (1)}$$

[1] Here we use quantitative possibilities ranging in [0,1], whereas in the QPT approach, only qualitative possibilities are considered. The link between plausibilities and possibilities ($pl = \Pi$) requires that possibilities be somehow quantified.

$$= \max_{x \in A} \Pi(c_i = x | C_{ij}) \quad \text{by (3)}$$

$$= \max_{x \in A} \Pi(C_{ij} | c_i = x) \quad \text{by (2) and (4)}$$

$$= \max_{x \in A, a \in C_{ij}} \Pi([a, a] | c_i = x) \quad \text{by (3)}$$

Table 2 presents the focal elements of the basic belief assignments derived from each informative C_{ij}.

Director	Criteria	C_{ij}	Focus : m	Focus : m
Mkt	Lear	[2,3]	1234 : .5	23 : .5
Mkt	Exp	4	345 : .5	4 : .5
Mkt	Com	4	345 : .5	4 : .5
Mkt	Crea	5	45 : .5	5 : .5
Fin D	Ana	4	345 : .5	4 : .5
Prod D	Ana	2	123 : .5	2 : .5
Prod D	Lea	4	345 : .5	4 : .5
Prod D	Dec	[1,2]	123 : .5	12 : .5
HR D	Lear	[2,4]	12345 : .5	234 : .5
HR D	Com	4	345 : .5	4 : .5
HR D	Dec	3	234 : .5	3 : .5
HR D	Crea	1	12 : .5	1 : .5

Table 2: List of the individual informative C_{ij} and the non zero basic belief masses of $m^{L_{s_i}}[C_{ij}]$ induced on the space $L_s = \{1, 2, 3, 4, 5\}$. The basic belief masses are presented as a pair where the first term is the list of the elements of L_s that belong to the focal element and the second term is the value of the mass itself.

The DM has some weighted opinions about the reliability of the assessment provided by Director j on criteria i, represented by the coefficients γ_{ij} (numerically rescaled into $g_{ij} = 0, 1, 2, 3$ for $\gamma_{ij} = \emptyset, a, b, \mathbb{1}$, respectively). DM also has some prior opinions about the importance that he should give to Director j's opinions when it comes to evaluating a candidate for a position like the open one, opinions that were coded by α_j (numerically rescaled into $s_j = 1, 2, 3$ for $\alpha_j = r, s, \mathbb{1}$, respectively). All these opinions are transformed into discounting factors that express how much belief DM should give to the beliefs induced by the data produced by Director j on criteria i. The discounting factors d_{ij} are decreasing when the α and γ values increase.

Discounting factors are meta-beliefs, i.e., beliefs over beliefs. They were introduced in [Shafer, 1976], and their formal nature explained in [Smets, 1993a].

Their real assessment is done as for any belief. For the purpose of the example, we use the values obtained from the relation:

$$d_{ij} = 1 - .95 \times (g_{ij}/3) \times (.75 + .25 \times (s_j - 1)/2).$$

So $d_{ij} = 1$ in the worst case where $g_{ij} = 0$ and 0.05 in the best case where $g_{ij} = 3$ and $s_j = 3$. The d_{ij} used here are of course arbitrary. In real application, their evaluation would be part of the whole assessment procedure.

Given the coefficients d_{ij}, DM discounts $m^{L_{s_i}}[C_{ij}]$ obtained from $pl^{L_{s_i}}[C_{ij}]$ derived just above into $m^{L_{s_i}}[E_{ij}]$ where:

$$m^{L_{s_i}}[E_{ij}](A) = (1 - d_{ij}) \times m^{L_{s_i}}[C_{ij}](A) \text{ if } A \neq [1, 5],$$

$$m^{L_{s_i}}[E_{ij}]([1, 5]) = (1 - d_{ij}) \times m^{L_{s_i}}[C_{ij}]([1, 5]) + d_{ij}$$

The basic belief assignment $m^{L_{s_i}}[E_{ij}]$, presented in Table 3 represents DM's beliefs about the actual value of c_i given the piece of evidence E_{ij} that is equal to what Director j has stated and DM's opinions about j's opinions (the α's and the γ's).

Director	Criteria	d_{ij}	Focus : m	Focus : m	Focus : m
Mkt	Lear	.367	1234 : .317	23 : .317	12345 : .367
Mkt	Exp	.050	345 : .475	4 : .475	12345 : .050
Mkt	Com	.050	345 : .475	4 : .475	12345 : .050
Mkt	Crea	.050	45 : .475	5 : .475	12345 : .050
Fin D	Ana	.287	345 : 356	4 : .356	12345 : .288
Prod D	Ana	.287	123 : .356	2 : .356	12345 : .288
Prod D	Lea	.525	345 : .238	4 : .238	12345 : .525
Prod D	Dec	.762	123 : .119	12 : .119	12345 : .763
HR D	Lear	.169		234 : .416	12345 : .584
HR D	Com	.169	345 : .416	4 : .416	12345 : .169
HR D	Dec	.723	234 : .139	3 : .139	12345 : .723
HR D	Crea	.169	12 : .416	1 : .416	12345 : .169

Table 3: List of discounting factors d_{ij} for the individual informative C_{ij} and the non zero basic belief masses of $m^{L_{s_i}}[E_{ij}]$ induced on the space $L_s = \{1, 2, 3, 4, 5\}$. The basic belief masses are presented as a pair where the first term is the list of the elements of L_s that belong to the focal element and the second term is the value of the mass itself.

Then DM combines the Directors' opinions over Criteria i by applying the (unnormalized) Dempster's rule of combination to the discounted basic belief assignments. The resulting basic belief assignment $m^{L_{s_i}}[\&_j E_{ij}]$, presented in Table 4, represents DM's beliefs about the actual value of c_i.

$$m^{L_{s_i}}[\&_j E_{ij}] = \oplus_j m^{L_{s_i}}[E_{ij}]$$

Ana	Lear	Exp	Com	Dec	Crea
Focus : m	Focus : m	Focus : m	Focus : m	Focus : m	Focus : m
∅ : 0.381	∅ : 0.075	4 : 0.475	4 : 0.475	∅ : 0.016	∅ : 0.790
2 : 0.102	3 : 0.075	345 : 0.475	345 : 0.475	2 : 0.016	1 : 0.021
3 : 0.127	23 : 0.166	12345 : 0.050	12345 : 0.050	12 : 0.086	12 : 0.021
123 : 0.102	4 : 0.162			3 : 0.122	5 : 0.080
4 : 0.102	34 : 0.111			23 : 0.016	45 : 0.080
345 : 0.102	234 : 0.149			123 : 0.086	12345 : 0.008
12345 : 0.083	1234 : 0.097			234 : 0.106	
	345 : 0.051			12345 : 0.551	
	12345 : 0.112				

Table 4: For each criteria, list of non zero basic belief masses of $m^{L_{s_i}}[\&_j E_{ij}]$ obtained by combining the discounted basic belief assignments obtained from each director. The basic belief masses are presented as a pair where the first term is the list of the elements of L_s that belong to the focal element and the second term is the value of the mass itself.

3.2 DM's beliefs on the real value of the Goodness Score

The real problem for DM is not to assess the actual value of each criterion, but to assess if candidate K is 'good' for the position to be filled. So we introduce a 'Goodness Score', that will vary from 1 to 5, 1 for 'very bad', 5 for 'very good'. The relation between the Goodness Score and the value x_i of criterion i depends only on β_i, the level of importance of criterion i. The values of $f_i(x_i)$ are tabulated in Table 5. E.g., when $\beta_i = g$, DM accepts as 'good' (score 5) a candidate for whom the criterion value x_i is 4 or 5. In practice, these values must be assessed through an analysis of the compensatory relation between the scores. E.g., it is as 'good' to have a score 3 for a criterion which has importance $\mathbb{1}$ than to have a score 3 when the importance is g and to have a score 2 when the importance is e.

