Empirical Models for the Dempster-Shafer-Theory

Mieczysław Alojzy Kłopotek[13] and Sławomir Tadeusz Wierzchoń[12]

[1] Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland
[2] Institute of Computer Science, Białystok University of Technology, Poland
[3] Institute of Computer Science, University of Podlasie, Poland

Abstract. In spite of many useful properties, the Dempster-Shafer Theory of evidence (DST) experienced sharp criticism from many sides. The basic line of criticism is connected with the relationship between the belief function (the basic concept of DST) and frequencies [65,18]. A number of attempts to interpret belief functions in terms of probabilities have failed so far to produce a fully compatible interpretation with DST - see e.g. [34,18,14] etc. As a way out of those difficulties, in the paper we will explain our three model proposals: (1) "the marginally correct approximation", (2) "the qualitative model", (3) "the quantitative model". All of them fit the framework of DST, especially the Dempster rule of combination of evidence that was the hardest point and the point of failure of previously known attempts.

1 Introduction

The Mathematical Theory of Evidence or MTE [50] known also as the Dempster-Shafer Theory, DST, has been deeply investigated with respect to formal properties. Hence its relation to Choquet capacities [67], the axiomatic systems for propagation of uncertainty by local computations [55] (in so-called Markov trees) and for graphoidal representation of belief functions [54] have been studied. Optimal query answering systems have been elaborated [68].

However, the relationship between experimental observations and the DST belief functions is still a hot topic of research. Establishing empirical models for belief functions may contribute strongly to future application of DST in many areas, including also economy [13,21,62].

By an appropriate empirical model for DST we understand the following (see fig.1 below):

We observe ("measure") a real world $state_1$, encode it as a belief function Bel_1, know that a real world process may be represented by a DST reasoning process Bel_p, run the reasoning while the real world process runs in the real world transforming it into the $state_2$. We observe (in exactly the same way as before) the real world and encode the $state_2$ as belief function Bel_2. Bel_2 shall coincide or at least be one of alternative predictions of the reasoning process.

In the past a number of different attempts has been made to identify a model like this. In the paper we will outline major attempts like lower

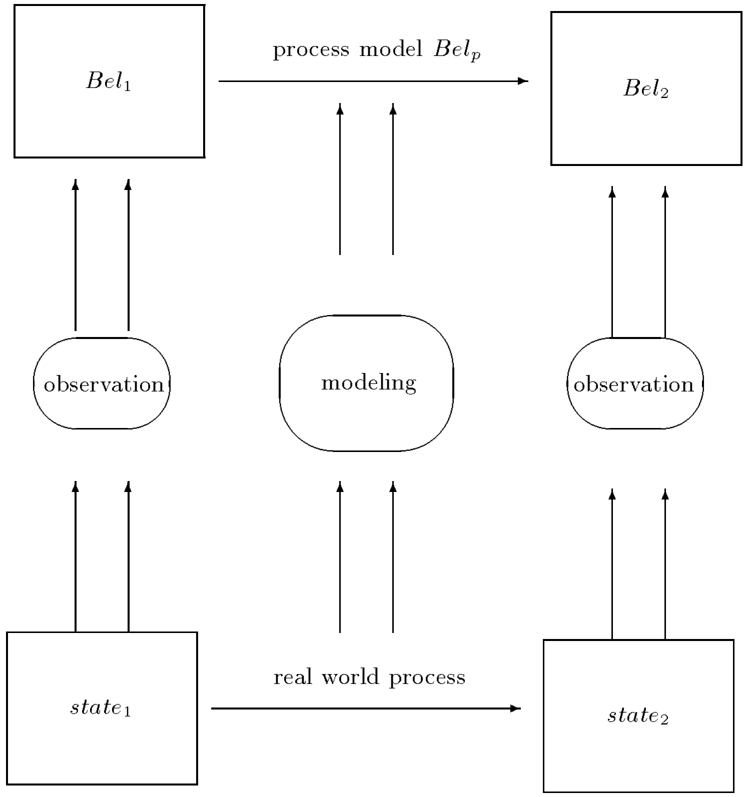

Fig. 1. Real World versus Dempster Shafer model

and upper probabilities, envelops of probability function families, probability structures approach, Shafer's Gamma mapping, rough set approaches etc. and point at their weaknesses and strengths. Though these approaches allow for derivation of a belief function describing the real world state properly, they fail in general to behave correctly under the reasoning (Bel_2 does not reflect a real world state even under "ideal" real world process).

It is still more dramatic if one attempts to identify graphoidal structures in the belief functions (needed in general to describe reasoning processes in a computationally efficient way). Conditional belief functions turn out to be negative-valued and hence cannot be interpreted by means of frequency-related concepts (like lower/upper conditional probabilities etc.). Still worse, combination of conditional belief functions in a graphoidal structure does not need to yield a proper joint belief function. Therefore the graphoidal structures of belief functions are poorly investigated and seldom used.

As a way out of those difficulties, in the chapter we will explain our three model proposals:

- "the marginally correct approximation".
- "the qualitative model"
- "the quantitative model"

The marginally correct approximation assumes that the belief function shall constitute lower bounds for frequencies, however only for the marginals, and not for the joint distribution. Then the reasoning process is expressed in terms of so-called Cano et al. conditionals - a special class of conditional belief functions that are positive. This approach implies modification of reasoning mechanism, because the correctness is maintained only by reasoning forward. Depending on the reasoning direction we need different "Markov trees" for the reasoning engine.

The qualitative approach is based on earlier rough set interpretations of DST, but makes a small and still significant distinction. All computations are carried out in strictly "relational" way that is indistinguishable objects in a database are merged (no object identities). The behavior under reasoning fits strictly to DST reasoning model. Factors of hypergraph representation can be expressed by relational tables. Conditional independence is well defined. However, there is no interpretation for conditional belief functions in this model.

The quantitative model assumes that during the reasoning process one attachs labels to objects hiding some of their properties. There is a full agreement with the reasoning mechanism of DST. Conditional independence and conditional belief functions are well defined. We have also elaborated processes that can give rise to well-controlled graphoidally structured belief functions. We elaborated also learning procedures for discovery of graphoidal structures from data.

The quantitative model seems to be the best fitting model for belief functions created so far.

1.1 Some Comments

The Dempster-Shafer Theory or the Mathematical Theory of Evidence (MTE) [50,9] is intended to be a generalization of bayesian theory of subjective probability [53]. This theory offers capability of representing ignorance in a simple and direct way, compatibility with the classical probability theory, compatibility with boolean logic and feasible computational complexity [45]. MTE may be applied for (1) representation of incomplete knowledge, (2) belief updating, (3) and for combination of evidence [43].MTE covers the statistics of random sets and may be applied for representation of incomplete statistical knowledge. Random set statistics is quite popular in analysis of opinion polls whenever partial indecisiveness of respondents is allowed

[11]. Practical applications of MTE include: integration of knowledge from heterogeneous sources for object identification [7], technical diagnosis under unreliable measuring devices [12], medical applications: [16,70], and for network reliability computation [42], reliability in real-time X-ray radioscopy and ultrasounds [10], multisensor image segmentation [2], safety control in large plants [20], map construction and maintenance [38], just to mention a few.

In spite of indicated merits, MTE experienced sharp criticism from many sides. The basic line of criticism is connected with the relationship between the belief function (the basic concept of MTE) and frequencies [65,18].

The problem of frequencies is not solely a scholar problem. It has significant knowledge engineering (expressive power, sources of knowledge, knowledge acquisition strategies, learning algorithms) and software engineering implications (internal representation, measure transformation procedures). First of all one should realize that a computer-based advisory system is rarely made for a single consultation. Hence we may (at least theoretically) obtain a statistics of cases for which the system has been applied. Life verifies also frequently enough the advices obtained from the advisory system. Hence after a long enough time one may pose at least partially the question whether or not the advices have been correct. A belief function (the basic concept of DST) without a modest frequency interpretation may be treated as a void answer in this context.

Therefore numerous probabilistic interpretations have been attempted since the early days of DST, see e.g. [9,51,34,19,40,42,35,58,14,18] and other.

But a number of attempts to interpret belief functions in terms of probabilities have failed so far to produce a fully compatible interpretation with DST - see e.g. [35,34,18,14] etc. As Smets [59] states "all these interpretations lead to different updating (conditioning) rules." [59, pp. 324-325]. Shafer [53] and Smets [59], in defense of MTE, dismissed every attempt to interpret MTE frequentistically. Shafer in [53] claims that probability theory developed over last years from the old-style frequencies towards modern subjective probability theory within the framework of bayesian theory. By analogy he claims that the very attempt to consider relation between DST and frequencies is old-fashioned and out of date. Wasserman opposes this view ([65], p.371) reminding "major success story in Bayesian theory", the exchangeability theory of de Finetti. It treats frequencies as special case of bayesian belief. "The Bayesian theory contains within it a definition of frequency probability and a description of the exact assumptions necessary to invoke that definition" [65]. Wasserman dismisses Shafer's suggestion that probability relies on analogy of frequency. .

1.2 Basics of the Dempster-Shafer Theory

Below we define the way we understand the basics of the Depster-Shafer theory (see [50]). On the one hand we include here the definitions of basic measures of uncertainty (belief, plausibility, mass function, commonality), operatorsd of marginalization and vacuous extension, the Dempster rule of evidence combination, rule of conditioning. Furthermore we consider as indispensible part of DST the Shafer's and Shenoy's representation of joint belief in terms of a decomposition and reasoning algorithms designed for that decomposition (see next subsection).

So let us start with basic notion.

Let Ξ be a finite set of elements called elementary events. Any subset of Ξ is a composite event, or hypothesis. Ξ be called also the frame of discernment.

Definition 1. [50] Let Ξ be a finite set of elements called elementary events. Any subset of Ξ be a composite event. Ξ be called also the frame of discernment.
A basic probability assignment (bpa) function (called also mass function) is any function m:$2^\Xi \rightarrow [0,1]$ such that

$$\sum_{A\in 2^\Xi} m(A) = 1 \qquad m(\emptyset)=0, \qquad \forall_{A\in 2^\Xi} \quad 0 \leq \sum_{A\subseteq B} m(B)$$

We say that a bpa is vacuous iff $m(\Xi)=1$ and $m(A)=0$ for every $A \neq \Xi$.

Definition 2. [50] A belief function be defined as Bel:$2^\Xi \rightarrow [0,1]$ so that $Bel(A) = \sum_{B\subseteq A} m(B)$ A plausibility function be Pl:$2^\Xi \rightarrow [0,1]$ with $\forall_{A\in 2^\Xi}\ Pl(A) = 1 - Bel(\Xi - A)$ A commonality function be Q:$2^\Xi \rightarrow [0,1]$ with $\forall_{A\in 2^\Xi} \quad Q(A) = \sum_{A\subseteq B} m(B)$

Definition 3. [50] The Rule of Combination of two Independent Belief Functions Bel_{E_1}, Bel_{E_2} Over the Same Frame of Discernment (the so-called Dempster-Rule), denoted

$$Bel_{E_1,E_2} = Bel_{E_1} \oplus Bel_{E_2}$$

is defined as follows: :

$$m_{E_1,E_2}(A) = c \cdot \sum_{B,C;A=B\cap C} m_{E_1}(B)\cdot m_{E_2}(C)$$

(c - constant normalizing the sum of m to 1)

Furthermore, let the frame of discernment Ξ be structured in that it is identical to cross product of domains Ξ_1, Ξ_2, $\ldots \Xi_n$ of n discrete variables $X_1, X_2, \ldots X_n$, which span the space Ξ. Let $(x_1, x_2, \ldots x_n)$ be a vector in the space spanned by the variables $X_1, , X_2, \ldots X_n$. Its projection onto the subspace spanned by variables $X_{j_1}, X_{j_2}, \ldots X_{j_k}$ ($j_1, j_2, \ldots j_k$ distinct indices from the set 1,2,...,n) is then the vector $(x_{j_1}, x_{j_2}, \ldots x_{j_k})$. $(x_1, x_2, \ldots x_n)$ is also called an extension of $(x_{j_1}, x_{j_2}, \ldots x_{j_k})$. A projection of a set A of such vectors is the set $A^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}$ of projections of all individual vectors from

A onto $X_{j_1}, X_{j_2}, \ldots X_{j_k}$. A is also called an extension of $A^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}$. A is called the vacuous extension of $A^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}$ iff A contains all possible extensions of each individual vector in $A^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}$. The fact, that A is a vacuous extension of B onto space $X_1, X_2, \ldots X_n$ is denoted by $A = B^{\uparrow X_1, X_2, \ldots X_n}$

Definition 4. Let m be a basic probability assignment function on the space of discernment spanned by variables $X_1, X_2, \ldots X_n$. $m^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}$ is called the projection (or marginalization) of m onto subspace spanned by $X_{j_1}, X_{j_2}, \ldots X_{j_k}$ iff

$$m^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}(B) = \sum_{A; B = A^{\downarrow X_{j_1}, X_{j_2}, \ldots X_{j_k}}} m(A)$$

Definition 5. Let m be a basic probability assignment function on the space of discernment spanned by variables $X_{j_1}, X_{j_2}, \ldots X_{j_k}$. $m^{\uparrow X_1, X_2, \ldots X_n}$ is called the vacuous extension of m onto superspace spanned by $X_1, X_2, \ldots X_n$ iff

$$m^{\uparrow X_1, X_2, \ldots X_n}(B^{\uparrow X_1, X_2, \ldots X_n}) = m(B)$$

and $m^{\uparrow X_1, X_2, \ldots X_n}(A) = 0$ for any other A.

Projections and vacuous extensions of Bel, Pl and Q functions are defined with respect to operations on m function. Notice that by convention if we want to combine by Dempster rule two belief functions not sharing the frame of discernment, we look for the closest common vacuous extension of their frames of discernment without explicitly notifying it.

Definition 6. (See [52]) Let B be a subset of Ξ, called evidence, m_B be a basic probability assignment such that $m_B(B) = 1$ and $m_B(A) = 0$ for any A different from B. Then the conditional belief function $Bel(.||B)$ representing the belief function Bel conditioned on evidence B is defined as: $Bel(.||B) = Bel \oplus Bel_B$.

1.3 Basics of belief function decomposition and reasoning

An integral part of DST is a particular type of inference engine - Shenoy-Shafer method of local computations (for presentation of this method the reader should consult the original paper of them [55]). The subsequent definitions of hypergraphs and operations on them are due to [55].

Hypergraphs: A nonempty set H of nonempty subsets of a finite set S be called a hypergraph on S. The elements of H be called hyperedges. Elements of S be called vertices. H and H' be both hypergraphs on S, then we call a hypergraph H' a *reduced hypergraph* of the hypergraph H, iff $H' \subseteq H$ holds, and for every $h \in H$ there exists such a $h' \in H'$ that $h \subseteq h'$.A hypergraph H *covers* a hypergraph H' iff for every $h' \in H'$ there exists such a $h \in H$ that $h' \subseteq h$.