Let $f_i(x)$ be the value of the Goodness Score when the score for Criteria i is x and let $f_i(A) = \{f_i(x) : x \in A\}$. Then the belief over the value of the criterion i is transformed into a belief $m^G[C_i]$ over the Goodness Score by:

$$m^G[C_i](B) = \sum_{A: f_i(A)=B} m^{L_{s_i}}[\&_j E_{ij}](A)$$

$$\text{for } B \subseteq [1, 2, 3, 4, 5]$$

β_i	1	2	3	4	5
e	1	3	4	5	5
f	1	2	4	5	5
g	1	2	3	5	5
$\mathbb{1}$	1	2	3	4	5

Table 5: Values of the goodness score $f_i(x)$ given the score x of criteria i (from 1 to 5) and its level of importance β_i

This just means that the basic belief mass $m^{L_{s_i}}[\&_j E_{ij}](A)$ given to A is transferred to the image of A under the transformation f_i that holds between the criterion value and the Goodness Score.

Ana Focus : m	Lear Focus : m	Exp Focus : m	Com Focus : m	Dec Focus : m	Crea Focus : m
$\emptyset$: 0.381	$\emptyset$: 0.075	4 : 0.475	4 : 0.475	$\emptyset$: 0.016	$\emptyset$: 0.790
2 : 0.102	4 : 0.075	345 : 0.475	345 : 0.475	2 : 0.016	1 : 0.021
3 : 0.127	34 : 0.166	12345 : 0.050	12345 : 0.050	12 : 0.086	12 : 0.021
123 : 0.102	5 : 0.162			3 : 0.122	5 : 0.080
4 : 0.102	45 : 0.162			23 : 0.016	45 : 0.080
345 : 0.102	345 : 0.149			123 : 0.086	12345 : 0.008
12345 : 0.083	12345 : 0.210			234 : 0.106	
				12345 : 0.551	

Table 6: For each criteria, list of non zero basic belief masses of $m^G[C_i]$ obtained on the Goodness Score using the appropriate f_i functions. The basic belief masses are presented as a pair where the first term is the list of the elements of L_s that belong to the focal element and the second term is the value of the mass itself.

The basic belief assignment $m^G[C_i]$ derived from the basic belief assignment $m^G[C_i]$ represents DM's beliefs about the value of the Goodness Score of the candidate given what DM collected about the criteria i. These basic belief assignments are then combined on i by Dempster's rule of combination.

$$m^G[\&_i C_i] = \oplus_i m^G[C_i]$$

The resulting basic belief assignment $m^G[\&_i C_i]$ represents DM's beliefs over the value of the Goodness Score of candidate K given the collected information and all DM's *a priori* about the various criteria importance and the directors competence.

3.3 Selecting a candidate

When it comes to decide which candidate to select, this last belief function is then transformed into a (pignistic) probability, denoted $BetP^G$, by the pignistic transformation (the meaning and the justification of this transformation are detailed in [Smets and Kennes, 1994]). We have:

$$BetP^G(A) = \sum_{B \subseteq [1,2,3,4,5]} \frac{|A \cap B|}{|B|} \frac{m^G[\&_i C_i](B)}{1 - m^G[\&_i C_i](\emptyset)}$$

This probability function $BetP^G$ over the actual value of the Goodness Score can then be used to order various candidates, using the classical methods developed in probability theory for such an ordering.

For the example under analysis, the non zero basic belief masses of $m^G[\&_i C_i]$ and of the pignistic probabilities $BetP^G$ are given in Table 7. In conclusion, there is a strong support that this candidate is Good (score 5). An expected Goodness Score can be computed, which value here is 4.61, and it could be used to compare K with other candidates.

focal sets	$\emptyset$	4	5	{4,5}
$m^G[\&_i C_i]$	.9858	.00276	.00988	.001537
$m^G[\&_i C_i] Normalized$	0	.34	.58	.08
$BetP^G$	0	.38	.62	

Table 7: Focal sets of $m^G[\&_i C_i]$ and their masses on the Goodness Score, and non zero values of $BetP^G$ on the singletons of G.

3.4 Contradiction analysis

The TBM offers the possibility to judge the magnitude of the conflict between the belief functions that enter into a combination. This is achieved by the basic belief mass $m(\emptyset)$ given to the empty set, the largest $m(\emptyset)$ the largest the conflict. The value of the conflict observed when combining the directors' opinions on a criteria are listed in Table 8.

Ana	Lear	Exp	Com	Dec	Crea
.38	.08	.00	.00	.02	.79

Table 8: For each criteria, value of the conflict between the basic belief assignments combined to derive $m^{L_{s_i}}[\&_j E_{ij}]$. They are equal to $m^{L_{s_i}}[\&_j E_{ij}](\emptyset)$ listed in table 4

There is no conflict among the directors when it comes to evaluate the criteria Exp, what is obvious as there is only one expressed opinion, and Com where both opinions are equal. There is a small conflict for Dec and Lear. A large conflict appears for what concern Ana, and still a larger one for Crea where the two directors were fully conflicting.

The overall conflict computed when combining the basic belief assignments on the Goodness Score is .9858 (see Table 7). This conflict should convince the DM that the opinions expressed by the directors are strongly conflicting and any conclusion should be taken very cautiously as probably something went wrong somewhere (as it just happens to be the case when looking to the data for the Ana and Crea criteria).

3.5 Sensitivity analysis

The TBM allows also to perform a sensitivity analysis. We can consider what would be the final pignistic probabilities and the mean Goodness Scores if we had obtained more precise assessments for each director and each criteria. Table 9 lists the mean Goodness Scores one would have obtained if the various imprecise assessments had been precise, one by one. The considered values are consistent with the collected intervals C_{ij}. The largest difference, hence the largest sensitivity, is observed for the data collected from the HR Director for the Lear criteria. So this is the first criteria that deserves to be assessed more precisely.

It is also possible to determine which data should be reconsidered in order to reduce the conflicts. In the present example, it is obvious (see Table 8) that the conflicts for Ana and Crea should be first settled. This would mean asking the directors to reconsider their evaluations. This sensitivity analysis is not further explored here as the problem is not at the level of the experts' opinions pooling, but at the experts' opinions themselves.

3.6 Comparing several candidates

We present the data collected for four candidates in Tables 10 to 13, the first being the one studied so far. We tabulate the collected data, the pignistic probabilities, the mean Goodness Scores, the overall conflict, and the 'crude mean' one would obtain by just adding the values of the mid point of each C_{ij} interval without taking in consideration any weight.

The mean Goodness Score is the highest for the first candidate, but the amount of overall conflict is so high that the next best candidate (the fourth) should also be considered (mean Goodness Score = 4.50). It is obvious that the collected data are not only too imprecise but also too much conflicting.

For illustrative purpose, we present the Crude mean scores. The fourth candidate would then be classified as the best candidate, and the first one ends up in a bad position. Which is the best decision can only be assessed by the expert who should considered the collected data and produce his/her own classification.