Hypertrees: t and b be distinct hyperedges in a hypergraph H, $t \cap b \neq \emptyset$, and b contains every vertex of t that is contained in a hyperedge of H other than t; if $X \in t$ and $X \in h$, where $h \in H$ and $h \neq t$, then $X \in b$. Then we call t a twig of H, and we call b a branch for t. A twig may have more than one branch. We call a hypergraph a hypertree if there is an ordering of its hyperedges, say $h_1, h_2, ..., h_n$ such that h_k is a twig in the hypergraph $\{h_1, h_2, ..., h_k\}$ whenever $2 \leq k \leq n$. We call any such ordering of hyperedges a hypertree construction sequence for the hypertree. The first hyperedge in the hypertree construction sequence be called the root of the hypertree construction sequence.

Variables and valuations:Let $\mathbf{V}$ be a finite set. The elements of $\mathbf{V}$ are called variables. For each $h \subseteq \mathbf{V}$ there is a set VV_h. The elements of VV_h are called valuations. Let VV=$\bigcup\{VV_h | h \subseteq \mathbf{V}\}$ be called the set of all valuations.

In case of probabilities a valuation on h will be a non-negative, real-valued function on the set of all configurations of h(a configuration on h is a vector of possible values of variables in h). In the belief function case a valuation is a non-negative, real-valued function on the set of all subsets of configurations of h.

Proper valuation: for each $h \subseteq \mathbf{V}$ there is a subset P_h of VV_h elements of which are called proper valuations on h. Let P be the set of all proper valuations.

Combination: We assume that there is a mapping $\odot : VV \times VV \rightarrow VV$ called combination such that:
(i) if G and H are valuations on g and h respectively, then $G \odot H$ is a valuation on $g \cup h$
(ii) if either G or H is not a proper valuation then $G \odot H$ is not a proper valuation
(iii) if both G and H are proper valuations then $G \odot H$ may be or not be a proper valuation

Factorization: Suppose A is a valuation on a finite set of variables $\mathbf{V}$, and suppose HV is a hypergraph on $\mathbf{V}$. If A is equal to the combination of valuations of all hyperedges h of HV then we say that A factorizes on HV.

Let us remind also the following: (See [52]) Let B be a subset of Ξ, called evidence, m_B be a basic probability assignment such that $m_B(B) = 1$ and $m_B(A) = 0$ for any A different from B. Then the conditional belief function $Bel(.||B)$ representing the belief function Bel conditioned on evidence B is defined as: $Bel(.||B) = Bel \oplus Bel_B$.

The clue behind the uncertainty propagation method of Shenoy and Shafer is that if we make a factorization in terms along a hypertree structure (a Markov tree) then calculation of the joint belief distribution of variables belonging to a single hyperedge allows for a divide and conquer approach in that one considers this hyperedge as the root of a Markov tree the removal of which splits the Markov tree into as much disjoint Markov subtrees as there are twigs on the root etc. The calculation in the root requires only info from its twigs and the twigs calculate the info from their own Markov subtrees without any concern about other twigs of the root and their Markov subtrees. See [55] for detailed information.

2 Lower/Upper Bound Interpretation

Mathematical Theory of Evidence (MTE) is blamed to leave frequencies outside its framework. In this section we consider this problem from the point of view of conditioning in the MTE. We describe the class of belief functions for which marginal consistency with observed frequencies may be achieved and conditional belief functions are proper belief functions,and deal with implications for approximation of general belief functions by this class and for inference models in MTE.

belief

2.1 Case-based Understanding of Belief Functions

In an attempt to overcome the reason for those numerous failures to interpret consistently Dempster's rule of combination, a new frequency interpretation of MTE has been proposed in [29,30]. It has been demonstrated there that, in general, Dempster's rule has a destructive impact on the data so that whatever we expect calculation of conditional probabilities (given an event) differs from whatever we obtain under conditioning (given an event) in MTE (see the definition of aposteriorical conditioning of Shafer in [51].

Let us have a look at SQL [64] construct to calculate joint probability distribution from a database in two variables P(X,Y):

create view Total (Counted) as select count(*) from Cases;

create view Probability (X,Y,Prob) as select X,Y,count(*)/Counted from Cases,Total group by X,Y;

Let us look at calculation of conditional probability $P(X,Y|X \in A_X)$. We proceed as follows: we select first the proper subset of cases from the database and proceed to calculate the unconditional probability for the selected cases.

create view SelectedCases (X,Y) as select X,Y from Cases where $X \in AX$*;*

create view TotalCond (Counted) as select count(*) from SelectedCases;

create view CondProbability (X,Y,CondProb) as select Y,X,count(*)/Counted from SelectedCases,TotalCond where $X \in AX$ *group by X, Y;*

If we calculate a conditional probability $P(X, Y | X \in A_X, X \in B_X)$ on a series of conditions $(X \in A_X, X \in B_X)$ we can proceed first by selecting cases for the first, then for the second condition etc., finding the intersection, and then calculating the probabilities (id - the identifier).

create view SelectedCasesA (id,X,Y) as select id,X,Y from Cases where $X \in AX$*;*

create view SelectedCasesB (id,X,Y) as select id,X,Y from Cases where $X \in BX$*;*

create view SelectedCases (id,X,Y) as select id,X,Y from SelectedCasesA intersection select id,X,Y from SelectedCasesB ;

create view TotalCond (Counted) as select count(*) from SelectedCases;

create view CondProbability (X,Y,CondProb) as select Y,X,count(*)/Counted from SelectedCases,TotalCond where $X \in AX$ *group by X, Y;*

Let us look at calculation of basic probability assignment (bpa) m(X,Y) from data :

create view Total (Counted) as select count(*) from Cases;

create view bpa (X,Y,m) as select X,Y,count(*)/TCounted from Cases,Total group by X,Y;

Let us look at calculation of aposterioric conditional bpa $m(X, Y || X \in A_X)$ We have to proceed as follows: we select first the proper subset of cases from the database and MODIFY the values for the variable X. Then we proceed to calculate the unconditional bpa for the selected and updated cases.

create view UpdatedCases (X,Y) as select $X \cap AX$*,Y from Cases where* $X \cap AX \neq \emptyset$*;*

create view TotalCond (Counted) as select count(*) from UpdatedCases;

create view Cond_bpa (X,Y,Condbpa) as select X,Y,count(*)/Counted from UpdatedCases,TotalCond ;

If we calculate a conditional bpa $m(||X \in A_X, X \in B_X)$ on a series of conditions $(X \in A_X, X \in B_X)$ we CANNOT proceed by selecting cases for the first, then for the second condition etc., finding the intersection, and then calculating the bpa, because this would yield wrong results due to side effects stemming from case modifications.

This has a serious impact if we try to factorize a belief function into simpler components, e.g. for purposes of propagation of uncertainty (methods of propagation of uncertainty are presented e.g. in [4,55]. It turns out that:

- It is, in general, impossible to factorize a joint belief distribution into components being conditional belief functions [1] (see eg. [4] for a discussion why two different notions of conditioning are needed for MTE: the posteriori-conditionals as conditionals in the sense of Shafer, and a priori-conditionals invented by Cano et al.,)
- A priori-conditional belief functions as proposed by Cano et al. in general do not exists (see a discussion on non-existence of a-priori conditionals in MTE presented by [54]) [2]
- What is more, it is often impossible to factorize a belief function Bel in variables p, q, r into two factors one Bel_1 in variables p, r and the second Bel_2 in variables q, r even if in conditional distribution given any value of the variable r, variables p, q are independent (see [63], that is when for every subset r_i of the domain of the variable r the following holds:

$$Bel(||r_i)^{\downarrow p,q} = Bel(||r_i)^{\downarrow p} \oplus Bel(||r_i)^{\downarrow q}$$

Therefore in papers [23,28,29] we have presented another approach to factorization of belief functions in terms of anticonditional belief functions. It turns out, however, that anticonditional belief functions are in general not belief functions but only pseudo-belief functions (that is ones with non-negative commonality functions). This means in practice, however, that anticonditional belief functions have no direct counterparts in the physical world, as the basic probability assignment may take negative values.

One can be tempted to suggest, that one shall then resign from modeling the joint belief distribution and instead try to find a marginally consistent decomposition of the joint belief distribution.

The natural question then arises:

- What is the class of Dempster-Shafer (DS) belief functions for which apriori-conditional belief functions exist ?
- How can general belief functions and uncertainty propagation for them be approximated by this class of belief functions and uncertainty propagation for them ?
- How can the belief functions and the reasoning with them be related to frequencies (cases)?

---

[1] Notice that in the domain of probability distributions, EVERY probability distribution may be represented by a composition of conditional probability distributions as so-called bayesian network

[2] In probability distributions, a-priori and a-posteriori conditionals from the point of view of frequencies coincide, and they always exist: Let us assume P(X,Y) is a frequency based (case-based) probability distribution of X,Y. A posteriori conditional distribution of variable X given variable Y=y is interpreted as uncertainty distribution of variable X $P'(X|Y=y)$ if we restrict ourselves only to cases for which Y=y. Apriori conditional distribution of X given Y is a function $P(X|Y)$ in X,Y defined as $P(X|Y) = P(X \cap Y)/P(Y)$ for all values of y for which $P(Y=y) > 0$. Obviously: $P'(X=x|Y=y) = P(X=x|Y=y)$ if $P(Y=y) > 0$.

2.2 Correct Approximation of Apriorical Conditional Belief Functions

Shafer [51] suggested that a belief function may emerge if we observe a variable X with domain of values Ξ_X indirectly (via a mapping f) by observing actually another variable Y with a domain Ξ_Y such that the mapping $f : \Xi_Y \to \Xi_X$ is not a function. Hence, in some cases, if we observe a single value ξ_Y of the variable Y, we can tell at most that the value of the variable X belongs to a non-empty set of elements $\{\xi_{X_1}, \xi_{X_2}, \ldots, \xi_{X_n}\}$. So a probability distribution P(Y) in Y translates into a basic belief assignment function, from which one calculates easily belief distribution Bel^X in X. In this way we could get a case-based distribution in variable X.

Variable Y	Corresponding set values of X (A)	frequency	$m^X(A)$	$Bel^X(A)$
y_1	$\{x_1\}$	10	.10	.10
y_2	$\{x_1, x_2\}$	20	.20	.30
y_3	$\{x_2, x_3\}$	30	.30	.70
y_4	$\{x_3\}$	40	.40	.40
–	$\{x_1, x_2, x_3\}$	–	.0	1.0

This could be extended simply to multivariate belief distributions. However, we encounter one interpretational problem:
Let us consider the following frequency table in variables X and Z:

X	Z	frequencies
$\{x_1, x_2\}$	$\{z_1, z_2\}$	20

What shall be the focal points of the multivariate belief function $Bel^{X,Z}$ in two variables X,Z ? Marginal consistency will be achieved either if we assume

$$m_1^{X,Z}(\{(x_1, z_1), (x_1, z_2), (x_2, z_1), (x_2, z_2)\}) = 1$$

and if we take

$$m_2^{X,Z}(\{(x_1, z_1), (x_2, z_2)\}) = 1$$

and if we suppose

$$m_3^{X,Z}(\{(x_1, z_2), (x_2, z_1)\}) = 1$$

and for many other belief functions. But which one is the best ? Let us try to calculate the conditional belief function $Bel_i(||X = x_1)^{\downarrow Z}$. (From Shafer's formula: $Bel_i(||X = x_1)^{\downarrow Z} = (Bel_i \oplus Bel_{X=x_i})^{\downarrow Z}$ with $Bel_{X=x_i}$ being a belief function with the only focal point $m_{X=x_i}(\{(x_1, z_1), (x_1, z_2)\}) = 1$). We get three totally different results - three belief functions with differing focal points: $m_1(||X = x_1)^{\downarrow Z}(\{z_1, z_2\}) = 1$, $m_2(||X = x_1)^{\downarrow Z}(\{z_1\}) = 1$, $m_3(||X = x_1)^{\downarrow Z}(\{z_2\}) = 1$.

You can easily see that it is next to impossible to decide which of these conditionals is the correct one from the point of view of the observed data. So which belief function shall be treated as the most representative for the data ? We suggest here the first one (Bel_1) for the following reasons:

1. if we observe X, Z separately, we have no reason to assume that x_1 must co-occur with z_1 but never with z_2 etc - we assume we have no more information than actually visible from the data,
2. the methods of uncertainty propagation suggested both by Cano et al. [4] and Shenoy & Shafer [55] implicitly assume that the joint belief in values of observed variables $X_1, X_2, ...X_n$ is the composition of the values of individual variables:

$$Bel_{observations} = Bel_{X_1=A_1} \oplus Bel_{X_2=A_2} \oplus \ldots \oplus Bel_{X_n=A_n}$$

   with $A_i \subseteq \Xi_{X_i}$ being a subset of the domain of the i^{th} variable. Violation of this assumption would invalidate the respective method of uncertainty propagation.

Let us consider now the following data in three (logical) variables X,Y,Z, giving the belief function Bel_{and}.

X=	Y=	Z=	frequency	$m_{and}(A_X \times A_Y \times A_Z)$
$\{t_X\}$	$\{t_Y\}$	$\{t_Z\}$	10	.10
$\{t_X\}$	$\{f_Y\}$	$\{f_Z\}$	10	.10
$\{f_X\}$	$\{t_Y\}$	$\{f_Z\}$	10	.10
$\{f_X\}$	$\{f_Y\}$	$\{f_Z\}$	10	.10
$\{t_X\}$	$\{t_Y, f_Y\}$	$\{t_Z, f_Z\}$	10	.10
$\{f_X\}$	$\{t_Y, f_Y\}$	$\{f_Z\}$	10	.10
$\{t_X, f_X\}$	$\{t_Y\}$	$\{t_Z, f_Z\}$	10	.10
$\{t_X, f_X\}$	$\{f_Y\}$	$\{f_Z\}$	10	.10
$\{t_X, f_X\}$	$\{t_Y, f_Y\}$	$\{t_Z, f_Z\}$	20	.20

Let us consider calculation of an apriori conditional belief function in the sense of Cano et al. [4] such that it would imply the value of Z given X,Y. A look at the data would suggest that variables X,Y and Z are connected by the logical equation: $X \wedge Y = Z$ so that one might suggest a belief function with focal point

$$m_{\&}(\{(t_X, t_Y, t_Z), (t_X, f_Y, f_Z), (f_X, t_Y, f_Z), (f_X, f_Y, f_Z)\}) = 1$$

is the apriori conditional connecting X,Y and Z. But this is a wrong conclusion, because:

$$Bel_{and} \neq Bel_{and}^{\downarrow\{X,Y\}} \oplus Bel_{\&}$$

What is more, the Cano et al a priori conditional of Z given X,Y does not exist at all !!! Therefore, decomposition of a joint belief distribution into conditionals cannot in general be achieved. However, we must acknowledge that marginal consistency is actually achieved, that is :

$$(Bel_{and})^{\downarrow X} = (Bel_{and}^{\downarrow\{X,Y\}} \oplus Bel_{\&})^{\downarrow X}$$

$$(Bel_{and})^{\downarrow Y} = (Bel_{and}^{\downarrow\{X,Y\}} \oplus Bel_{\&})^{\downarrow Y}$$

$$(Bel_{and})^{\downarrow Z} = (Bel_{and}^{\downarrow\{X,Y\}} \oplus Bel_{\&})^{\downarrow Z}$$

Can we always construct a marginally consistent apriori-conditional ? We shall assume that we have found a marginally consistent apriori conditional $Bel^{|p}$ of Belief function Bel given set of variables p iff $Bel^{|p}$ is a belief function and for every variable X:

$$(Bel)^{\downarrow X} = (Bel^{|p} \oplus Bel^{\downarrow p})^{\downarrow X}$$

let us consider the following belief distribution in variables X,Z

X	Z	frequencies
$\{x_1\}$	$\{z_1\}$	40
$\{x_1, x_2\}$	$\{z_2\}$	60

It is easily to show that marginally consistent $Bel^{|X}$ does not exist.