Criteria	Director	Opinion	Mean Score	Difference
Lear	Mkt	2	4.55	.07
Lear	Mkt	3	4.62	.00
Lear	Fin	1	4.61	.00
Lear	Fin	2	4.57	.04
Lear	Fin	3	4.58	.03
Lear	Fin	4	4.65	.03
Lear	Fin	5	4.66	.04
Lear	HR	2	**4.29**	**.32**
Lear	HR	3	4.47	.14
Lear	HR	4	4.74	.13
Dec	Mkt	1	4.61	.00
Dec	Mkt	2	4.61	.00
Dec	Mkt	3	4.61	.00
Dec	Mkt	4	4.66	.04
Dec	Mkt	5	4.70	.08
Dec	Fin	1	4.61	.00
Dec	Fin	2	4.61	.00
Dec	Fin	3	4.61	.00
Dec	Fin	4	4.65	.03
Dec	Fin	5	4.68	.06
Dec	Prd	1	4.61	.00
Dec	Prd	2	4.61	.00

Table 9: For each ij, value of the mean Goodness Scores and their difference with the observed score when the imprecise C_{ij}'s take one of the compatible precise values. The Score are recomputed while keeping all C_{ij}'s unchanged except the one considered.

4 The QPT approach

The problem stated in Section 2 raises three main questions: i) the representation of the precise or imprecise score of the candidate provided by each expert for each criterion, including the expert's confidence in his assessment; ii) the fusion of expert opinions; iii) the multiple criteria aggregation. The first step is easily handled in qualitative possibility theory [Dubois and Prade, 1998a] which offers a representation framework for handling imprecise values pervaded with qualitative uncertainty. This might be related to the processing of fuzzy marks for students' evaluation for which an (ad hoc) treatment was proposed recently [Biswas, 1995]. An alternative would have been to map the ordinal scale on a suitable cardinal scale by a method such as "Macbeth" [Bana and Vansnick]. However this approach requires in fact additional information.

In the following, $\vee$ and $\wedge$ denote max and min on a given ordinal scale. $\neg x$ for any x in a given ordinal scale $L = \{\emptyset, s_1, \ldots, s_k, \mathbb{1}\}$ denotes the value

C_{ij}	Mkt D	Fin D	Prod D	HR D	G.Sc.	BetP		
Ana		4	2		1	.00	G. Mean	4.61
Lear	[2,3]	[1,5]	4	[2,4]	2	.00	Confl.	.99
Exp	4				3	.00	Crude M.	3.08
Com	4			4	4	.38		
Dec	[1,5]	[1,5]	[1,2]	3	5	.62		
Crea	5			1				

Table 10: Assessment of a candidate by each director on each criteria. On the right, BetP are the values of the pignistic probabilities given to the Goodness Score (G. Sc.), G. Mean is the mean of the Goodness Score, Confl. is the value of the overall conflict, and Crude M. is the" crude mean.

C_{ij}	Mkt D	Fin D	Prod D	HR D	G.Sc.	BetP		
Ana		[2,4]	[1,4]		1	.00	G. Mean	3.98
Lear	[2,4]	[1,3]	[1,4]	2	2	.00	Confl.	.91
Exp	[4,5]				3	.06	Crude M.	3.13
Com	[4,5]			[4,5]	4	.89		
Dec	[1,3]	[1,2]	4	[4,5]	5	.04		
Crea	3			[4,5]				

Table 11: Assessment of a candidate by each director on each criteria. See legend of 10.

corresponding to the reversed scale (i.e. $\neg\emptyset = \mathbb{1}, \neg s_i = s_{k-i+1}$). It is the counterpart to $1 - x$ on the $[0, 1]$ interval scale. To simplify the notation we have denoted the top and bottom elements of each scale by the same symbols; however this does not mean that they are the same.

4.1 Representing imprecise and uncertain scores

The evaluation of each criterion for candidate K and a given expert j may be imprecise (either due to the fact that it is unclear to what precise extent K satisfies criterion i or due to the possible lack of competence in i of the expert j assessing the value). Each evaluation will be represented by a possibility distribution (discounted in case of limited expertise) restricting the more or less possible values of this evaluation. Let $\pi^j_{c_i(K)}$ (π^j_i for short) denote the possibility distribution restricting the possible values of $c_i(K)$ according to expert j.

The possibility distribution (π.d.f. for short) of the true value of each score on criterion i according to expert j, taking into account his competence, is computed in the following way. Interval-valued scores (including single values)

C_{ij}	Mkt D	Fin D	Prod D	HR D	G.Sc.	BetP		
Ana		4	[2,3]		1	.00	G. Mean	3.44
Lear	[2,3]	[2,4]	5	[2,4]	2	.00	Confl.	.82
Exp	[3,4]				3	.61	Crude M.	3.08
Com	[2,4]			[1,4]	4	.33		
Dec	[3,5]	1	3	[2,4]	5	.05		
Crea	[3,5]			3				

Table 12: Assessment of a candidate by each director on each criteria. See legend of 10.

C_{ij}	Mkt D	Fin D	Prod D	HR D	G.Sc.	BetP		
Ana		[4,5]	[1,5]		1	.00	G. Mean	4.50
Lear	[1,3]	[4,5]	4	[2,3]	2	.00	Confl.	.75
Exp	[4,5]				3	.03	Crude M.	3.21
Com	[3,5]			[2,4]	4	.44		
Dec	[1,3]	[4,5]	3	[2,3]	5	.53		
Crea	[1,5]			[1,5]				

Table 13: Assessment of a candidate by each director on each criteria. See legend of 10.

are modeled by a possibility distribution taking the value $\mathbb{1}$ in the interval and $\emptyset$ outside. Blanks (absence of answers) are interpreted as a possibility distribution being $\mathbb{1}$ everywhere (modeling "unknown"). The confidence level γ_{ij} is taken into account by a discounting process, defined as ($\tilde{\pi}$ denotes the original possibility distribution function)

$$\pi_i^j(s) = \tilde{\pi}_i^j(s) \vee \neg\gamma_{ij}, \quad \forall s \in L_s \tag{5}$$

Note that the certainty level $\gamma_{ij} \in L_\gamma$ is turned into a possibility level $\neg\gamma_{ij}$ over the scores not compatible with $\tilde{\pi}_i^j$. Thus L_γ is L_π reversed (this is the usual equivalence between certainty of A and impossibility of not A).

Moreover, as in the TBM approach (3.1), it is admissible here to fuzzify the measurements of each director, *because* the score scale {1,2,3,4,5} may be thought as a discretized continuum, which means, e.g., that 3 is close to 4 in some sense, as when a director says '4' we cannot fully exclude neither 3 nor 5. We use the same fuzzification as in section 3.1, except that the values immediately close to the assessments receive the possibility 'a' (instead of .5 as in section 3.1, values which are further away remain with a zero possibility. This fuzzification takes place before applying (5). Thus $\tilde{\pi}_i^j$ should be the fuzzified original possibility distribution in expression (5). This fuzzification and the discounting lead to Table 14. In each cell the π.d.f is enumerated on L_s.

	Mkt D	Fin D	Prod D	HR D
Ana	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\emptyset\emptyset a\mathbb{1}a$	$a\mathbb{1}a\emptyset\emptyset$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$
Lear	$a\mathbb{1}\mathbb{1}aa$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$aaa\mathbb{1}a$	$a\mathbb{1}\mathbb{1}\mathbb{1}a$
Exp	$\emptyset\emptyset a\mathbb{1}a0$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$
Com	$\emptyset\emptyset a\mathbb{1}a$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\emptyset\emptyset a\mathbb{1}a0$
Dec	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}bbb$	$bb\mathbb{1}bb$
Crea	$\emptyset\emptyset\emptyset a\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$	$\mathbb{1}a\emptyset\emptyset\emptyset$

Table 14: Opinions of directors on each criterion after fuzzification and discounting.