What is then the class of belief functions possessing marginally consistent conditionals ? What is the class of marginally consistent conditional belief functions ?

It is easy to show that if there exists a marginally consistent conditional belief function $Bel^{|p}$ of the belief function Bel given set of variables p then there exists another marginally consistent conditional belief function $Bel_{\Xi}^{|p}$ of the belief function Bel given set of variables p such that $(Bel_{\Xi}^{|p})^{\downarrow p}$ has only one focal point $(m_{\Xi}^{|p})^{\downarrow p}(\Xi_p) = 1$ with Ξ_p being the joint domain of variables from set p. In other words, Cano et al. apriori-conditionals represent completely the class of marginally consistent conditional belief functions.

Hence it is nearly obvious that in general case-based (separately measured) belief functions do not possess marginally consistent conditional belief functions.

Therefore it is in general of primary interest to find an appropriate approximation of general belief functions by means of decompositions into Cano's et al apriori-conditionals. Let us say that an approximation Bel' of belief function Bel is correct iff for every set A $Bel'(A) <= Bel(A)$. An approximation Bel' of belief function Bel is marginally correct iff for every variable X and every set A $Bel'^{\downarrow X}(A)) <= Bel^{\downarrow X}(A)$.

Let us consider the following algorithm for calculation of a correct approximation of the function Bel in variables X,Y:

0. We initialize the basic probability assignment function m_{cond} defined over variables X,Y with 0 for every subset of $\Xi_X \times \Xi_Y$.
1. For each set $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$ we calculate the quantity $g(A_X, A_Y) := m(A_X \times A_Y)/m^{\downarrow X}(A_X)$.
2. i:=1, q:=0
3. For each set $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$ we select a set $r_i(A_X) \subseteq \Xi_Y$ with $g(A_X, r_i(A_X)) > 0$.
   $g_{i,r,min}$ be the minimum of $g(A_X, r_i(A_X))$ over all $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$
4. For the relation r_i we select a function $a_i : \Xi_X \to 2^{\Xi_Y}$ such that for each set $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$ we have $r_i(A_X) \subseteq \bigcup_{\xi_X \in A_X} a_i(\xi_X)$
5. We update the function m_{cond} as follows: We calculate the set $A = \bigcup_{\xi_X \in \Xi_X} \{\xi_X\} \times a_i(\xi_X)$. Then $m_{cond}(A) := m_{cond}(A) + g_{i,r,min}$.
6. We update the g function as follows: For each set $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$ $g(A_X, r_i(A_X)) := g(A_X, r_i(A_X)) - g_{i,r,min}$
7. For each set $A_X \subseteq \Xi_X$ with $m^{\downarrow X}(A_X) > 0$

$$q := \frac{q + g_{i,r,min} * m^{\downarrow X}(A_X) * card(A_X)}{card(\bigcup_{\xi_X \in A_X} a_i(\xi_X))}$$

8. i:=i+1. If g is equal zero everywhere then terminate, otherwise continue with step 3.

The marginal quality of an approximation constructed by the above algorithm be the quantity q. The quality q can range from zero to one. If the quality is equal one then we have constructed a marginally consistent conditional of Bel given X. If Bel is a probabilistic belief function (with focal points being singleton sets), then a marginally consistent Cano's conditional of Bel given X always exists. If Bel is a general belief function possessing a marginally consistent Cano's conditional of Bel given X, then the construction by the above algorithm of the conditional is characterized by the fact that in step 4 we have $r_i(A_X) = \bigcup_{\xi_X \in A_X} a_i(\xi_X)$. However, the construction task as such is hard. In particular, we cannot assume that if there exists a Cano's conditional of Bel given X, and if and if for i=1,...,k we managed to obtain $r_i(A_X) = \bigcup_{\xi_X \in A_X} a_i(\xi_X)$, then we will get a $r_i(A_X) = \bigcup_{\xi_X \in A_X} a_i(\xi_X)$ for i=k+1. Consider the following (counter)example: Let Bel be a belief function separately measurable in X,Y, marginally consistent with $Bel_1 \oplus Bel_2$ where Bel_1 being a belief function in X, Bel_2 in X,Y with focal points:

A_X	$m_1(A_X)$
$\{x_1\}$	p_1
$\{x_2\}$	p_2
$\{x_3\}$	p_2
$\{x_1, x_2\}$	p_{12}
$\{x_1, x_3\}$	p_{13}
$\{x_2, x_3\}$	p_{23}

$A_{X,Y}$	$m_2(A_{X,Y})$
$\{(x_1, y_1), (x_2, y_1)(x_3, y_2)\}$	1/3
$\{(x_1, y_1), (x_2, y_2)(x_3, y_2)\}$	1/3
$\{(x_1, y_2), (x_2, y_1)(x_3, y_2)\}$	1/3

Obviously then Bel_2 is the marginally consistent Cano conditional of Bel. Let us select r_1 as follows:

A_X	$r_1(A_X)$
$\{x_1\}$	$\{y_1\}$
$\{x_2\}$	$\{y_1\}$
$\{x_3\}$	$\{y_1\}$
$\{x_1, x_2\}$	$\{y_1\}$
$\{x_1, x_3\}$	$\{y_1\}$
$\{x_2, x_3\}$	$\{y_1\}$

We can then easily construct function a_1 as $a_1(x_1) = y_1$, $a_1(x_2) = y_1$, $a_1(x_3) = y_1$ so that obviously $r_1(A_X) = \bigcup_{\xi_X \in A_X} a_1(\xi_X)$. However, if we increase in step 8 i to i=2 and reenter step 3, then it will not be possible any more to construct such an a_2 that $r_2(A_X) = \bigcup_{\xi_X \in A_X} a_2(\xi_X)$ because necessarily a fragment of r_2 will be:

A_X	$r_2(A_X)$
...	...
$\{x_1, x_2\}$	$\{y_1, y_2\}$
$\{x_1, x_3\}$	$\{y_1, y_2\}$
$\{x_2, x_3\}$	$\{y_1, y_2\}$

Hence in general finding a marginally consistent Cano's conditional, even if it exists, is hard and requires backtracking. So probably one will be satisfied already if one finds a high quality marginally correct approximate conditional belief distribution (Application of genetic algorithms is advised here.).

2.3 Impact on Reasoning

Let Bel be a belief function in X,Y and $Bel^{|X}$ its marginally correct approximate conditional belief distribution. Then $Bel^{\downarrow X} \oplus Bel^{|X}$ is a marginally correct approximation of Bel. Then Shafer's conditioning $Bel(||A_X \times \Xi_Y)$ with A_X being any non-empty subset of the domain Ξ_X of X is marginally correctly approximated by $(Bel^{\downarrow X} \oplus Bel^{|X})(||A_X \times \Xi_Y)$, that is given the "decomposition" of Bel into its marginal on X and the approximately correct conditional on X we can reason approximately correctly about the a posteriori distribution of Y given any event of observation of the intrinsic value of X. However, we cannot do it in the reverse direction: we cannot derive the aposterioric distribution of X by observation of Y because in general $Bel(||\Xi_X \times A_Y)$ with A_Y being any non-empty subset of the domain Ξ_Y of Y is NOT marginally correctly approximated by $(Bel^{\downarrow X} \oplus Bel^{|X})(||\Xi_X \times A_Y)$. What is worse, even if the conditionals are marginally consistent, we will not achieve marginal correctnesses, not to say marginal consistency.

This means that if we have a marginally consistent factorization (in terms of Cano's conditionals) of a belief function then we can in general use neither

Shafer & Shenoy nor Cano's at al. framework for propagation of uncertainty because the results will be inconsistent with the data. (If we are reasoning counter the stream of calculations of conditionals, we may fail to obtain correct result). A way out of this problem, for the framework of Cano, is to consider separate marginally correct approximations in each direction of reasoning. That is, given the polytree representing correctly dependencies among variables, we need to transform this polytree, for every variable the value of which we want to infer, into a such a (belief) network graph that every arrow points towards that variable. In case of plain directed trees the result of this transformation is a polytree (with undirected backbone identical with that of the original tree) such that edges connecting the target variable with its neighbous are reoriented to point at the target variable and every other variable is connected via a directed path with the target variable. In case of general polytree the result of the transformation is a network with more edges than the original polytree. Edges are added when an edge is reversed which originally pointed at a node with many ingoing edges. E.g. if we had the situation that $X->Z<-Y$ and we want to invert the edge $Z<-Y$, then we need to add the edge $X->Y$. In case of general polytree we obtain in this way networks with considerably different undirected backbones[3].

If we assume that we adopt a structure (factorization) of Cano type, that is in form of a generalized polytree (singly connected "bayesian" network), then products of above transformations are directly convertible into a Shenoy's and Shafer's Markov tree. Let us now shift to Shenoy's and Shafer's belief propagation in Markov tree: The general principle there is "message passing" - if a node of Markov-tree gets information from its all but one neighbors , then it sends, to the remaining node, a "message", that is $\oplus$ combination of those messages plus its own factor of the belief function factorization. In the original Shenoy/Shafer algorithm, this node's own factor of the belief function factorization is exactly the same independly to which neighbor the message is sent. We propose to have separate hypertrees for each target variable and to reason within each of them in one direction only (resp. modifications of propagation algorithm are known). Then it is guaranteed that the results of reasoning (a posteriori marginal distributions) will be marginally correct approximations with respect to the intrinsic distribution.

Let us now consider the meaning of Cano et al. conditionals. These are in fact sets of mappings between sets of variables selected with some probability. This gives a new meaning to the belief function. Instead of thinking in the way probability functions do that is that given some value of one variable, the conditional probability distribution assigns a value to another variable, we can think of objects that are assigned with some

---

[3] In case of a general belief network transformations aiming at creation of directed pathes from every node towards the target variable would result in still more complicated networks so that merits of reasoning in such networks would be questionable

probability a belief

2.4 Summary of Lower/Upper Bound Approximation

1. The case-based derivation of aposteriori MTE belief function requires combination of operations of case selection with operation of case updating.
2. Therefore, marginally correct inference rules for MTE derived from data can be applied only unidirectionally - reversal of direction of reasoning leads to contradictions with the data. For this reason, original Cano et al. and Shenoy/Shafer uncertainty propagation methods are not suitable for case-based MTE belief functions.
3. Hence in case of direct dependence of a set of n variables, n inferences networks - one for each variable as dependent on the remaining ones - have to be established and used depending on target variable.

3 Qualitative Interpretation

Though a tendency to consider belief functions as subjective uncertainty measures is visible [53], the need for case-based interpretation as a pre-condition for practical applicability has been explicitly stressed [65]. Still this interpretation should be qualitative rather than quantitative in nature [59]. However, these requirements seem not to be met so far.

Apart from manageable interpretation, the MTE caused troubles of purely numerical nature. Unlike ordinary probabilities, which assign mass to each possible outcome, belief functions assign mass to each subset of the outcome space. As a consequence the amount of memory space required to store a belief function in a computer will grow exponentially with the size of the outcome space we consider. Searching for economical representation of a belief function, the researches made use of the already known concepts of sets factorisation analysed in the theory of relational databases. In our search for non-quantitative interpretation of MTE, our attention was attracted by the nature of the join operator of relational databases [3] or in general the multivalued dependency [8] the study of which led to invention of local computation method for uncertainty propagation of Shenoy and Shafer [54] for MTE. This in turn made the rough-set theoretic interpretation of MTE belief functions of Skowron and Busse [58] best choice for further investigation, as it was purely-case based and relational, though it is frequency based. The rough-set interpretation sheds some light onto what the concept of "evidence" may mean in experimental terms. The "evidence" there is the information part of a database record and it "supports" the decision part of a record. The complexity of combination of evidence according to the Dempster rule in [58] gave an impulse for search of a simpler way to accomplish it. It turns out that with frequency approach updating of decision parts of cases is needed (Dempster's combination is destructive). Out of this experience we decided to abandon the frequencies and concentrated on purely relational operations.

The section recalls traditional rough set theoretic frequency interpretation of DST from [58] and explains our insights of destructive nature of Dempster's combination with respect to frequencies. Then it presents our new qualitative interpretation. The section ends with some concluding remarks.

Throughout the section, as relational data tables are subject of rough set theory, SQL [64] query language is used to express semantics of DST measures and operators in terms of decision tables, both with respect to traditional and our rough sets based interpretation of DST, as SQL has the capability of expressing purely relational and frequentistic data processing.

3.1 Rough Set Theory. The Traditional Interpretation of Belief Functions

Skowron and Grzymala-Busse [58] and others studying rough sets developed more specifically the proposal of Shafer with respect to frequency interpretation of DST.

Let us introduce the following denotation concerning decision tables. Let a tuple μ mean a function $\mu : A \rightarrow DOM(A)$, with A being a set of attributes A_j, $DOM(A_j)$ being the domain of the attribute A_j, $DOM(A) = \bigcup_{A_j \in A} DOM(A_j)$. A be called the scheme of μ, $A = S(\mu)$. A relational table TAB be any set of tuples with identical scheme. This common scheme be denoted by $S(TAB)$. Let $\mu[R]$ with $R \subseteq S(\mu)$ denote the restriction of the tuple μ to the scheme R: $\mu[R] = \{(A_j, a_{jk}) | A_j \in R \wedge a_{jk} = \mu(A_j)\}$. The restriction of a relational table TAB to R, denoted $TAB[R]$, be defined $TAB[R] = \{\mu[R] | \mu \in TAB\}$. A relational join of two relational tables TAB_1, TAB_2 be defined as: $TAB_1 \otimes TAB_2 = \{\mu_1 \cup \mu_2 | \mu_1 \in TAB_1, \mu_2 \in TAB_2 \wedge \forall_{A_j \in S(\mu_1) \cap S(\mu_2)} \ \mu_1(A_j) = \mu_2(A_j)\}$.

A decision table is a relational table in which we split the scheme into two distinct parts: the information part **I** and the decision part **D**.

Let $card(SET)$ denote the cardinality of the set SET.

Let us assume that a decision table TAB of decisions D (atomic values) under conditions (information) **I** (atomic value vectors) is available. However, **I** may not contain the complete information to make decision D. This gives rise in a natural way to a mapping Γ assigning different values of D to the same value of **I**. Under these circumstances the belief in a set A (subset of the domain of D) $Bel(A)$ may be derived from a case database as follows: $Bel(A) = 1 - card(\{\mu | \mu \in TAB \wedge \mu(D) \notin A\}) / card(TAB)$, which may be implemented the SQL query language [64] as:

create view Total (Counted) as select count(*) from TAB;
*select count(*)/Counted from TAB,Total where not (***I** *in (select* **I** *from TAB where not (D in A)));*

Skowron and Grzymała-Busse[58] elaborated also a notion of conditioning under rough set interpretation [58, p.219 ff.] as respective measures for subtables (that is tables consisting of cases selected by a criterion). Let us condition on D belonging to the set B selecting tuples fitting the condition B $Subtable = \{\mu | \mu \in TAB \wedge \mu(D) \in B$ which may be implemented as

*create view SubtableTAB(D,***I** *) as select X,***I** *from TAB where not (***I** *in (select* **I** *from TAB where not (D in B)));*

The belief distribution for the subtable can be calculated from the view SubtableTAB in the same way as for the TAB database. However, it is a matter of a calculation exercise to show that their notion of conditionality does not agree with that of Shafer from Def.6.

Therefore, to achieve consistency with Shafer's conditioning from Def.6 we propose the following interpretation, derivable from our approach described in the paper [30]: Let $v \in B$. Then in SQL

update TAB set D=v where not (D in B) and **I** *in (select* **I** *from TAB where D in B);*
delete TAB where not (D in B);

Let c_1,c_2 be two cases from the database TAB such that $\mathbf{I}(c_1) = \mathbf{I}(c_2)$ but $D(c_1) \neq D(c_2)$ Let be $D(c_1) \in B$ and $D(c_2) \notin B$ Then after the above update and delete operations both cases c_1, c_2 are retained in the database TAB. c_1 has retained its value $D(c_1)$. But the case c_2 was subject to a metamorphose: its $D(c_2)$ has been changed (to v). This means that the Dempster rule of combination is "destructive". Preservation of frequency interpretation under conditioning enforces ignoring the intrinsic (observed) value of an attribute and replacement of it with some other value. It should be stressed at this point that the above SQL operation cannot be easily expressed in terms of sets and relations, because an update operation is engaged which may make distinct tuples identical.

A still more complex task is the interpretation of combination of two independent pieces of evidence. Skowron and Grzymała-Busse[58] elaborated a procedure consisting in transforming the combined decision tables into a kind of summary with multivalued decision columns (since non-relational) and then applying complex rational arithmetic to get finally a decision table (derived by so-called Ψ-independent combination) implementing Dempster combination of independent belief functions (consult [58] for details).

3.2 New Interpretation

Below we present a slight modification of the rough set interpretation of belief functions that surprisingly turns out to be both simple, elegant and straight forward and at the same time fulfils the requirement that was not matched by any known interpretations: it is qualitative and not quantitative in nature and still case-based. Furthermore we demonstrate that our new interpretation corresponds strictly to the notion of multivalued dependency, that is combination of belief functions parallels the relational join operator from database theory.

The results of any experiment with multiple outcomes can evaluated along two dimensions: the quantitative and the qualitative one. If we say e.g. that in a series of coin tossing experiments we got 57 heads and 43 tails, then this is a quantitative evaluation. But if we say that there were heads and tails (and not e.g. edges) then we say something about the qualitative aspects of the experiment. In the quantitative evaluation we say that 57 % of all the cases we got heads, in the qualitative evaluation we say that 50 % of all possibilities of diverse outcomes are heads. In most real life cases we are more interested in the quantitative aspects. Sometimes, however, the qualitative side may be of more interest. E.g. if 40 witnesses say that they saw the suspect had killed the victim, but their testimonies are suspiciously similar, and 20 say they saw the contrary, and their testimonies made impressions of individuality, then we would say that there are 1:20 chances of the guilt of the suspect rather than 40:20, because the qualitative aspects (diversity) are more important than quantitative ones (frequency).

Behaviour of frequencies under reasoning was historically the foundation of probability theory. As probabilistic (frequency based) models of MTE fail in general, we considered just the diversity as a possible alternative for a model of MTE. The diversity is well handled by relational model of databases. Though at a first glance the count of cases in a relational database may appear identical with counting frequencies of objects / events, the difference starts as soon as we make a projection on the subset of attributes. It turns out that projected frequencies differ significantly from the counts of cases in a projected relational database. Therefore we say that our approach is non-frequency, non-quantitative, that is qualitative one.

Let us define the plausibility $Pl_{TAB}(SET)$ derived from a decision table TAB with decision variable D and the set **I** of information variables as: $Pl_{TAB}(SET) = card(\{\mu[\mathbf{I}] | \mu \in TAB \wedge \mu(D) \in SET\})/card(TAB[\mathbf{I}])$, implemented as

create view tmpTAB(No) as select count(distinct **I**) *from TAB;*
create view plTAB as select count(distinct **I**)*/No from TAB,tmpTAB where TAB.D in SET;*

Example 1 explains the detailed numerical procedure for calculation of Pl from the above SQL expression.

Theorem 1. *The function $Pl_{TAB}(SET)$ derived from a decision table TAB with decision variable D and the set* **I** *of information variables is plausibility function $Pl(SET)$ in the sense of Dempster-Shafer theory.*

Proof. In DST, the plausibility of a set SET is just the sum of basic probability assignments $m(A)$ such that $SET \cap A \neq \emptyset$. Let r be a record, $I(r)$ its information part and $D(r)$ its decision part. For the set SET, let us consider a subset R of all records from the decision table TAB such that: if $r \in R$ then $D(r) \in SET$ and for every $d \in SET$ there exists $r' \in R$ such that $D(r) = D(r')$ and there exists no record $r" \notin R$ such that $I(r") = I(r)$. Obviously, for two distinct sets SET' and $SET"$ their respective sets R' and $R"$ will share no records. Furthermore, $\bigcup_{SET \subseteq domain(D)} R(SET)$ will be the (relationally) identical with TAB. Hence we can consider the ratio number of records with distinct information part in $R(SET)$ divided by the number of records with distinct information parts in DT as the bpa function $m(SET)$ in the sense of DST. But the function $Pl_{TAB}(SET)$ counts (the relative share of) the records with distinct information part such that the decision part belongs to SET. Hence in practice it is just the sum of basic probability assignments $m(A)$ such that $SET \cap A \neq \emptyset$. Therefore it is a plausibility function.

Example 1. Let us first look at the relational table BUILD in tab. 1). The column D is the decision column, I is information part of the table. The domain of the decision variable D is {center, restaurant, school}.

Table 1. Example: Public offering: erection of buildings of a school, a restaurant and a shopping center. Decision table: BUILDing. "Information part" - the firm, "Decision part" - the object to be erected

	I	D
1.	ABD A.G.	center
2.	LQR Inc.	school
3.	PTS Ltd.	center
	PTS Ltd.	restaurant
4.	XYZ Inc.	center
5.	ZZZ Ltd.	restaurant
	ZZZ Ltd.	school

Let us calculate now the plausibility Pl({school, restaurant}) from this table. There are 7 cases (rows) in the dataset. But there are only 5 cases with distinct information part (firms) **I**. And there are only 3 cases with decision either *school* or *restaurant* with distinct information part **I** (LQR Inc., PTS Ltd., ZZZ Ltd).So the plausibility[4]is equal to Pl({school, restaurant})=3/5. One can check that Pl({school})=2/5 (LQR Inc., ZZZ Ltd) and Pl({restaurant})=2/5 (PTS Ltd., ZZZ Ltd).

Notice that from the calculational rules for Dempster-Shafer theory we can derive also relational views calculating other measures:

[4] Notice that under Skowron/Busse interpretation [58] we get $Pl = 4/7$ which is obviously a different value. The difference stems from the fundamental difference between frequency (Skowron/Busse) and relational(ours) view of the world.

Belief $Bel_{TAB}(SET) = 1 - card(\{\mu[\mathbf{I}] | \mu \in TAB \wedge \mu(D) \notin SET\})/card(TAB[\mathbf{I}])$, implemented as

create view belTAB as select 1-count(distinct **I***)/No from TAB,tmpTAB where not (TAB.D in SET);*

Commonality $Q_{TAB}(SET) = card(\{\mu[\mathbf{I}] | \forall_{d \in SET}\ \mu[\mathbf{I}] \cup \{(D,d)\} \in TAB\}) /card(TAB[\mathbf{I}])$, implemented as

create view tmp1TAB(CN) as select count(distinct D) from TAB where TAB.D in SET group by **I** *;*
create view qTAB as select count(*)/No from tmpTAB,tmp1TAB where CN=card(set); (card() is a function counting the elements of the set passed as its argument)

Basic belief assignment $m_{TAB}(SET) = card(\{\mu[\mathbf{I}] | \forall_{d \in SET}\ \mu[\mathbf{I}] \cup \{(D,d)\} \in TAB\} \wedge \forall_{d \notin SET}\ \mu[\mathbf{I}] \cup \{(D,d)\} \notin TAB\})/card(TAB[\mathbf{I}])$, implemented as

*create view tmp11TAB(***I***,D) as select TAB.***I***, TAB.D+XX.D from TAB, TAB XX where TAB.D in SET and XX.***I***=TAB.***I***;*
*create view tmp12TAB(***I***,CN) as select* **I***,count(distinct D) from tmp11TAB group by* **I***;*
*create view m as count(*)/No from tmpTAB, tmp12TAB where CN=card(SET)*card(SET);*

Theorem 2. *The functions $Bel_{TAB}(SET)$, $Q_{TAB}(SET)$, $m_{TAB}(SET)$ derived from a decision table TAB with decision variable D and the set* **I** *of information variables are belief, commonality, basic probability/belief assignment functions resp. $Bel(SET), Q(SET), m(SET)$ in the sense of Dempster-Shafer theory.*

Example 2. From table 1 we easily calculate that:
Commonality Q({school, restaurant}=1/5 (Number of firms ready to build either the school and the restaurant: ZZZ Ltd).
Belief Q({school, restaurant}=2/5 Number of firms ready to build nothing but the school or the restaurant (LQR Inc., ZZZ Ltd)
bpa - No of firms exactly offering erecting of: m({school, restaurant}=1/5 (ZZZ Ltd), m({restaurant}=0 (none), m({school}=1/5 (LQR Inc.)

Conditioning as Selection of a Subtable Let us define now the conditional belief function $Bel_{TAB}(.||B)$ representing the belief function Bel conditioned on evidence B as the belief function $Bel_{TAB\_B}(.)$ define over the view table

*create view TAB_B(***I***,D) as select* **I***,D from TAB where TAB.D in B;*

Table 2. Shafer's Conditioning. Relational Interpretation. Calculating $Bel(.||\{school, restaurant\})$ as a Bel from the table obtained by *select I,D from BUILD where D=school or D= restaurant*

	I	D
1.	LQR Inc.	school
2.	PTS Ltd.	restaurant
5.	ZZZ Ltd.	restaurant
	ZZZ Ltd.	school

Example 3. In our example, for the relational table BUILD from tab. 1, $Bel_{BUILD}(.||\{school, restaurant\})$ is just Bel calculated from the respective projection *select I,D from BUILD where D=school or D= restaurant* visible in tab. 2. It is easily seen that $Pl_{BUILD}(\{restaurant\}||\{school, restaurant\}) = 2/3$ (PTS Ltd.,ZZZ Ltd.) and $m_{BUILD}(\{restaurant\}||\{school, restaurant\}) = 1/3$ (PTS Ltd.).

Our conditional belief function matches perfectly the Shafer's definition of $Bel(.||B)$ cited above (Def.6).Notice that under Skowron/Busse interpretation (section 3), the matching of Shafer's conditionality definition had to be paid for with creating a physical copy of a relational table and updating it, whereas our notion works perfectly without any updates - only selection is used just as in probabilistic conditioning.

Table 3. Public offering: equipment for buildings of a school, a restaurant and a shopping center. Decision table: EQUIPment "Information part" - the firm, "Decision part" - the object to be equipped

	I2	D
1.	AAA GmbH	school
2.	BBB Ltd.	center
	BBB Ltd.	restaurant
3.	CCC Inc.	center
	CCC Inc.	restaurant

Combination as Relational Join Now let us discuss the most impressive property of the new interpretation: the Dempster's rule of combination interpreted as relational join.

Example 4. Let us consider the decision table EQUIP (tab. 3) and BUILD (tab. 1). We want to combine independent evidence from both tables to support a decision. Let us assume that independence of evidence means that no pair of firms (one from BUILD, one from EQUIP) refuse to cooperate on erecting and equipping an object. How many pairs of firms do we have to finish a set of objects mentioned in the offerings ? The answer lies in the relational table FINISH (tab. 4) obtained as a relational join of BUILD and EQUIP (over the common column D) so that the new decision table has as its decision column D and as its information part I,I2:

create table FINISH (I,I2,D);
insert into table FINISH from
select I,I2,D from BUILD, EQUIP where BUILD.D=EQUIP.D;

Notice that in BUILD, there were 5 cases with distinct information part, in EQUIP - 3, and in BUEQ there are only 10. We have here $Pl_{FINISH}(\{school, restaurant\}) = 8/10$ and $Bel_{FINISH}(\{school, restaurant\}) = 3/10$

You can easily check that $Bel_{FINISH} = Bel_{BUILD} \oplus Bel_{EQUIP}$.

Generally, we can formulate the theorem:

Theorem 3. *If the decision tables DT1(**I1**,D) and DT2(**I2**,D) with non-overlapping information parts **I1**,**I2** are combined by relational join operation $DT1 \otimes DT2$, implemented as*

*select **I1**,**I2**,DT1.D from DT1,DT2 where DT1.D=DT2.D;*

*yielding table DT12(**I**,D) with **I**=**I1**∪**I2**, then $Bel_{DT12} = Bel_{DT1} \oplus Bel_{DT2}$.*

Proof. This can be demonstrated by considering the "fate" of records counted on calculation of m_i. If R_1 is the set of records counted when calculating $m_1(A)$ in DT1, and if R_2 is the set of records counted when calculating $m_2(B)$ in DT2, then upon join only records $\mu = (\mu_1[\mathbf{I1}], \mu_2[\mathbf{I2}], \mu_1[D])$ with $\mu_1 \in R_1$, $\mu_2 \in R_2$, $\mu_1[D] = \mu_2[D] \in A \cap B$ will be created, hence they will be counted in support of $m(A \cap B)$. Furthermore, their number will be exactly equal to the product of the number of distinct records in R_1 times the number of distinct records in R_2, so that the Dempster formula will be matched perfectly upon normalization.

Relational Marginalization and Decombination A further intriguing property, not present in any interpretation of MTE known so far, is the relationship between relational marginalization and MTE factorization ("decombination") of belief functions.

Example 5. Notice that BUILD and EQUIP in our example are both in first normal form and the domain of the attribute D is identical in both tables. Therefore we know from elementary properties of relational data tables that marginalization of FINISH over I,D

Table 4. Combination of Independent Evidence. Decision table FINISH obtained as *select I,I2,D from BUILD, EQUIP where BUILD.D=EQUIP.D*

	I	I2	D
1.	ABD A.G.	BBB Ltd.	center
2.	ABD A.G.	CCC Inc.	center
3.	LQR Inc.	AAA GmbH	school
4.	PTS Ltd.	BBB Ltd.	center
	PTS Ltd.	BBB Ltd.	restaurant
5.	PTS Ltd.	CCC Inc.	center
	PTS Ltd.	CCC Inc.	restaurant
6.	XYZ Inc.	BBB Ltd.	center
7.	XYZ Inc.	CCC Inc.	center
8.	ZZZ Ltd.	BBB Ltd.	restaurant
9.	ZZZ Ltd.	CCC Inc.	restaurant
10.	ZZZ Ltd.	AAA GmbH	school

select distinct I,D from FINISH;

is exactly identical with BUILD. and marginalization of FINISH over I2,D

select distinct I2,D from FINISH;

is exactly identical with EQUIP.