4.2 Merging expert opinions

As already said, the global assessment requires two types of combination: a *multiple criteria aggregation* problem, and the *fusion* of expert evaluations. So, depending on the way the problem is presented, we may either think of i) first computing the global evaluation of K according to each expert and then to fuse these evaluations into a unique one, or ii) on the contrary, first fuse the expert evaluations for each criterion, and then aggregate the "global" results pertaining to each criterion. In general, the two procedures are not equivalent. (i.e., expert opinion fusion and multiple criteria aggregation do not commute, and the same holds for the TBM analysis). So it is important to understand what is the meaningful order between the fusion and aggregation operations, or if this remains unclear, to choose fusion and aggregation modes which commute.

At this point, it is worth emphasizing that expert opinion fusion and multiple criteria aggregation are two operations which do not convey the same intended semantics. The fusion of expert opinions aims at finding out what the possible values of the genuine score of K for a given criterion are, and possibly to detect conflicts between experts. Hopefully, some consensus should be reached at least on values which are excluded as possible values of the score. The aggregation of multiple criteria evaluations aims at assessing the global worth of the candidate from his scores on the different criteria; then different aggregation attitudes may be considered, e.g., conjunctive ones where each criterion is viewed as a constraint to satisfy to some extent, or compensatory ones where trade-offs are allowed.

In the following, we choose to merge expert's opinions on each criterion first, and then to perform a multiple criteria aggregation, since it might seem more natural to use the experts first to properly assess the score according to each criterion. Proceeding in the other way would assume that each expert is looking for a global evaluation of the candidate (may be using his own criteria aggregation attitude) and the decision maker is only there for combining and weighting expert's evaluations.

Let us first merge the distributions pertaining to a single criterion. When there is no major conflict between the assessments to be merged, a conjunctive combination can be performed which singles out values of common agreement. In QPT, possibility distributions are then combined by the min operation (here denoted $\wedge$), after having been discounted if necessary. When there is a strong conflict, i.e., here when

$$\text{for some } i, \text{conflict}_i = \vee_s \wedge_j \pi_i^j(s) = 0, \tag{6}$$

disjunctive combination is advisable in such a case[2] [Dubois and Prade, 1994a]. Conjunctive fusion can be applied only if there is no conflict in the above sense. A disjunctive combination means that the opinion of an important expert will not be forgotten, even if it conflicts with another important one. See [Dubois and Prade, 1994a] for a general introduction to the logical view of information fusion and its encoding in the possibilistic framework by weighted conjunctions and disjunctions.

However, unreliable estimates should be discounted both in the conjunctive and in the disjunctive merging. The reliability w_{ij} attached to π_i^j should be both upper bounded by the confidence α_j of DM in the expert j and the self-confidence γ_{ij} of the expert. This leads to take the weight w_{ij} as the conjunctive-like combination of α_j and γ_{ij}, $\forall i = 1, \dots, 6$.

The weighted conjunction, applied to i without conflict, will be :

$$\pi_i'(s) = \bigwedge_j [(\neg w_{ij} \vee \pi_i^j(s)], \ \forall s \in L_s. \tag{7}$$

while the weighted disjunction is defined as:

$$\pi_i'(s) = \bigvee_j \left[w_{ij} \wedge \pi_i^j(s) \right], \quad \forall s \in L_s, \tag{8}$$

As said above, $w_{ij} = \gamma_{ij} \otimes \alpha_j$ where $\otimes$ is a conjunction operator from $L_\gamma \times L_\alpha$ to L_γ. Table 15 defines $\otimes$ on the basis of an implicit commensurateness hypothesis of L_γ and L_α. Due to the idempotency of $\wedge$ and $\vee$, the apparently redundant treatment of the information encoded by the γ_{ij} which are taken into account both in (5), and in (7) - (8) for building the w_{ij} is innocuous. Table 16 gives the weights for all criteria and experts.

In fact, it is not clear if the weights are absolute or relative. Here we have assumed they are absolute. In case they are relative, we could use a "nonmonotonic" conjunction or disjunction for discounting the part of the information provided by less important experts which is in conflict with what is provided by the more important ones; see [Dubois and Prade, 1994a] for details.

[2] Here we assume that a conflict occurs when the intersection of π.d.f.'s is empty. Note that since the underlying space is ordered, and a fuzzification has been performed at the previous step, conflicts are less often present (however one still occurs for criteria $i = 6$ (Crea)).

$\otimes$	$\emptyset$	u	v	$\mathbb{1}$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
a	$\emptyset$	a	a	a
b	$\emptyset$	a	b	b
$\mathbb{1}$	$\emptyset$	a	b	$\mathbb{1}$

Table 15: Definition of $\otimes$

w_{ij}	Mkt D	Fin D	Prod D	HR D
Ana	$\emptyset$	a	a	$\emptyset$
Lear	b	a	a	b
Exp	$\mathbb{1}$	$\emptyset$	$\emptyset$	$\emptyset$
Com	$\mathbb{1}$	$\emptyset$	$\emptyset$	b
Dec	a	a	a	a
Crea	$\mathbb{1}$	$\emptyset$	$\emptyset$	b

Table 16: Weights w_{ij} for all criteria and directors

We apply formula (7) to π.d.f's of Table 14 (or (8) if (6) holds). We obtain the π.d.f's of Table 17 (left). Moreover, in case of disjunctive combination, translating the lack of answer of a director (here due to a total lack of competence) by the π.d.f. expressing total ignorance (possibility = 1 on each score) is not innocuous. Indeed in case of disjunctive combination, asserting the whole set $\{1,...5\}$ amounts to state that the expert considers for some reason that the score can take any value in the set and that all of them are plausible: namely the candidate is capable of the best as well as the worst w.r.t. the criterion. This is not the same as being fully ignorant about the candidate. So in (8), the scope of $\vee$ should be limited to the j's for which we have an answer.

We notice that some distributions are no longer normalized (i.e., no score is fully possible at level $\mathbb{1}$). In the case of the disjunctive combination, this is only because the weights are not normalized, i.e. $\vee_j w_{ij} < \mathbb{1}$ for some i's, which means that the DM cannot be fully confident in any director for assessing some criteria (namely Ana, Lear and Dec). This is a little paradoxical, since it would be natural that the DM employs at least one fully reliable expert per criterion! In the case of the conjunctive combination, resulting distributions are unnormalized in case of partial conflicts. Thus, we should either use the unnormalized distributions to fully keep track of the problem, or normalize them in a suitable way. Here, we choose the second solution, and propose the following approach.[3] We consider that the maximum of the distribution,

[3]Normalization has some advantage for the next step of the procedure. Due to the use

Ana	$aaaaa$
Lear	$abbaa$
Exp	$\emptyset\emptyset a \mathbb{1} a$
Com	$\emptyset\emptyset a \mathbb{1} a$
Dec	$bbbbb$
Crea	$ba\emptyset a \mathbb{1}$

Ana	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$
Lear	$b\mathbb{1}\mathbb{1}bb$
Exp	$\emptyset\emptyset a \mathbb{1} a$
Com	$\emptyset\emptyset a \mathbb{1} a$
Dec	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}$
Crea	$ba\emptyset a \mathbb{1}$

Table 17: Merged opinions of directors on criteria (left); Merged opinions after normalization (right)

denoted h, reflects the uncertainty level, considered to be $\neg h$ in the information. This means that the amount of conflict is changed into a level of uncertainty and the minimum level of the modified distributions will be $\neg h$. Then, we make an additional hypothesis that the scale is an interval scale (which is questionable!) so that the profile of the distribution is not changed. Specifically:

$$\pi_i(s) = \pi_i'(s) + \neg(\vee_s \pi_i'(s)) \tag{9}$$

It means that $+$ in (9) is defined by $s_i + s_j = s_{\min(k+1,i+j)}$ on a scale $\{s_0 = \emptyset, s_1, \ldots, s_k, s_{k+1} = \mathbb{1}\}$. The result is shown in Table 17. A slightly different method would be to keep possibility degrees equal to $\emptyset$ at $\emptyset$, since $\pi_i'(s) = \emptyset$ means that all experts agree on the fact that the value s is impossible.

4.3 Criteria aggregation

The way of aggregating the criteria evaluations is not at all specified in the statement of the problem. The aggregation of preferences expressed by the criteria and their levels of importance is not viewed here as another multiple source information fusion problem as in Section 3 where the "Goodness" of the candidate was estimated by the conjunctive combination of the basic belief assignments modeling the Goodness of the candidate according to each criterion. In the following the aggregation of criteria is viewed as a problem different from data fusion, where we are not trying to estimate the true value of a parameter, but rather to express how the levels of satisfaction of each criterion contribute to the global level of satisfaction, just allowing for trade-offs. However the aggregation function (which is not at all part of the QPT model) is almost unspecified here. Moreover, the evaluations attached to each criterion are not pointwise here, but rather imprecise and pervaded with uncertainty.

Only qualitative levels of importance β_i are provided for each criterion i. Even with ordinal scales, different attitudes can be thought of. The aggregation

of the extension principle in the final step (see (12)), we need to normalize, otherwise the resulting π.d.f. will be truncated by the height of the smallest distribution, and thus would continue to keep track of the lack of reliability of a part of the information, but would have its profile modified, since the nuances between more or less high degrees of possibility are lost.

may be purely conjunctive (based on weighted "min" operation), or somewhat compensatory (using a median operation for instance). It might be also disjunctive if at least one important criterion has to be satisfied, then it is modeled by a weighted maximum. More general aggregation attitudes can be captured by Sugeno's integral [Sugeno, 1977];[Grabisch et *al.*, 1995]. Let us assume that the DM has a somehow compensatory attitude.

Here we use the only associative qualitative compensatory aggregation operator, namely the median. Indeed, $\min(x, y) \leq \text{median}(x, y, \alpha) \leq \max(x, y)$ for any α belonging to the domain of x and y. Thus we shall take α to be the middle point of the scale, i.e., 3, since it will be applied to $L_s = \{1, 2, 3, 4, 5\}$; Due to its associativity, the median operation can be easily generalized to the aggregation of n terms, namely $\text{median}(x_1, x_2, \alpha)$ becomes $\text{median}(\wedge_k x_k, \vee_k x_k, \alpha)$. However we have still to take into account the importance levels of the criterion. So the global score s after aggregation will be taken to be equal to $\text{median}(c, d, \alpha)$ where c (resp. d) is a weighted conjunction (resp. disjunction) computed as follows.

With our notations, in the case of precise scores, the global score combination with the weighted minimum method is obtained by:

$$c = \bigwedge_{i=1}^{6} [(\neg\beta_i)\tilde{\vee}c_i] \tag{10}$$

where $\tilde{\vee}$ is a disjunctive operator from $L_\beta \times L_s$ to L_s. Table 18, left part, gives the definition of this operator.

Computed with the weighted maximum, the global score would be:

$$d = \bigvee_{i=1}^{6} [(\beta_i)\tilde{\wedge}c_i] \tag{11}$$

where $\tilde{\wedge}$ is given in right part of Table 18.

$\tilde{\vee}$	$\emptyset$	e	f	g	$\mathbb{1}$
1	1	2	3	4	5
2	2	2	3	4	5
3	3	3	3	4	5
4	4	4	4	4	5
5	5	5	5	5	5

$\tilde{\wedge}$	$\emptyset$	e	f	g	$\mathbb{1}$
1	1	1	1	1	1
2	1	2	2	2	2
3	1	2	3	3	3
4	1	2	3	4	4
5	1	2	3	4	5

Table 18: Definition of $\tilde{\vee}$ and $\tilde{\wedge}$

The "fuzzy" evaluations provided by the π.d.f.'s should now be aggregated. A natural manner to proceed is to extend multiple criteria aggregation techniques to such non-scalar evaluations (this can be done both if this step is done before or after the fusion step). At the technical level, this is done using an extension principle [Zadeh, 1975] which makes it possible to extend any function/operation f to any fuzzy arguments. Namely,

$$f(\pi_1, \dots, \pi_m)(s) = \vee_{s=f(s^1,\dots,s^m)} \pi_1(s^1) \wedge \dots \wedge \pi_m(s^m)$$

where $\pi_i(s^i)$ is the possibility degree of score s^i according to π.d.f. π_i.

Here f is the aggregation function $f(c_1, ..., c_6, \beta_1, ..., \beta_6) = \text{median}(c, d, 3)$. We have for any $s \in L_s$ and candidate K,

$$\pi_K(s) = \bigvee_{s=\text{median}(3, \bigwedge_{i=1}^{6}[(\neg\beta_i)\tilde{\vee}c_i], \bigvee_{i=1}^{6}[\beta_i \tilde{\wedge} c_i]} (\pi_1(c_1) \wedge \dots \wedge \pi_6(c_6)) \tag{12}$$

Applying this formula to our data, we finally obtain the distribution given in Table 19.

score	1	2	3	4	5
possibility degree	$\emptyset$	$\emptyset$	$\mathbb{1}$	$\mathbb{1}$	$\emptyset$

Table 19: Possibility distribution π_K of the final score

The result can be interpreted by saying that the candidate K is certainly not a very good candidate nor a bad or average one, and K is between medium and good. Let us briefly comment the result. The most important criteria are Com and Crea, where the scores are respectively 4, and 1 or 5 (with possibility $\mathbb{1}$). This explains why the global value 5 is impossible, since it is not reached by one important criterion. Moreover the score on criterion Dec which is rather important, is around 3, while the other important criterion Ana has a score from 2 to 4. This explains why only intermediate values 3 and 4 are possible for the global score, with a median-based , i.e ., a compensatory operation, at the multiple criteria aggregation step. In this peculiar example the π.d.f.'s which is obtained for the score is {0,1}-valued. In Table 20, where other candidates are considered, π.d.f.'s with intermediate possibility degrees are obtained as well.

Results for the three other candidates considered in Section 3.6 are displayed in Table 20. The result obtained above is a possibility distribution π_K restricting the possible values of the global evaluation of K. Then a fuzzy ranking method should be used for assessing to what extent it is certain, and to what extent it is possible that K_1 has a score larger than K_2, on the basis of π_{K_1} and π_{K_2} in case of several candidates [Dubois and Prade, 1983].

Remark. Note that if weighted min combinations would be used both in the expert opinion fusion and in the multi-criteria aggregation then the two combinations commute, at least with precise scores. In the general case the multiple criteria aggregation and the expert opinion fusion are not expected to commute, as suggested by the simple following example. Let us suppose we have two criteria and two experts (c_{ij} is the grade for the criterion i according to expert j):

$$c_{11} = 1; \;\; c_{12} = 5.$$
$$c_{21} = 5; \;\; c_{22} = 1.$$

If we first aggregate the criteria by some median operation, we may find 3 for each expert, since according to each expert the candidate is very good on one criterion and very bad on the other. Then fusing the global expert opinions we find 3. If we rather start by fusing the expert opinions, we find "1 or 5" in both cases, since they are conflicting. Note that the result will be the same at this step for the following very different data set:

$$c_{11} = 1; \;\; c_{12} = 5.$$
$$c_{21} = 1; \;\; c_{22} = 5.$$

We have just forgotten, when we will apply some extended aggregation operation to "1 or 5" and "1 or 5" that the pairs (1,1) and (5, 5) are impossible (with the first set of data).