In general:

Theorem 4. *If the information part* **I** *of the decision table DT(***I***,D) can be split into two such parts* **I1**,**I2** *that* $\mathbf{I1} \cup \mathbf{I2} = \mathbf{I}$ *and* $\mathbf{I1} \cap \mathbf{I2} = \emptyset$ *and the relation DT is identical with DT1⊗DT2, implemented*

select **I1**,**I2**,*DT1.D from DT1,DT2 where DT1.D=DT2.D;*

*where DT1 and DT2 are DT1=DT[***I1***,D], DT2=DT[***I2***,D], implemented*

create view DT1 as select distinct **I1**,*D from DT;*
create view DT2 as select distinct **I2**,*D from DT;*

that is there is a multivariate dependency between **I1** *and* **I2** *given D, then* $Bel_{DT} = Bel_{DT^{\downarrow \mathbf{I1},D}} \oplus Bel_{DT^{\downarrow \mathbf{I2},D}}$.

Proof. Follows directly from theorem 3.

Let consider the unnormalized DST measures of decision tables m'_{TAB}, Bel'_{TAB}, Pl'_{TAB}, Q'_{TAB}, such that $f'_{TAB} = f_{TAB} \cdot card(TAB^{\downarrow I})$ (card - number of distinct rows, f - m or Bel or Pl or Q) and the unnormalized combination operator $\oplus'$ such that $Bel'_{E_1,E_2} = Bel'_{E_1} \oplus' Bel'_{E_2}$ is defined as follows: $m'_{E_1,E_2}(A) = \sum_{B,C;A=B\cap C} m'_{E_1}(B) \cdot m'_{E_2}(C)$.

What may be more surprising, a kind of a reverse theorem holds:

Theorem 5. *The information part I of the decision table DT(I,D) can be split into two such parts* **I1**,**I2** *that* $\mathbf{I1} \cup \mathbf{I2} = I$ *and* $\mathbf{I1} \cap \mathbf{I2} = \emptyset$ *and* $Bel'_{DT} = Bel'_{DT^{\downarrow \mathbf{I1},D}} \oplus' Bel'_{DT^{\downarrow \mathbf{I2},D}}$ *if and only if the relation DT is identical with DT1⊗DT2, implemented*

select **I1**,**I2**,*DT1.D from DT1,DT2 where DT1.D=DT2.D;*

*where DT1 and DT2 are DT1=DT[*__I1__*,D], DT2=DT[*__I2__*,D], implemented*

create view DT1 as select distinct **I1**,*D from DT;*
create view DT2 as select distinct **I2**,*D from DT;*

that is there is a multivariate dependency between **I1** *and* **I2** *given D,*

Proof. (An outline.) The if-part parallels exactly theorem 4. We need only pay attention to the fact that we never normalize.
The only-if-part follows from numerical calculations for Q-values of all the tables considered. If a record is counted in DT when calculating $Q'_{DT}(SET)$, then it is also counted when calculating both $Q'_{DT1}(SET)$ and $Q'_{DT2}(SET)$. If we form a join $DT1 \cdot DT2$ then $Q'_{DT1 \cdot DT2}(SET) = Q'_{DT1}(SET) \cdot Q'_{DT2}(SET)$. And this is the maximum value Q' can take in DT12. If there is ANY deviation from multivariate dependency concerning records with decision part in SET, then value of $Q'_{DT12}(SET)$ is smaller than $Q'_{DT1}(SET) \cdot Q'_{DT2}(SET)$. This proves our claim.

Remak: We can conclude that Dempster's rule of combination is equivalent with relational join and the Dempster-Shafer independence of evidence means multivalued dependence of evidence. We can also simulate other rules of combination of evidence. In the above, we assumed that given the decision, we cannot conclude from the information part **I1** the value of the information part **I2** in DT. This meant qualitative independence. Now let us assume the contrary in another decision table DT': given the decision d, we can totally predict **I2** from **I1** for all records r with D(r)=d in DT' or we can totally predict **I1** from **I2** for all records r with D(r)=d in DT'. It is immediately clear that in this case for any set of decisions the unnormalized plausibility is calculated as $Pl'_{DT'}(A) = max(Pl'_{DT'^{\downarrow \mathbf{I1},D}}(A), Pl'_{DT'^{\downarrow \mathbf{I2},D}}(A))$. We can conclude for normalized plausibility that we deal here with the know rule of combination of dependent evidence: $Pl'_{DT'}(A) = max(\alpha \cdot Pl'_{DT'^{\downarrow \mathbf{I1},D}}(A), (1-\alpha) \cdot Pl'_{DT'^{\downarrow \mathbf{I2},D}}(A))$ where α ranges from 0 to 1 (depending on proportions between the numbers of distinct information parts **I1** and **I2**).

Multivariate Beliefs and Multidecision Tables We can extend our consideration to tables with multiple decision variables. In a straight forward way we can extend our definition of DST measures to such tables and consider multivariate belief distributions (in all the decision variables). It is trivial to see that dropping a decision variables does not diminish the diversity of

Table 5. Multivariate MTE. The table MADEOF and its projection *select distinct I,D from MADEOF*

Decision table MADEOF

	I	D	D2
1.	ABD A.G.	center	wooden
2.	LQR Inc.	school	stone
	LQR Inc.	school	wooden
3.	PTS Ltd.	center	stone
	PTS Ltd.	center	wooden
	PTS Ltd.	restaurant	stone
4.	XYZ Inc.	center	stone
5.	ZZZ Ltd.	restaurant	stone
	ZZZ Ltd.	restaurant	wooden
	ZZZ Ltd.	school	wooden

m({(center,wooden)})=1/5
m({(center,stone)})=1/5

Projected onto I,D:

	I	D
1.	ABD A.G.	center
2.	LQR Inc.	school
3.	PTS Ltd.	center
	PTS Ltd.	restaurant
4.	XYZ Inc.	center
5.	ZZZ Ltd.	restaurant
	ZZZ Ltd.	school

$m^{\downarrow D}$({(center)})=
m({(center,stone)})+
m({(center,wooden)})=2/ 5

Table 6. Variable Independence

D3 - heating

	I2	I3	D	D3
1.	AAA GmbH	EC	school	electric
2.	AAA GmbH	GC	school	gas
3.	BBB Ltd.	EC	center	electric
	BBB Ltd.	EC	restaurant	electric
4.	BBB Ltd.	GC	center	gas
	BBB Ltd.	GC	restaurant	gas
5.	CCC Inc.	EC	center	electric
	CCC Inc.	EC	restaurant	electric
6.	CCC Inc.	GC	center	gas
	CCC Inc.	GC	restaurant	gas

$$Bel = Bel^{\downarrow D} \oplus Bel^{\downarrow D2}$$

because the above table represents a cross product of the tables (without common columns)

	I2	D
1.	AAA GmbH	school
2.	BBB Ltd.	center
	BBB Ltd.	restaurant
3.	CCC Inc.	center
	CCC Inc.	restaurant

and

	I3	D3
1.	EC	electric
2.	GC	gas

Table 7. Conditional Variable Independence

I4 - painting company, D4 - color, D5 - finish,

	I2	I4	D	D5	D4
1.	AAA GmbH	Messer	school	wood	green
	AAA GmbH	Messer	school	wood	red
	AAA GmbH	Messer	school	plastic	green
	AAA GmbH	Messer	school	plastic	red
2.	BBB Ltd.	Messer	center	metallic	white
	BBB Ltd.	Messer	center	metallic	yellow
	BBB Ltd.	Messer	center	marble	white
	BBB Ltd.	Messer	center	marble	yellow
3.	BBB Ltd.	Gabel	restaurant	wood	red
4.	CCC Inc.	Messer	center	metallic	white
	CCC Inc.	Messer	center	metallic	yellow
	CCC Inc.	Messer	center	laminated	white
	CCC Inc.	Messer	center	laminated	yellow
5.	CCC Inc.	Gabel	restaurant	plastic	red

In Bel of the above table variables D4 and D5 are conditionally independent given D in the sense of Shenoy's valuation-based systems because the above table is a relational join of the tables below (with D as a common column)

	I2	D	D5
1.	AAA GmbH	school	wood
	AAA GmbH	school	plastic
2.	BBB Ltd.	center	metallic
	BBB Ltd.	center	marble
	BBB Ltd.	restaurant	wood
3.	CCC Inc.	center	metallic
	CCC Inc.	center	laminated
	CCC Inc.	restaurant	plastic

&

	I4	D	D4
1.	Gabel	restaurant	red
2.	Messer	school	green
	Messer	school	red
	Messer	center	white
	Messer	center	yellow

the information part. Hence dropping a decision variable Di from the set of decision variables **D** has the same effect as dropping a variable in the belief function. That is for any set B of decision vectors in variables **D**-{Di}: $m_{TAB^{\downarrow \mathbf{D}-\{Di\}}}(B) = m_{TAB}^{\downarrow \mathbf{D}-\{Di\}}(B) = c \cdot \sum_{A;B=A^{\downarrow \mathbf{D}-\{Di\}}} m(A)_{TAB}$ (c - normalizing factor) See table 5 for an example.

The operator of projection $\downarrow$ should be understood as the DST projection operator applied to a belief function.

Let DTM be a decision table with decision variables D1 and D2. Let the information part consist of two disjoint parts I1 and I2. Let us consider the following views:

create view DTM1 (I1,D1) as select distinct I1,D1 from DTM;
create view DTM2 (I2,D2) as select distinct I2,D2 from DTM;
create view DTM12 as select distinct I1,I2,D1,D2 from DTM1,DTM2;

If now the table DTM12 is relationally identical with DTM, then we shall say that the decision variables D1 and D2 are independent in the decision table DTM. It is not surprising that: $Bel_{DTM} = Bel_{DTM}^{\downarrow D1} \oplus Bel_{DTM}^{\downarrow D2}$. This means that independence of decision variables in a decision table implies independence of variables in the corresponding belief function. See table 6 for an example.

Let DTX be a decision table with decision variables D1, D2 and D3. Let the information part consist of two disjoint parts I1 and I2. Let us consider the following views:

create view DTX1 (I1,D1,D3) as select distinct I1,D1,D3 from DTX;
create view DTX2 (I2,D2,D3) as select distinct I2,D2,D3 from DTX;
create view DTX12 as select distinct I1,I2,D1,D2,D3 from DTX1,DTX2 where DTX1.D3=DTX2.D3;

If now the table DTX12 is relationally identical with DTX, then we shall say that the decision variables D1 and D2 are independent given D3 in the decision table DTX. It is not surprising that: $Bel_{DTX} = Bel_{DTX}^{\downarrow D1,D3} \oplus Bel_{DTX}^{\downarrow D2,D3}$. But this means that the variables D1 and D2 are independent given D3 in the belief function Bel_{DTX} in the sense of Shenoy's VBS [54]. See table 7 for an example.

These results mean that analysis of independence and conditional independence of variables in a belief function corresponding to a decision table may serve as an indicator of presence or absence of independence or multi-valued dependence in the decision table.

3.3 Summary of Qualitative Approach

A novel case-based interpretation of DST belief functions which is qualitative in nature and has the potential to represent a number of DST operations has been presented. The interpretation is based on rough sets (in connection with decision tables), but differs from previous interpretations of this type e.g. [58] in that it counts the diversity rather than frequencies in the decision table. The interpretation has the property that given a definition of the DST measure of objects in the interpretation domain (decision table) we can perform operations in the interpretation domain (e.g. combining decision tables) and the measure of the resulting object is derivable from measures of component objects via DST operator (e.g. combination). We demonstrated this property for Dempster rule of combination, marginalization, Shafer's conditioning, independent variables, Shenoy's notion of conditional independence of variables. Other known case-based (frequency or probabilistic) interpretations fall short of this property. E.g. in [58] complex rational number arithmetic, unnatural for decision tables, is needed to achieve compatibility of the final decision table with Dempster rule of combination. In [34] (probabilistic interpretation) only lower and upper bounds are found for Dempster rule. In [14] (probability structures interpretation) the belief function obtained from the Dempster rule is a potential, but not necessary result of the corresponding probabilistic structure combination operation. See also [59] for discussion of other interpretations.

As probabilistic (frequency based) models of MTE seem to fail in general, we looked for alternatives. The result of any experiment with multiple outcomes can evaluated along two dimensions: the quantitative and the qualitative one. In most real life cases we are more interested in the quantitative aspects. Sometimes, however, the qualitative side may be of more interest. E.g. legal applications we would treat suspiciously similar testimonies as a single argument in favour or against a hypothesis: we would rely on the count of diversity of arguments rather than on their actual counts. And MTE was claimed to be applicable just in legal reasoning [53]. Therefore we considered just the diversity as a possible alternative for a model of MTE and this idea turned out to very fruitful. The diversity is well handled by relational model of databases. Though at a first glance the count of cases in a relational database may appear identical with counting frequencies of objects / events, the difference starts as soon as we make a projection on the subset of attributes. It turns out that projected frequencies differ significantly from the counts of cases in a projected relational database. Therefore we say that our approach is non-frequency, non-quantitative, that is qualitative one.

The new interpretation may be directly applied in the domain of multiple decision decision tables: independence of decision variables or Shenoy's conditional independence in the sense of DST may serve as an indication of possibility of decomposition of the decision table into smaller but equivalent tables. Furthermore it may be applied in the area of Cooperative Query Answering [44]. The problem there is that a query posed to a local relational database system may contain an unknown attribute. But possibly other co-operating db systems know it and may explain it to the queried system in terms of known attributes, shared by the various systems. The uncertainties studied in the decision tables arise here in a natural way and our interpretation may be used to measure these uncertainties in terms of DST (as diversity of support). Furthermore, if several co-operating systems respond, then the queried system may calculate the overall uncertainty measure using DST combination of measures of individual responses.

Now we can ask how to understand then a DST belief function in the light of our experience. One possibility is to consider the belief function as a measure of diversity of support. This is an obvious departure from frequency interpretations proposed by Shafer and others. No mater how frequently the same piece of evidence is presented, it is counted once. This insight may encourage to revise other known interpretations of DST. In the "legal" interpretations e.g. [53], the witnesses in favor of a hypotheses should be counted separately, if their statements differ in unimportant details permitting to deduce that their statements are personal and not studied in. In the "probability of provability" approach [39] not a probability of correctness, but rather the number of distinct valid proofs of a statement should be counted. In the "possible world semantics" [46] the worlds should not be assigned a probability, but rather distinct possible worlds should be counted that differ in non-essential details. Then the operation of combination of independent evidence in the "legal" interpretation is just mixing compatible statements of two sets of witnesses (which saw the same event from different perspective) and counting different possible combinations. In the "probability of provability" approach the combination would mean putting together conclusion compatible proofs stemming from distinct domains (e.g. macro and microphysical observations) and counting legal combinations. In the possible worlds semantics we may combine worlds spanned over disjoined sets of dimensions.