4.4 Qualitative expected values

.

The assessment problem can be viewed as a *decision* to be made *under uncertainty*. Namely, the relevance of the criteria defines a candidate profile, i.e., a kind of utility function, while the value of the candidate K is ill-known with respect to each criterion (once expert opinions have been fused). Viewing the problem in this way, it is natural to compute to what extent it is certain (or it is possible) that the candidate K (whose level assessment may be pervaded with imprecision and uncertainty for each criteria) satisfies the criteria at the required level; see [Dubois and Prade, 1995] for an axiomatic view of the corresponding qualitative decision procedure. It corresponds to a fuzzy pattern matching problem [Dubois et *al.*, 1988] in practice.

For each criterion i, we build a satisfaction profile μ_i, e.g., $\mu_i(s) = \neg\beta_i \vee s$. μ_i means that the greater the score, the better the candidate, and that the satisfaction degree is lower bounded by $\neg\beta_i$ which is all the greater as i is less important. Then the certainty that K satisfies the profile is given by

$$\wedge_i \wedge_s \left(\mu_i(s) \vee \neg\pi_{c_i(K)}(s)\right) \tag{13}$$

where $\pi_{c_i(K)}$ is supposed to be normalized. The possibility that K satisfies the profile is

$$\wedge_i \vee_s \left(\mu_i(s) \wedge \pi_{c_i(K)}(s)\right) \tag{14}$$

where $\pi_{c_i(K)}$ has been obtained by fusing the expert opinions first, at the level of each criterion. The possibility degree (very optimistic) should be only used for breaking ties in case of equality of the certainty degrees for different candidates.

The aggregation by $\wedge_i$ of the elementary certainty and possibility degrees can be justified in the possibility framework (definition of a join possibility distribution of non-interactive variables [Zadeh, 1975], *when* the global requirement is interpreted in terms of a weighted conjunction of elementary requirements pertaining to each criterion). The expression (13) and (14) can be viewed as possibilistic 'lower and upper expectations'.

When a median-based aggregation is used, the possibility and the necessity measures are no longer $\wedge_i$-decomposable as in (13) and (14) and we would have to compute directly the possibility and the necessity that the global satisfaction profile (again computed from the μ_i by application of the extension principle) is satisfied giving the joint π.d.f. $\wedge_i \pi_{C_i(K)}$.

What is computed remains in the spirit of the approach detailed before. Instead of obtaining a possibility distribution we obtain two scalar evaluations for which it should be possible to show that they summarize this possibility distribution (in a sense to be made precise at the theoretical level). Anyway both approaches first compute the $\pi_{c_i(K)}$'s and are based on the choice of an aggregation function.

Note also that we are not obliged to combine the different evaluations $\pi_{c_i(K)}$ pertaining to each criterion (once we have fused the expert grades). We may summarize each possibility distribution by a "lower expected value", and then compare lexicographically tuples made of the scalar evaluations thus obtained for each criterion (the scalar evaluations being ordered in the tuple according to the importance of the criteria), for different candidates.

5 General discussion

The use of the two approaches for dealing with the same (class of) problems has raised two types of issues which are now briefly considered; namely first a comparison of the approaches, and second how each approach could be validated.

5.1 Outline of a comparison

The reader may observe some agreement between the results obtained by the belief function approach and by the possibilistic approach, when considering the possibility distributions and the mass functions which are obtained, before computing the expected values (see Table 20). This is not too surprising if we consider the two flowcharts summarizing the two approaches, which are rather similar (see Tables 22 and 23). However the remaining discrepancies between results are largely due to the different views which are chosen at the multiple criteria aggregation step.

In QPT-based approach, merging opinions and aggregating criteria are envisaged as two different problems. Merging opinions is performed by min-based combination of the π.d.f.'s, as prescribed by QPT, when there is no conflict. When a conflict exists, it means that at least one of the opinions is totally wrong,

K	G. Sc.	1	2	3	3-4	4	4-5	5	Conclusions	Mean G.
1	QPT	$\emptyset$	$\emptyset$	$\mathbb{1}$		$\mathbb{1}$		$\emptyset$	best	
	TBM					.33	.99	.57	4 or 5	4.61
2	QPT	$\emptyset$	$\emptyset$	$\mathbb{1}$		b		$\emptyset$		
	TBM			.05	.94	.87			4, maybe 3	3.98
3	QPT	$\emptyset$	b	$\mathbb{1}$		$\mathbb{1}$		$\emptyset$		
	TBM			.49	.93	.21			3 or 4	3.44
4	QPT	a	b	$\mathbb{1}$		$\mathbb{1}$		$\emptyset$		
	TBM					.37	.95	.47	4 or 5	4.50

Table 20: Results for the 4 candidates $K = 1, 2, 3, 4$ which data are presented in Tables 10 to 13. Goodness Scores from 1 to 5, with the QPT outcomes graded $0, a, b, 1$, and the value of bel^G for the intervals really supported. Score 4-5 means the score is 4 or 5. Presented conclusions are those derived from 'common sense' given the QPT or the TBM outcomes. For QPT, case 1 > case 3 > case 4 and case 1 > case 2, but case 2 and cases 3 or 4 are not comparable. Last column is the mean Goodness Score computed in the TBM.

and a max-based disjunction is performed in order to save the provided information. Then the weighting of the conjunction or of the disjunction amounts to a preliminary discounting or truncation of the π.d.f.'s according to the confidence in the sources. Aggregating criteria supposes to know if trade-offs exist or not. There exists a large panoply of different possible aggregation attitudes, even when dealing with qualitative scales. This contrasts with the view used in the TBM-based approach where only Dempster rule of combination, which is conjunctive, is used here. However belief functions viewed as set functions could have been used for describing multiple criteria aggregations based on Choquet integrals, where *sets* of criteria can be weighted [Grabisch and Roubens, 2000]. In case other aggregation operations would be chosen at the multiple criteria aggregation step, with the QPT approach, results which are rather different could be obtained. For instance, choosing a weighted min-combination at the aggregation step, would lead to results where low scores would not be excluded (see Table 21 for the results which would be obtained for the first candidate).

score	1	2	3	4	5
possibility degree	b	$\mathbb{1}$	$\mathbb{1}$	$\mathbb{1}$	0

Table 21: Possibility distribution π of the final score

The other discrepancies between the results obtained in Sections 3 and 4 with the two approaches may have several reasons.

- The QPT approach using a qualitative possibility scale cannot use any direct counterpart of the product used in the TBM approach. Moreover using the possibility theory approach with a [0,1] scale will open the door to the use of adaptative merging operators [Dubois and Prade, 1994b] which provide a softer adaptation between the conjunctive and the disjunctive attitudes (using the degree of conflict as a weighting factor).

- Moreover, when introducing the Goodness Score in the belief function approach, the underlying idea is that we are all the less demanding for reaching the maximum level of satisfaction as the criterion is less important. This is not the understanding chosen in the possibilistic approach (remember that the problem is not completely specified). Rather, the unimportant criteria were considered as somewhat satisfied even if they were not at all satisfied. In fact there are different possible ways of understanding the weighting of the criteria, in a qualitative conjunctive setting. Let $c_i(K)$ be the supposedly precise value of score of K, according to criterion i. Indeed, we can i) modify $c_i(K)$ into $\neg\beta_i \tilde{\vee} c_i(K)$ as we did, or ii) modify $c_i(K)$ into $\begin{cases} 5 & \text{if} \quad c_i(K) \geq \beta_i \\ c_i(K) & \text{if} \quad c_i(K) < \beta_i \end{cases}$ (β_i is interpreted as a threshold level to reach in order to be fully satisfied, taking advantage of the commensurateness hypothesis underlying Table 18), iii) or we may even think of $\neg\beta_i$ as a "bonus" to be added to $c_i(K)$. Once the type of weighted conjunction has been chosen, it has to be extended to the π.d.f.'s π_i, since the precise value of $c_i(K)$ is not available. Choosing the option (ii) above in the QPT approach instead of (i) would be similar to the idea of Goodness Score as defined in the TBM approach.

- The value assessment problem considered here is probably not the best type of examples for exhibiting the benefit of one of the specificities of the TBM approach, namely the capability of putting masses on subsets of values which are not singletons. More generally, this capability might be of interest both at the merging step and at the aggregation step (then leading to a Choquet integral-based evaluation as already mentioned).

The example also raises the issue of the difference between a value asserted by the expert as "unknown" (interval [1, 5] in the example), and the absence of answer by the expert for a criterion evaluation. Even if in both cases, the result is that the evaluation is actually unknown, the reason of the absence of answer may be either that the director feels himself totally incompetent and does not answer (as already discussed), or that the criterion does not really apply to Mr. K (according to the expert). This will call for a richer evaluation framework where the evaluation assessment can take its value in an extended domain $\{1, 2, 3, 4, 5\} \cup \{\texttt{does\_not\_apply}\}$. Then it creates further difficulties, especially when comparing candidates...

For summarizing the main differences between the belief function (TBM) and the possibility-based approaches (QPT) on the value assessment problem, it is clear that the possibilistic method can handle poor data expressed in a qualitative, non-numerical, way whereas the belief function framework may more easily capture reinforcements and compensatory effects. In the QPT approach, the propagation of imprecision and uncertainty in the combination process has been emphasized, while the TBM approach has privileged the decision step by computing expected values.

5.2 Brief discussion on the validation issues

The validation of an approach used in practice involves several aspects according to which an approach may be judged and compared to others.

A first type of validation is the practical one (not always easy to perform). It consists in checking the 'correctness' of the results provided by the system by comparing them to those produced by a panel of experts (w.r.t. to the information actually provided to the system). One must nevertheless be careful that the experts might not provide a golden standard. This approach produces a measure of concordance between two systems, not always a measure of quality. Furthermore a system may provide correct results without being considered as genuinely 'useful' by the experts. Anyway this type of validation is usually done on systems which are already developed at an operational level. The real pragmatic validation should be carried on as it is done in medicine, through a clinical trial where several methods are compared on many real examples and their values assessed by comparing the final results.

But other aspects should be also considered before. In the following we distinguish between those pertaining to normative, empirical, computational and explanatory issues.

From a normative validation point of view, we can see our problem as made of two main steps : i) the combination of pieces of information coming from different experts, ii) then a decision step made on the basis of the imprecise/uncertain information obtained at the end of step i with respect to a (multiple criteria) value function, as pointed out at the end of Section 4. 'Normative' refers to the existence of postulates which, once accepted, necessarily lead to a specific method. Regarding the combination of uncertain information, there are no genuine normative approaches available although there exist characterization theorems in various frameworks. Thus the min-based combination is the only conjunctive idempotent attitude, whereas Dempster's rule of combination is not idempotent (even though there exist other combination rules that are more cautious and idempotent). Concerning the decision step, the situation is a bit different since there exist well established normative frameworks. Regarding decision under uncertainty, one has been recently proposed for von Neumann-Morgensten and Savage-like justifications of the possibility theory-based approach [Dubois and Prade, 1995], [Dubois et al., 1998b]. For the TBM, the pignistic transformation produces a probability used for decision making; its justification is detailed in [Smets and Kennes, 1994] and the avoidance of any Dutch-book is explained in [Smets, 1993b].

Regarding empirical validation, we face a cognitive problem. Is the representation framework used cognitively meaningful? Are the operations performed on the representation meaningful? Preliminary (positive) elements of answers concerning the possibility theory framework can be found in

[Raufaste and Da Silva Neves, 1998].

Concerning the computational issue, the two methods presented are clearly computationally manageable in practice (even if one is simpler). The fact that Dempster's rule of combination is NP complex is not a real issue here as the kind of data expected are simple and the involved belief functions have few components, making the computation tractable in practice.

Lastly, it is also important to judge the (potential) explanation capabilities. Are the results provided by the system easy to explain to the user if necessary? E.g., why the value of the candidate is finally assessed as such? Such explanation capabilities are presented in [Farreny and Prade, 1992] and in [Xu and Smets,1996] for the QPT and the TBM, respectively. Generally, the qualitative framework of possibility theory allows for a logical reading and processing of the evaluations [Benferhat et *al.*, 1997].

6 Concluding remarks

In this paper, we have taken advantage of a generic value assessment problem for discussing the different facets of the problem and raising the various difficulties and hypotheses which should be made at each step for computing a meaningful evaluation through two models, the QPT and the TBM.

The purpose of the paper was twofold: first identifying and discussing various facets of a value assessment problem, and second showing how two different approaches (which have in common the capability of modeling imprecision) can handle the problem in manners which are in fact quite parallel.

The statement of the problem contains no numerical data, but only ordinal assessments. Owing to the qualitative framework of possibility theory , the QPT provides a natural approach. However note that it is important to consider the nature of the scales which are used in order to know what operations are meaningful on them. Moreover commensurateness hypotheses are necessary.

With the TBM, numbers (the beliefs) are needed. They are in fact analogous to those a probability approach would ask for. The difference between the TBM and a probability approach is that the TBM accepts and uses the data just as they are, without introducing extraneous data. E.g., a probability approach would allocate (equal) probabilities to each of the individual values of the scores when the score is only known as a non-degenerate interval.

Sensitivity analysis can be performed, where either the values of the parameters (TBM) or the aggregation operators (QPT) are slightly modified and the robustness of the conclusions can be assessed. Furthermore the DM can also determine the sensitivity of the results if more precise data were collected. Given this information, the DM could efficiently ask to a specific expert to provide a better estimate of the candidate score for a specific criterion, avoiding getting "useless" data.

7 Acknowledgments

This work has been supported by the European working group FUSION (Esprit Project 20001: Fusion of Uncertain Data from Sensors, Information Systems and Expert Opinions). Special thanks are due to Simon Golstein who helped in stating the candidate value assessment problem considered in this paper. The authors are indebted to Patrice Perny for valuable comments on the meaning of the absence of commutation of the multiple criteria aggregation and the expert opinion fusion operations.

References

[Benferhat et *al.*, 1997] Benferhat S., Dubois D., Prade H. (1997) From semantic to syntactic approaches to information combination in possibilistic logic. In: Aggregation and Fusion of Imperfect Information, (B.Bouchon-Meunier, ed.), Physica. Verlag, 141-161.

[Biswas, 1995] Ranjit Biswas, An application of fuzzy sets in student's evaluation. Fuzzy Sets and Systems. 74:187-194, 1995.

[Bana and Vansnick] Carlos A. Bana e Costa, Jean-Claude Vansnick, A theoretical framework for measuring attractiveness by a categorical based evaluation technique (MACBETH). XIth Int. Conf. on MCDM, August 1994, Coimbra, Portugal,15–24

[Club CRIN Flou, 1997] Club CRIN Logique Floue. Evaluation subjective - Méthodes, Applications et Enjeux. Association ECRIN, 32 bd de Vaugirard, 75015 Paris.