Please notice also that the rough-set interpretation sheds some light onto what the concept of "evidence" may mean. The "evidence" are just different sets of information attributes and the "independence" means (deterministic) unpredictability of attribute values of the one set from the other set. This should not be confused with predictability of the decision variable. Nor with stochastic predictability which may be present. In the "legal" interpretation, independence would be measured with non-predictability of insignificant details. In the "provability" interpretation the independence may be measured by mutual non-derivability of the sets of underlying axioms. In the possible world semantics by possibility of putting together projections onto separated sets of dimension axes.

Further studies on interpreting other known DST operators in the spirit of qualitative interpretation presented in this section are needed and may reveal new potential applications of DST to real world problems.

4 Quantitative Interpretation

4.1 Introduction

The new interpretation of DST runs informally as follows: It is assumed that each object of a population is described by a set-valued attribute (with non-empty set as its value) and a set-valued label (having non-empty intersection with the attribute). Intuitively, the attribute represents objective observations, the label means the part of potential observations which one is ready to accept. Whenever belief (or mass or plausibility function) is calculated from data, it refers to the intersection of the label and the attribute value (and not to attribute value alone). Combination of evidence means a combined selection of a subpopulation and change (reduction) of the label of an object. Thus Dempster rule of combination is understood as data modification.

We introduce the "measurement function" M() (Def.7) which follows the idea of "lazy" observation of value of the object. It asks if any value from a set of values is the potential value of the object. It may be easier (cheaper) in practice to obtain an answer to the question about a set of values ("Are you busy ?") than to interrogate separately for each individual value ("Are you doing your lessons in mathematics?", "Are you washing your hands ?"

etc.) and the answer may be still sufficient for one's purposes (I want someone to go for a walk with me). The way from the M() function to Shafer's frequency interpretation [52] (Γ-function) is straight forward. But many authors [35,18] complained about Shafer's frequency interpretation due to its difficulties with Dempster's rule of combination. Therefore we introduce a generalization of Shafer's concept by adding the notion of "labeling" Def.8 and introduce the modified measurement function $M_l()$ Def.9. Using this notion we define frequency-based m, Bel and Pl functions in Def.10 Please notice that Def.10 relates belief and plausibility functions in a simple way to "lazy measurement". We show that the simple and general labeling processes (Def.11, Def.13) match the behaviour of the Dempster's rule of combination in terms of frequencies within the realm of frequency-based m, Bel and Pl functions under modified measurement of object values (see Theorem 8, Theorem 10).

Let us assume that we know that objects of a population can be described by an intrinsic attribute X taking exclusively one of the n discrete values from its domain $\Xi = \{v_1, v_2, ..., v_n\}$. Let us assume furthermore that to obtain knowledge of the actual value taken by an object we must apply a measurement method (a system of tests) M

Definition 7. X be a set-valued attribute taking as its values non-empty subsets of a finite domain Ξ. By a measurement method of value of the attribute X we understand a function:

$$M : \Omega \times 2^{\Xi} \rightarrow \{TRUE, FALSE\}$$

where Ω is the set of objects, (or population of objects) such that

- $\forall_{\omega;\omega\in\Omega} \quad M(\omega, \Xi) = TRUE$ (X takes at least one of values from Ξ)
- $\forall_{\omega;\omega\in\Omega} \quad M(\omega, \emptyset) = FALSE$
- whenever $M(\omega, A) = TRUE$ for $\omega \in \Omega$, $A \subseteq \Xi$ then for any B such that $A \subset B$ $M(\omega, B) = TRUE$ holds,
- whenever $M(\omega, A) = TRUE$ for $\omega \in \Omega$, $A \subseteq \Xi$ and if $card(A) > 1$ then there exists B, $B \subset A$ such that $M(\omega, B) = TRUE$ holds.
- for every ω and every A either $M(\omega, A) = TRUE$ or $M(\omega, A) = FALSE$ (but never both).

$M(\omega, A)$ tells us whether or not any of the elements of the set A belong to the actual value of the attribute X for the object ω.

The measuring function $M(\omega, A)$, if it takes the value TRUE, states for an object ω and a set A of values from the domain of X that the X takes for this object (at least) one of the values in A.

With each application of the measurement procedure some costs be connected, increasing roughly with the decreasing size of the tested set A so that we are ready to accept results of previous measurements in the form of pre-labeling of the population. So

Definition 8. A *label* L of an object $\omega \in \Omega$ is a subset of the domain Ξ of the attribute X.
A *labeling* under the measurement method M is a function $l : \Omega \rightarrow 2^{\Xi}$ such that for any object $\omega \in \Omega$ either $l(\omega) = \emptyset$ or $M(\omega, l(\omega)) = TRUE$.
Each *labelled object* (under the labeling l) consists of a pair (ω, L_ω), ω - the object, $L_\omega = l(\omega)$ - its label.
By a *population under the labeling* l we understand the predicate $P : \Omega \rightarrow \{TRUE, FALSE\}$ of the form $P(\omega) = TRUE \ iff \ l(\omega) \neq \emptyset$ (or alternatively, the set of objects for which this predicate is true)
If for every object of the population the label is equal to Ξ then we talk of an *unlabeled population* (under the labeling l), otherwise of a *pre-labelled* one.

Let us assume that in practice we apply a modified measurement method M_l being a function:

Definition 9. Let l be a labeling under the measurement method M. Let us consider the population under this labeling. The modified measurement method

$$M_l : \Omega \times 2^{\Xi} \rightarrow \{TRUE, FALSE\}$$

where Ω is the set of objects, is defined as

$$M_l(\omega, A) = M(\omega, A \cap l(\omega))$$

(Notice that $M_l(\omega, A) = FALSE$ whenever $A \cap l(\omega) = \emptyset$.)

For a labeled object (ω, L_ω) (ω - object, L_ω - its label) and a set A of values from the domain of X, the modified measurement method tells us that X takes one of the values in A if and only if it takes in fact a value from intersection of A and L_ω. Expressed differently, we discard a priori any value not in the label.

Please pay attention also to the fact, that given a population P for which the measurement method M is defined, the labeling l (according to its definition) selects a subset of this population, possibly a proper subset, namely the population P' under this labeling. $P'(\omega) = P(\omega) \wedge M(\omega, l(\omega))$. Hence also M_l is defined possibly for the "smaller" population P' than M is.

Let us define the following functions referred to as labelled Belief, labelled Plausibility and labelled Mass Functions respectively for the labeled population P: The predicate $\underset{\omega}{\mathfrak{Prob}}^{P(\omega)} \alpha(\omega)$ shall denote the probability of truth of expression $\alpha(\omega)$ over ω given population $P(\omega)$.

Definition 10. Let P be a population and l its labeling. Then

$$Bel_P^{M_l}(A) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)} \neg M_l(\omega, \Xi - A)$$

$$Pl_P^{M_l}(A) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)} M_l(\omega, A)$$

$$m_P^{M_l}(A) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)}\left(\bigwedge_{v;v\in A} M_l(\omega,\{v\}) \wedge \bigwedge_{v;v\in \Xi - A} \neg M_l(\omega,\{v\})\right)$$

$$Q_P^{M_l}(A) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)}\left(\bigwedge_{v;v\in A} M_l(\omega,\{v\})\right)$$

Theorem 6. *$m_P^{M_l}$, $Bel_P^{M_l}$, $Pl_P^{M_l}$ and $Q_P^{M_l}$ are the mass, belief, plausibility and commonality Functions in the sense of DST.*

Proofs of all the theorems are given in [32] Let us now assume we run a "(re-)labelling process" on the (pre-labelled or unlabeled) population P.

Definition 11. Let M be a measurement method, l be a labeling under this measurement method, and P be a population under this labeling. The *(simple) labelling process* on the population P is defined as a functional $LP : 2^{\Xi} \times \Lambda \rightarrow \Lambda$, where Λ is the set of all possible labelings under M, such that for the given labeling l and a given nonempty set of attribute values L ($L \subseteq \Xi$), it delivers a new labeling l' ($l' = LP(L,l)$) such that for every object $\omega \in \Omega$:

1. if $M_l(\omega, L) = FALSE$ then $l'(\omega) = \emptyset$

(that is l' discards a labeled object $(\omega, l(\omega))$ if $M_l(\omega, L) = FALSE$

2. otherwise $l'(\omega) = l(\omega) \cap L$ (that is l' labels the object with $l(\omega) \cap L$ otherwise.

Remark: It is immediately obvious, that the population obtained as the sample fulfills the requirements of the definition of a labeled population.

The labeling process clearly induces from P another population P' (a population under the labeling l') being a subset of P (hence perhaps "smaller" than P) labelled a bit differently. If we retain the primary measurement method M then a new modified measurement method $M_{l'}$ is induced by the new labeling.

Definition 12. "labelling process function" $m^{LP;L} : 2^{\Xi} \rightarrow [0,1]$: is defined as:

$$m^{LP;L}(L) = 1$$

$$\forall_{B;B\in 2^{\Xi}, B\neq L}\, m^{LP;L}(B) = 0$$

It is immediately obvious that:

Theorem 7. *$m^{LP;L}$ is a Mass Function in sense of DST.*

Let $Bel^{LP,L}$ be the belief and $Pl^{LP,L}$ be the Plausibility corresponding to $m^{LP,L}$. Now let us pose the question: what is the relationship between $Bel_{P'}^{M_{l'}}$, $Bel_P^{M_l}$, and $Bel^{LP,L}$.

Theorem 8. *Let M be a measurement function, l a labeling, P a population under this labeling. Let L be a subset of Ξ. Let LP be a labeling process and let $l' = LP(L, l)$. Let P' be a population under the labeling l'. Then $Bel_{P'}^{M_{l'}}$ is a combination via Dempster's Combination rule of Bel^{M_l}, and $Bel^{LP;L}$., that is:*

$$Bel_{P'}^{M_{l'}} = Bel_P^{M_l} \oplus Bel^{LP;L}$$

Let us define a more general (re-)labeling process. Instead of a single set of attribute values let us take a set of sets of attribute values $L^1, L^2, ..., L^k$ (not necessarily disjoint) and assign to each one a probability $m^{LP,L^1,L^2,...,L^k}(A_i)$ of selection.

Definition 13. Let M be a measurement method, l be a labeling under this measurement method, and P be a population under this labeling (Note that the population may also be unlabeled). Let us take a set of (not necessarily disjoint) nonempty sets of attribute values $\{L^1, L^2, ..., L^k\}$ and let us define the probability of selection as a function $m^{LP,L^1,L^2,...,L^k} : 2^{\Xi} \to [0, 1]$ such that

$$\sum_{A;A\subseteq\Xi} m^{LP,L^1,L^2,...,L^k}(A) = 1$$

$$\forall_{A;A\in\{L^1,L^2,...,L^k\}} m^{LP,L^1,L^2,...,L^k}(A) > 0$$

$$\forall_{A;A\notin\{L^1,L^2,...,L^k\}} m^{LP,L^1,L^2,...,L^k}(A) = 0$$

The *(general) labelling process* on the population P is defined as a (randomized) functional $LP : 2^{2^{\Xi}} \times \Phi \times \Lambda \to \Lambda$, where Λ is the set of all possible labelings under M, and Φ is a set of all possible probability of selection functions, such that for the given labeling l and a given set of (not necessarily disjoint) nonempty sets of attribute values $\{L^1, L^2, ..., L^k\}$ and a given probability of selection $m^{LP,L^1,L^2,...,L^k}$ it delivers a new labeling l" such that for every object $\omega \in \Omega$:

1. a label L, element of the set $\{L^1, L^2, ..., L^k\}$ is sampled randomly according to the probability distribution $m^{LP,L^1,L^2,...,L^k}$; This sampling is done independently for each individual object,

2. if $M_l(\omega, L) = FALSE$ then $l"(\omega) = \emptyset$
(that is l" discards an object $(\omega, l(\omega))$ if $M_l(\omega, L) = FALSE$

3. otherwise $l"(\omega) = l(\omega) \cap L$ (that is l" labels the object with $l(\omega) \cap L$ otherwise.)

Again we obtain another ("smaller") population P" under the labeling l" labelled differently. Also a new modified measurement method $M_{l"}$ is induced by the "re-labelled" population. Please notice, that l" is not derived deterministicly. Another run of the general (re-)labeling process LP may result in a different final labeling of the population and hence a different subpopulation under this new labeling.

Theorem 9. $m^{LP,L^1,\ldots,L^k}$ *is a Mass Function in sense of DST.*

Let $Bel^{LP;L^1,\ldots,L^k}$ be the belief and $Pl^{LP,L^1,\ldots,L^k}$ be the Plausibility corresponding to $m^{LP,L^1,\ldots,L^k}$. Now let us pose the question: what is the relationship between $Bel_{P"}^{M_{l"}}$, $Bel_P^{M_l}$, and $Bel^{LP,L^1,\ldots,L^k}$.

Theorem 10. *Let* M *be a measurement function,* l *a labeling,* P *a population under this labeling. Let* LP *be a generalized labeling process and let* $l"$ *be the result of application of the* LP *for the set of labels from the set* $\{L^1, L^2, ..., L^k\}$ *sampled randomly according to the probability distribution* $m^{LP,L^1,L^2,\ldots,L^k}$*;. Let* $P"$ *be a population under the labeling* $l"$*. Then the expected value over the set of all possible resultant labelings* $l"$ *(and hence populations* $P"$*) (or, more precisely, value vector) of* $Bel_{P"}^{M_{l"}}$ *is a combination via Dempster's Combination rule of* $Bel_P^{M_l}$*, and* $Bel^{LP,L^1,\ldots,L^k}$*., that is:*

$$E(Bel_{P"}^{M_l'}) = Bel_P^{M_l} \oplus Bel^{LP,L^1,\ldots,L^k}$$

4.2 Independence and shades of conditionality

Let us turn now to multivariable setting of DST. As we defined combination rule of belief function in terms of frequencies, we may be tempted to define independence of belief functions (and conditional independence) following the remark of Shafer [51] that independence in DST shall be "understood traditionally". However, several peculiarities emerge. First, from probability theory, we associate independence of two variables X,Y both with unability to predict X,Y from one another and to represent joint distribution as a combination of marginal distributions. However, if, in DST, we parallel the first concept (mutual non-predictability) as in Def.14, we do not necessarily arrive at the second one (joint distribution from marginals) - see Theorem 11,Theorem 12. Hence, we need a stronger concept of independence, rooted in composite measurement of variables Def.15, which can lead then to decomposability of joint belief distribution see Theorem 13. A similar complexity emerges if we want to parallel the concept of a constant variable from probability. Given definition of independence of a distribution from a variable X from Def.16, we see from Theorem 15 and the subsequent ones that only the assumption of composite measurement of variables justifies dropping such a variable X from consideration.
If one turns to conditional independence, one has again to do with splitting of familiar probabilistic concepts: In probability, for the joint probability distribution $P(X,Y)$, distributions $f(X;Y)$ of X given a value of variable Y mean exactly the same as a function $g(X,Y)$ such that $P(X,Y) = g(X,Y) \cdot P(Y)$ that is $f(X,;Y) = g(X,Y) = P(X|Y)$. But in DST, Shafer's conditional belief function $Bel(||Y) = Bel \oplus Bel_Y$ does not have the property that $Bel = Bel(||Y) \oplus Bel^{\downarrow X}$, so that two different concepts of conditionality

emerge: the Shafer's conditional belief as defined in Def.6 and the anticonditional belief function as defined in Def.18, introduced in this paper. If X,Y anre independent given Z then $P(X|Z,Y) = P(X|Z)$. This does hold for DST only existentially as described in Theorem 17 and Theorem 18. Shafer's conditional belief is the result of reasoning process - it appears "a posteriori"; whereas anticonditional belief function serves the purpose of decomposition (factorization) of a joint belief distribution - it is used "a priori".