[Dubois et *al.*, 1997] Didier Dubois, Henri Prade and Régis Sabbadin, A possibilistic logic machinery for qualitative decision. Proc. of the AAAI'97 Spring Symp. Series on Qualitative Preferences in Deliberation and Practical Reasoning, Stanford, CA, March, 47–54, 1997.

[Dubois et *al.*, 1988] Didier Dubois, Henri Prade and Testemale, Weighted fuzzy pattern matching. Fuzzy Sets and Systems, 28(3), 313–331, 1988.

[Dubois et *al.*, 1999a] Didier Dubois, Michel Grabisch and Henri Prade, Assessing the value of a candidate. A qualitative possibilistic approach. In Symbolic and Quantitativ e Approaches to reasoninn and Uncertainty Proc. (ECSQARU 99). LNAI 1638, Springler Veralg, 137-147,1999.

[Dubois et *al.*, 1999b] Didier Dubois, Michel Grabisch, Henri Prade and Philippe Smets, Assessing the value of a candidate. Comparing belief functions and possibility theories. Proc. XVth Conf. Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, 170-177, 1999.

[Dubois and Prade, 1983] Didier Dubois and Henri Prade, Ranking fuzzy numbers in the setting of possibility theory, Inf. Sci., 30, 1983.

[Dubois and Prade, 1994a] Didier Dubois and Henri Prade, Possibility theory and data fusion in poorly informed environments. Control Engineering Practice, 2(5), 811–823, 1994.

[Dubois and Prade, 1994b] Didier Dubois and Henri Prade, Decision making under fuzzy constraints and fuzzy criteria - mathematical programming versus rule-based approach. In Fuzzy Optimization - Recent Advances. M. Delgado, J. Kacprzyk, J.L. Verdegay and M.A. Vila, eds. Physica-Verlag, 21-32, 1994.

[Dubois and Prade, 1995] Didier Dubois and Henri Prade, Possibility theory as a basis for qualitative decision theory. Proc. of the 14th Inter. Joint Conf. on Artificial Intelligence (IJCAI'95), Montréal, Canada, 1924-1930, 1995.

[Dubois and Prade, 1997] Didier Dubois and Henri Prade, Fuzzy criteria and fuzzy rules in subjective evaluations - a general discussion. In Proc. 5th European Congress on Intelligent Technologies and soft Computing (EUFIT 97), Aachen, Germany, 975-978, 1997.

[Dubois and Prade, 1998a] Didier Dubois and Henri Prade, Possibility theory : qualitative and quantitative aspects, In *Quantified Representation of Uncertainty and Imprecision*(Ph. Smets eds.), Vol. 1 of the Handbook of Defeasible Reasoning and Uncertainty Management Systems (D.M. Gabbay, Ph. Smets, series eds.), Kluwer Academic Publishers, 169–226,1998.

[Dubois et al., 1998b] Didier Dubois, Henri Prade and Regis Sabbadin, Qualitative decision theory with Sugeno integrals, in Proc. of the 14th Conf. Uncertainty in Artificial Intelligence (UAI'98), Madison, USA, July 24-26, 1998, Morgan Kaufmann, 121-128.

[Farreny and Prade, 1992] Henri Farreny and Henri Prade, Positive and negative explanations of uncertain reasoning in the framework of possibility theory, in Fuzzy Logic for the Managment of Uncertainty (L.A. Zadeh, J. Kacpryk, eds.), Wiley, 1992, 319-333.

[Grabisch et *al.*, 1995] Michel Grabisch, Hung T.Nguyen and Elbert A. Walker, Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Kluwer Acad. Publ., 1995.

[Grabisch and Roubens, 2000] Grabisch M., Roubens M. (2000) Application of the Choquet integral in multiattribute decision making. In Fuzzy Measures and Integrals. Theory and Applications. Physica-Verlag, 348-374.

[Raufaste and Da Silva Neves, 1998] Eric Raufaste and Rui Da Silva Neves, Empirical evaluation of possibility theory in human radiological diagnosis", in Proc. of the 13 th Europ. Conf. on Artificial Intelligence (ECAI'98), Brighton, UK, Aug. 23-28, 1998, John Wiley & Sons, 124-128.

[Shafer, 1976] Glenn Shafer, A mathematical theory of evidence. Princeton Univ. Press. Princeton, NJ., 1976.

[Smets, 1993a] Philippe Smets, Belief Functions : the Disjunctive Rule of Combination and the Generalized Bayesian Theorem. International Journal of Approximate Reasoning, 9 : 1-35, 1993.

[Smets, 1993b] Philippe Smets, No Dutch Book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. SIS, Workshop on Probabilistic Expert Systems, (ed. R. Scozzafava), Roma, 1993, 181-204.

[Smets and Kennes, 1994] Philippe Smets and Robert Kennes, The transferable belief model.Artificial Intelligence, 66 : 191-234, 1994.

[Smets, 1998] Philippe Smets, The Transferable Belief Model for Quantified Belief Representation. Handbook of defeasible reasoning and uncertainty management systems. Gabbay D. M. and Smets Ph. Eds. Vol. 1, Kluwer, Doordrecht, 1998, 267-301.

[Sugeno, 1977] Michio Sugeno, Fuzzy measures and fuzzy integrals: A survey. In: Fuzzy Automata and Decision Process (eds.), North-Holland, 89-102,1977.

[Xu and Smets,1996] Hong Xu and Philippe Smets, Some Strategies for Explanations in Evidential Reasoning Explaining the origin of a computed belief and finding which pieces of evidence are influencing the results. IEEE Transactions on Systems, Man and Cybernetics, 26A : 599-607, 1996.

[Zadeh, 1975] Lotfi A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Inform. Sci. Part 1, 8 (1975) 199–249; Part 2, 8 (1975) 301-357; Part 3 (1975) 43-80, 1975.

The TBM flowchart

	Start	
Quantitative Poss	$\downarrow$	
	Poss(Expert says C_{ij}| actual value c_i)	
Poss = pl	$\downarrow$	$Poss(A\|B) = pl[B](A) = pl[A](B)$
	$m^{\text{actual value } c_i}$[What expert i states about criteria j]	
Numerical discounting	$\downarrow$	discounting
	$m^{\text{actual value } c_i}$[What i states about j, My opinion about i for j]	
	$\searrow\downarrow\swarrow$	
Distinctness	$\oplus$	conjunctive combination
	$m^{\text{actual value } c_i}$[What all state about j, My opinion about experts]	
f:criteria$\rightarrow$goodness	$\downarrow$	extension principle
	m^{Goodness}[All I know about criteria c_i]	
	$\searrow\downarrow\swarrow$	
Distinctness	$\oplus$	conjunctive combination
	m^{Goodness}[All I know]	
	$\downarrow$	pignistic transformation
	BetP over Goodness of candidate K	
Numerical scale	$\downarrow$	expected utility
	Expected Goodness of candidate K	

Table 22: The TBM flowchart. On the left, the needed assessments and assumptions. On the right of the arrows, the used operators.

The QPT flowchart

Start

Qualitative Possibilities $\downarrow$ representation of fuzzified data

Poss(actual value c_i | Expert says C_{ij})

Qual. opinion on i for j $\downarrow$ discounting

Poss(actual c_i | Expert says C_{ij}, My opinion about i for j)

$\searrow\downarrow\swarrow$

Merging expert opinions: weighted min (max if conflict)

Disj. combi. if conflict $\downarrow$

Normalization of merged opinions

Define aggregation rule $\downarrow$ extension principle on median-based aggregation.

Satisfaction pattern : Poss(score | All I know & aggregation rule)

$\downarrow$

Optimistic and pessimistic expectations about global value of K.

Table 23: The QPT flowchart. On the left, the needed assessments and assumptions. On the right of the arrows, the used operators.