Let us introduce the notion of quantitative independence for DS-Theory. We will fix the measurement method M we use and the population P we consider so that respective indices will be usually dropped.

Definition 14. Two disjoint sets of variables p, q are (mutually, marginally) independent when for objects of the population knowledge of the truth value of $M_l^{\downarrow p}(\omega, A^{\downarrow p})$ for all $A \subseteq \Xi_p \times \Xi_q$ does not change our prediction capability of the values of $M_l^{\downarrow q}(\omega, B^{\downarrow q})$ for any $B \subseteq \Xi_p \times \Xi_q$, that is

$$\underset{\omega}{\mathfrak{Prob}}^{P(\omega)}(M_l^{\downarrow q}(\omega, B^{\downarrow q}) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)}(M_l^{\downarrow q}(\omega, B^{\downarrow q})|M_l^{\downarrow p}(\omega, A^{\downarrow p}))$$

Theorem 11. *If sets of variables p, q are quantitatively independent, then for any $C \subseteq \Xi_q$, $A \subseteq \Xi_p$*

$$m^{\downarrow q}(C) \cdot m^{\downarrow p}(A) = \sum_{F; F^{\downarrow p} = A, F^{\downarrow q} = C} m(F)$$

Theorem 12. *If sets of variables p, q are quantitatively independent, then for any $C \subseteq \Xi_q$, $A \subseteq \Xi_p$*

$$Bel^{\downarrow q}(C) \cdot Bel^{\downarrow p}(A) = Bel(A \times C)$$

This actually expresses the relationship between marginals of two independent variables and their joint belief distribution. This relationship has one dismaying aspect: in general, we cannot calculate the joint distribution from independent marginals (contrary to our intuition connected with joint probability distribution).

In practical settings, however, we frequently have to do with some kind of composite measurement method, that is:

Definition 15. Two disjoint sets of variables p, q are measured compositely iff for $A \subseteq \Xi_p, C \subseteq \Xi_q$:

$$M(\omega, A \times C) = M(\omega, A \times \Xi_q) \wedge M(\omega, \Xi_p \times C)$$

and for $M(\omega, B)$, $B \subseteq \Xi_p \times \Xi_q$:

$$M(\omega, B) = \bigvee_{A,C; A \subseteq \Xi_p, C \subseteq \Xi_q, A \times C \subseteq B} M(\omega, A \times C)$$

Under these circumstances, it is easily shown that whenever $m(B) > 0$, then there exist A and C such that: $B = A \times C$.

So we obtain:

Theorem 13. *If sets of variables p, q are quantitatively independent and measured compositely, then*

$$m(A \times C) = m^{\downarrow p}(A) \cdot m^{\downarrow q}(C)$$

Hence the Belief function can be calculated from Belief functions of independent variables under composite measurement.:

Theorem 14. *If sets of variables p, q are quantitatively independent and measured compositely, then*

$$Bel = Bel^{\downarrow p} \oplus Bel^{\downarrow q}$$

Let us justify now the notion of empty extension:

Definition 16. The joint distribution over set of variables $X = p \cup q$ with sets p, q disjoint is independent of the variable set p when for objects of the population for every A,$A \subseteq \Xi_p \times \Xi_q$ knowledge of the truth value of $M_l^{\downarrow p}(\omega, A^{\downarrow p})$ does not change our prediction capability of the values of $M_l(\omega, A)$, that is

$$\underset{\omega}{\mathfrak{Prob}}^{P(\omega)}(M_l(\omega, A)) = \underset{\omega}{\mathfrak{Prob}}^{P(\omega)}(M_l(\omega, A) | M_l^{\downarrow p}(\omega, A^{\downarrow p}))$$

Theorem 15. *The joint distribution over $X = p \cup q$ with disjoint sets of variables p, q, measured compositely, is independent of the variable set p only if $m^{\downarrow q}(\Xi_q) = 1$ that is the whole mass of the distribution marginalized onto q is concentrated at the only focal point Ξ_q.*

Theorem 16. *If for $X = p \cup q$, p, q disjoint, $Bel = (Bel^{\downarrow p})^{\uparrow X}$ that is Bel is the vacuous extension of some Bel defined only over q, then the Bel is independent of the variable set q.*
If for a Bel over $X = p \cup q$ with disjoint sets p, q measured compositely Bel is independent of q, then $Bel = (Bel^{\downarrow p})^{\uparrow X}$.

Definition 17. Let Bel be defined over $X = p \cup q$, p, q disjoint. We shall speak that Bel is *compressibly independent* of q iff $Bel = (Bel^{\downarrow p})^{\uparrow X}$.

REMARK: $m^{\downarrow q}(\Xi_q) = 1$ does not imply empty extension as such, especially for non-singleton values of the variable p. As previously with marginal independence, it is the composite measurement that makes the empty extension a practical notion.

Let us introduce a concept of conditionality related to the above definition of independence. Traditionally, conditionality is introduced to obtain a kind of independence between variables de facto on one another. So let us define that:

Definition 18. For belief function Bel defined for discourse space over set of variables X let $s \subseteq X$. We define (anti-)conditional belief function $Bel^{|s}(A)$ as

$$Bel = Bel^{\downarrow s} \oplus Bel^{|s}$$

Let us notice at this point that the (anti-)conditional belief as defined above does not need to be unique, hence we have here a kind of pseudoinversion of the $\oplus$ operator. Furthermore, the conditional belief does not need to be a belief function at all, because some focal points m may be negative. But it is then the pseudo-belief function in the sense of the DS-theory, that is m,Bel,Pl take values from the interval [-1,1], and only the Q-measure remains non-negative. Please recall the fact that if $Bel_{12} = Bel_1 \oplus Bel_2$ then $Q_{12}(A) = c \cdot Q_1(A) \cdot Q_2(A)$, c being a proportionality factor (as all supersets of a set are contained in all intersections of its supersets and vice versa). Hence also for our conditional belief definition:

$$Q(A) = c \cdot (Q^{\downarrow s})^{\uparrow X}(A) \cdot Q^{X|s}(A)$$

We shall talk later of unnormalized conditional belief iff

$$Q_*^{|s}(A) = Q(A)/(Q^{\downarrow s})^{\uparrow X}(A)$$

Let us now reconsider the problem of independence, this time of a conditional distribution of $(p \cup q \cup r | p \cup r)$ from the third variable set r.

Theorem 17. *Let $X = p \cup q \cup r$, p, q, r pairwise disjoint, and let Bel be defined over X. Furthermore let $Bel^{|p \cup r}$ be a conditional Belief conditioned on variables p, r. Let this conditional distribution be compressibly independent of r. Let $Bel^{\downarrow p \cup q}$ be the projection of Bel onto the subspace spanned by p, q. Then there exists $Bel^{\downarrow p \cup q | p}$ being a conditional belief of that projected belief conditioned on the variable set p such that this $Bel^{|p \cup r}$ is the empty extension of $Bel^{\downarrow p \cup q | p}$*

$$Bel^{|p \cup r} = ((Bel^{\downarrow p \cup q})^{|p})^{\uparrow X}$$

Let us notice that under the conditions of the above theorem

$$Bel = Bel^{|p \cup r} \oplus Bel^{\downarrow p \cup r} = Bel^{\downarrow p \cup q | p} \oplus Bel^{\downarrow p \cup r}$$

and hence for any $Bel^{\downarrow p \cup r | p}$

$$Bel = Bel^{\downarrow p \cup q | p} \oplus Bel^{\downarrow p} \oplus Bel^{\downarrow p \cup r | p}$$

and therefore

$$Bel = Bel^{\downarrow p \cup q} \oplus Bel^{\downarrow p \cup r | p}$$

This means that whenever the conditional $Bel^{|p \cup r}$ is compressibly independent of r, then there exists a conditional $Bel^{|p \cup q}$ compressibly independent of q. But this fact combined with the previous theorem results in:

Theorem 18. *Let $X = p \cup q \cup r$, p, q, r pairwise disjoint, and let Bel be defined over X. Furthermore let $Bel^{|p \cup r}$ be a conditional Belief conditioned on variables p, r. Let this conditional distribution be compressibly independent of r. Then the empty extension onto X of any $Bel^{\downarrow p \cup q | p}$ being a conditional belief of projected belief conditioned on the variable p is a conditional belief function of X conditioned on variables p, r. Hence for every $A \subseteq \Xi$*

$$\frac{Q(A)}{Q^{\downarrow p \cup r}(A^{\downarrow p \cup r})} = \frac{Q^{\downarrow p \cup q}(A^{\downarrow p \cup q})}{Q^{\downarrow p}(A^{\downarrow p})}$$

In this way we obtained some sense of conditionality suitable for decomposition of a joint belief distribution.

4.3 Belief from data

In this section we will show how the ideas of previous section will influence practical implementation of frequency-based belief functions in a relational database. We assume that attributes in the database may be set-valued and that set-theoretic operators are implemented for attributes (including ordering of sets).
First let us assume that we have a convenient situation that for each object $\omega \in \Omega$ and each value $\xi \in \Xi$ we have calculated $M(\omega, \xi)$ and, in the database Cases, the database attribute X stores for each object ω the set of those values ξ for which $M(\omega, \xi) = TRUE$ holds. Let us assume that we have also the database attribute *Lab* initially containing Ξ for each object.

Then let us construct a view bpa (in SQL[64]) reflecting the mass function for the population:

create view ModifiedCases (XMod) as select $X \cap Lab$ *from Cases where* $X \cap Lab \neq \emptyset$*;*
create view Total (Counted) as select count(*) from ModifiedCases;
create view bpa (X,m) as select XMod,count(*)/Counted from ModifiedCases,Total group by XMod;

(m in view bpa is our mass function assigned to the set X of the view bpa)
Now let us look what happens when we "condition" our databases Cases on the event that X stems from ACertainSet. We have to update our database Cases as follows:

update Cases set $Lab = Lab \cap ACertainSet$ *;*

After this operation the view bpa defined above provides us with the resulting mass function $m' = m(||ACertainSet) = m \oplus m_{ACertainSet}$. Later on we may have a further "conditioning" of our databases Cases on the event that X stems from ACertainSet2. Then

update Cases set $Lab = Lab \cap ACertainSet2$ *;*

After this operation the view bpa defined above provides us with the resulting mass function $m" = m'(||ACertainSet2) = m(||ACertainSet, ACertainSet2) = m' \oplus m_{ACertainSet2}$. Calculations in case of general labeling process are more complicated and hard to express in interactive SQL (embedded SQL has to be used). We shall draw attention of the Reader to the fact that in cases of conditioning the group by phrase of an SQL query does not refer to the original value of X, but to its modification by the label Lab, which is updated.

Let us see how laziness of measurement method may be exploited for case-based calculation of belief function.
Let us assume that we have, in the database Cases, the database attribute id being object identity in the external world and a set-valued attribute *Lab* initially containing Ξ for each object and that we have a built-in function M such that for each object ω and each set $A \subseteq \Xi$ it provides us with the measurement value (by accessing e.g. the object by a robot handle) $M(\omega, A) =$ TRUE or FALSE. We have seen previously that the function Bel can be calculated by lazy measurement. For a concrete set A we can calculate Bel(A), after defining the view

create view Total (Counted) as select count(*) from Cases where Lab $\neq \emptyset$*;*

by issuing the query:

select count(*)/Counted from Cases,Total where (if $A \cap Lab \neq \emptyset$ *then* $\neg M(id, (\Xi - A) \cap Lab)$ *else false) ;*

If we want now to condition the Cases by the value set ACertainSet, then we have to update the label by

update Cases set Lab=if $M(id, Lab \cap ACertainSet)$ *then* $Lab \cap ACertainSet)$ *else* $\emptyset$*;*

Then the query above yields for a given set A $Bel'(A)$ with $Bel' = Bel(||ACertainSet)$.
Clearly, this "lazy" procedure with dynamic testing of M() is worth using only if there is really a dramatic difference in costs of applying M to sets of decreasing sizes.

4.4 Independence from data

The preceding sections defined precisely what is meant by marginal independence of two variables in terms of the relationship between marginals and the joint distribution, as well as concerning the independence of a joint distribution from a single variable.

For the former case we can establish frequency tables with rows and columns corresponding to cardinalities of focal points of the first and the second marginal, and inner elements being cardinalities from the respective sum on DS-masses of the joint distribution. Clearly, cases falling into different inner categories of the table are different and hence χ^2 test is applicable. The match can be χ^2-tested. The following formula should be followed for calculation

$$\sum_{\substack{A;\\ A\subseteq \Xi_p,\\ m^{\downarrow p}(A)>0}} \sum_{\substack{B;\\ B\subseteq \Xi_q,\\ m^{\downarrow q}(B)>0}} \frac{((\sum_{C;C\subseteq\Xi,A=C^{\downarrow p}B=C^{\downarrow q}} m(C)) - m^{\downarrow p}(A)\cdot m^{\downarrow q}(B))^2}{m^{\downarrow p}(A)\cdot m^{\downarrow q}(B)}$$

The number of df is calculated as

$$(card(\{A; A\subseteq \Xi_p, m^{\downarrow p}(A)>0\})-1)\cdot(card(\{B; B\subseteq \Xi_q, m^{\downarrow q}(B)>0\})-1)$$

In case of independence of a distribution from one variable (set) one needs to calculate the marginal of the distribution of that variable (set), say s. Then the measure of discrepancy from the assumption of independence is given as:

$$1 - m^{\downarrow s}(\Xi_s)$$

Statistically we can test, based on Bernoullie distribution, what is the lowest possible and the highest possible value of $1 - m^{\downarrow s}(\Xi_s)$ for a given significance level of the true underlying distribution.

4.5 Conditional independence from data

In case of independence between the conditional distribution and one of conditioning variables, however, it is useless to calculate the pseudoinversion of $\oplus$, as we are working then with a population and a sample the size of which is not properly defined (by the "anti-labeling"). But we can build the contingency table of the unconditional joint distribution for the independent variable on the one hand and the remaining variables on the other hand, and compare the respective cells on how they match the distribution we would obtain assuming the independence. The number of degrees of freedom for the χ^2 test would then be the number of focal points of the joint distribution, minus the number of focal points within each of the multi-variable marginals plus one (for covering twice the total sum of 1 on all focal points).

If we test conditional independence of variables X and Y on the set of variables Z, then we have to compare empirical distribution $Bel^{\downarrow X,Y,Z}$ with $Bel^{\downarrow X,Z|Z} \oplus Bel^{\downarrow Y,Z|Z} \oplus Bel^{\downarrow Z}$. The traditional χ^2 statistics is computed (treating the latter distribution as expected one). If the hypothesis of equality is rejected on significance level $\alpha = 0.05$ then X and Y are considered dependent, otherwise independent.

4.6 Summary of Quantitative Approach

This section provides a frequency interpretation of MTE that covers simultaneously static and dynamic aspects of MTE. It is hoped that, though MTE may in fact be suitable for entirely different modes of uncertainty, the study of this interpretation will shed some light onto the nature of Dempster rule of combination of evidence.

The interpretation introduced here does not differ too much in defining the belief function in terms of frequency (or probability) from various other probabilistic approaches proposed in the past [51,58,18,14,34,35]. To pinpoint the essential new quality let us first remind the opinion expressed by Smets on probabilistic approaches proposed in the past. Smets states that "Far too often, authors concentrate on the static component (how beliefs are allocated?) and discover many relations between TBM (transferable belief model of Smets) and ULP (upper lower probability) models, inner and outer measures (Fagin and Halpern [14]), random sets (Nguyen [37]), probabilities of provability (Pearl [39]), probabilities of necessity (Ruspini [46]) etc. But these authors usually do not explain or justify the dynamic component (how are beliefs updated?), that is, how updating (conditioning) is to be handled (except in some cases by defining conditioning as a special case of combination). So I (that is Smets) feel that these partial comparisons are incomplete, especially as all these interpretations lead to different updating rules." [59][pp. 324-325]. We make here a break-through. The dynamic component of DST (Dempster rule) fits our frequency framework. And this is achieved by a simple trick: the label attached to the object. This label is modified by the re-labeling process, it prevents from reading the intrinsic value of the set-valued attribute when calculating the belief function. Thus it reveals two major aspects of DST which were apparently ignored in the past: (1) Dempster rule both selects and modifies objects of a population - differing just from proper probabilistic conditioning which only selects objects, (2) and belief function does not describe the reality as such (is not objective) but rather intermixes the description of real world with personal prejudices (vicious "labels" we attach to our surroundings) - just being a mixture of objective and subjective view of the world.

The new interpretation of DST has also consequences for multivariable considerations of DST. It has been demonstrated that there exists a much richer spectrum of shades of independence in DST than in probability calculus. Especially, non-predictability of one variable from marginal value of the other and vice versa does not imply that their joint distribution may be calculated from their marginals, unless they are compositely measured. This is due to the subtle structure of the space of discourse in DST. Furthermore, there is a split of concepts of conditionality, and hence complexities in conditional independence, compared to probability calculus, as Shafer's conditional belief function $Bel(||Y) = Bel \oplus Bel_Y$ does not have the property that $Bel = Bel(||Y) \oplus Bel^{\downarrow X}$. Hence two different concepts of conditionality emerge: the Shafer's conditional belief as defined in Def.6 and the anti-conditional belief function as defined in Def.18, introduced in this paper. Shafer's conditional belief is the result of reasoning process; whereas anticonditional belief function serves the purpose of decomposition (factorization) of a joint belief distribution. It turns out, however, that there exist a multitude of anticonditional belief functions for a given belief distribution, and even in cases, where variable X appears not to be influenced by a variable Y given a third variable Z, there may exist an anticonditional belief distribution of X given Y,Z such that X is influenced by Y. However important existential properties of anticonditional belief functions could be demonstrated in Theorem 17 and Theorem 18, relating anti-conditional belief distributions in all variables X,Y,Z and of a subset of variables X,Z after ignoring non-influential variable Y. These properties may serve as a foundation for decomposition of joint belief distribution into factors consisting of anti-conditional belief functions.These considerations extend radically the class of belief functions for which such decompositions can be considered, because previously Shenoy in [54] dealt only with so-called positive normal belief functions, which had unique anticonditional beliefs.

The existence of frequency interpretation and of statistical independence tests enables to acquire structural knowledge on Dempster-Shafer belief distribution from data. Several algorithms have been elaborated so far allowing to learn from data tree-structured, polytree-structured and network-structured decomposition of belief functions [?,24,25,28,29,31].

Also algorithms for the reverse process (sample generation from structured belief functions) have been elaborated for this quantitative model of belief functions [?,28,29,33]

5 Discussion and Concluding Remarks

The 32 years since the first publication of Dempster that initiated the Dempster-Shafer Theory of Evidence were scene of fierce discussion of the intrinsic meaning and significance of this new formalism of representation of uncertain knowledge. At least four distinct lines of development may be distinguished: (1) imprecise probabilities (with intuition saying that belief and plausibility were the lower and upper bound on some set of probabilities), (2) uncertainty about possible worlds (with intuition that basic probability assignment function assigning probabilities to completely deterministic possible states of the world) (3) uncertainty about propositional knowledge possessed by agents (beliefs being probabilities of provability), (4) uncertainty about proper decisions under incomplete knowledge of causal variables in decision tables (with plausibility being the relative count of cases supporting a set of decision).

All of these directions of development indicated some sets of expectations and intuitions about the applicability of the DST, but also were accompanied by fierce criticism, counter examples and misunderstandings. Some of them we have outlined in section 1.

But here we want to concentrate on the lesson we have learnt from this turbulent development. Nearly all of the approaches made the following crucial assumption: there is only one intrinsic state of the world, or one intrinsic probability distribution that one wants to approximate from incomplete and/or uncertain observations or enlightened insights. We think that this assumption may always lead to successful counter examples as can be seen from careful inspection of literature.

In order to be able to apply DST consistently, we insist that for the set A Bel(A) does not represent agent's belief that any element of A holds (without knowing which one), but rather that m(A) reflects the belief that exactly all elements of A hold or, alternatively, that one is unable to get the insight which of elements of A hold. What does it mean for a business in practical terms ? Let us assume we need to employ somebody with M.Sc. either in mathematics (e1), or computer science (e2), or physics (e3). "Normally", events canel, e2 and e3 may be considered exclusive (so may be a basis for creation of a probability distribution to describe the possible stream of candidates).However, in some cases there are people having a degree e.g. both in mathematics and physics (hence we have to correct ourselves to generate a DST belief distribution over {e1,e2,e3} to describe the possible stream of candidates).

Furthermore this assumption is crucial in that it requires the revision of the intuition behind "belief revision" or the Dempster's conditioning rule. Usually the rhetorics of "independent" opinion (that the value domain is in fact restricted to some set) is used to justify the way of working of Dempster rule. This rhetorics may be justified if the "opinion" is logical in nature that is represented with a belief function with the whole mass concentrated on a single set. As soon as this is not the case, the counter examples can be transferred to nearly any interpretation of DST suggested so far, because any attempt of reasoning (belief revision) will materialize in assumption of underlying cases having values different for the intrinsic ones.

Our interpretations escape this pitfall. The application of the Dempster rule can be seen as follows: Assume two departments in our firm seek for new employees: one with M.Sc. in physics or computer science, the other in computer science or mathematics. The ingoing stream of candidates applying for a position in the firm is randomly served either by an officer of the one or the other department. In the distributed firmdatabase those that didn't fit the needs of the department of the actual serving officer are discarded, while those fitting are classified as having M.Sc. in physics, or computer science or both, in case of the first department, and in case of the second: as either having M.Sc. in computer science or mathematics or both. This database will of course reflect only the beliefs of the firm and not the real state of the world, because not only not all candidates having degrees both in mathematics and computer science will be registered as such (depending on the department, they will be register as either ones with M.Sc. in computer science or ones with M.Sc. both in mathematics and computer science), but also qualifications like M.Sc. both in mathematics and physics will be completely ignored, not to say about e.g. both M.Sc. in mathematics and history. But on the other hand notice that all candidates counted as support for m(comp.sci., mathem) will actually have both M.Sc. in mathematics and computer science.

Equipped with these basic assumptions we were capable to escape the most frequent point of attack on the Dempster-Shafer Theory of evidence - the claim of missing empirical interpretation.

In the chapter we have explained three different proposals of empirical models, meeting different expectations posed on DST:

- "the marginally correct approximation".
- "the qualitative model"
- "the quantitative model"

The marginally correct approximation assumes that the belief function shall constitute lower bounds for frequencies, however only for the marginals, and not for the joint distribution. Then the reasoning process is expressed in terms of so-called Cano et al. conditionals - a special class of conditional belief functions that are positive. This approach implies modification of reasoning mechanism, because the correctness is maintained only by reasoning forward. Depending on the reasoning direction we need different "Markov trees" for the reasoning engine.

Note that lower/upper bound interpretations have a long tradition for DST [9,34] and has been heavily criticezed [18]. The one that we presented here differs from the known ones significantly as we insist on different reasoning schemes (hypertrees) depending on which are our target variables values of which are to be inferred. This assures overcoming basic difficulties with lower/upper bound interpretations. In this way also we were able to outline an area where the intuition of lower/upper approximation is applicable.

The qualitative approach is based on earlier rough set interpretations of DST, but makes a small and still significant distinction. All computations are carried out in strictly "relational" way that is indistinguishable objects in a database are merged (no object identities). The behavior under reasoning fits strictly to DST reasoning model. Factors of hypergraph representation can be expressed by relational tables. Conditional independence is well defined. However, there is no interpretation for conditional belief functions in this model.

Rough set interpretations [58] were primarily interested in interpreting the belief function in terms of decision tables. However, the Dempster-rule of evidence combination was valid there only for "extended decision tables", not easily derived from original ones. In our interpretation, both the original tables and the resultant tables dealt with when simulating Dempster-rule are conventional decision tables and the process of decision table merging in a natural one (relational join operator).

With our interpretation we show what kind of "possible worlds semantics" may be used consistently with DST: the basic hint when measuring uncertainty is to think in terms of "diversity" (number of different worlds) and not in terms of "frequency" (number of worlds).

The quantitative model assumes that during the reasoning process one attachs labels to objects hiding some of their properties. There is a full agreement with the reasoning mechanism of DST. Conditional independence and conditional belief functions are well defined. We have also elaborated processes that can give rise to well-controlled graphoidally structured belief functions. We elaborated also learning procedures for discovery of graphoidal structures from data.

The quantitative model seems to be the best fitting model for belief functions created so far.

This frequency model differs from what was previously considered [59–61] in that it assumes that reasoning in DST is connected with updating of variables for individual cases. This is different from e.g. reasoning in probability where reasoning means only selection of cases. And just in this way failures of previous approaches could be overcome. A new flavor of the word "belief" comes with this interpretation: we maintain in our section of database only data that we believe are important for a particular local application (like in the above case of two departments looking for job candidates) without bothering of sufficiency of the data for the whole database.

Many authors [53,59] question the need for an empirical model for DST and point rather to theoretical properties of DST considered within an axiomatic framework seeking parallels with the probability theory. Though it is true that the probability theory may be reasoned with within the framework of Kolmogorov axioms and quite useful results derived in this way, one shall still point out that the applicability of probability theory is significanlty connected with frequencies. Both frequencies considered as "naive probabilities", ore ones being probabilities "in the limit". Statistics is clearly an important part of the probabilistic world.

We suspect that a similar relation exists between the axiomatic framework of DST and empirical models for DST. We have just shown that at least three empirical types models with properties postulated in section 1 exist. They may constitute some hints for future search for more application areas of DST. And, as the general human experience demonstrates, practical applications would inevitably lead to further theoretical progress.

References

1. A. Acid , L.M. deCampos, A. Gonzales, R. Molina, N. Perez de la Blanca: Learning with CASTLE. In:Kruse R., Siegel P., eds, *Symbolic and Quantitative*

Approaches to Uncertainty. Lecture Notes In Computer Science 548, Springer-Verlag (1991), 99-106

2. A. Bendjebbour, W. Pieczynski: *Traitement du Signal*, Vol. 14, No.5, 1997, pp. 453-464.
3. C. Beeri, R. Fagin, D. Maier, Y. Yannakakis: On the desirability of acyclic database schemes, *J. Assoc. Comput. Mach.*, vol. 30, no 3, 1983, 479-513.
4. Cano J., Delgado M., Moral S.: An axiomatic framework for propagating uncertainty in directed acyclic networks, *Int. J. of Approximate Reasoning.* 1993:8, 253-280.
5. C.K. Chow, C.N. Liu: Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory* **IT-14**(1968), 462-467
6. G.F. Cooper, E. Herskovits: A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* **9** (1992), 309-347.
7. deKorvin A., Kleyle R., Lea R.: The object recognition problem when features fail to be homogeneous, *IJAR* 1993:8, 141-162.
8. C. Delobel: Normalization and hierarchical dependencies in the relational data model, *ACM Transactions on Database Systems*, vol. 3, no 3, 1978, 201-222.
9. A.P. Dempster: Upper and lower probabilities induced by a multi-valued mapping. *Ann. Math. Stat..* **38**(1967), 325-339.
10. A DromignyBadin,S Rossato,Y.M. Zhu: Radioscopic and ultrasonic data fusion via the evidence theory, *Traitement du Signal*, Vol. 14, No.5, 1997, pp.499-510.
11. Dubois D., Prade H.: Evidence, knowledge and belief functions, *International Journal of Approximate Reasoning*, 1992:6, 295-319.
12. Durham S.D., Smolka J.S., Valtorta M.: Statistical consistency with Dempster's rule on diagnostic trees having uncertain performance parameters, *International Journal of Approximate Reasoning* 1992:6, 67-81.
13. Dutta, S. K., K. Harrison, and R. P. Srivastava, "The Audit Risk Model Under the Risk of Fraud," in Applications of Fuzzy Sets & The Theory of Evidence to Accounting II, Vol. 7, edited by P. Siegel, K. Omer, A. Korvin, and A. Zebda, published by Jai Press Inc., 1998, pp. 221-244.
14. R. Fagin, J.Y. Halpern: Uncertainty, belief, and probability. *Comput. Intell.* **7**(1991), 160-173.
15. Geiger D., Verma T., Pearl J.: d-Separation: From theorems to algorithms. In:Henrion M., Shachter R.D., Kamal L.N. and Lemmer J.F., eds, *Uncertainty in Artificial Intelligence 5.* Elsevier Science Publishers B.V. (North-Holland), 1990, 139-148.
16. Gordon J., Shortliffe E.H.: The Dempster-Shafer theory of evidence, in: Shafer G., Pearl J.,eds, *Readings in Uncertain Reasoning.* Morgan Kaufmann Pub.Inc., San Mateo CA, (1990), 529-539.
17. J.W. Grzymała-Busse: *Managing Uncertainty in Expert Systems*, Kluwer, 1991.
18. J.Y. Halpern, R. Fagin: Two views of belief: belief as generalized probability and belief as evidence.*Artificial Intelligence* **54**(1992), 275-317
19. R. Hummel, M. Landy: A statistical viewpoint on the theory of evidence. *IEEE Trans.PAMI* (1988), 235-247.
20. M. Itoh, T. Inagaki: Evidential theoretic revision of beliefs under a mixed strategy for safety-control of a large-complex plant. *Transactions of the Society of Instrumental and Control Engineering*, Vol. 34, no.5, 1998, pp. 415-421
21. Jafray, J.Y. (1989) Coherent bets under partially resolving uncertainty and belief functions. Theory and Decisions, 26,99-105

22. M.A. Kłopotek: Partial dependency separation - a new concept for expressing dependence/independence relations in causal networks. *Demonstratio Mathematica.* Vol XXXII No 1,1999, pp. 207-226.
23. M.A. Kłopotek: Beliefs in Markov trees - From local computations to local valuation. In:Trappl R., ed., *Cybernetics and Systems Research.* Vol.1., World Scientific Publishers (1994), pp. 351-358
24. M.A. Kłopotek: Learning belief network structure from data under causal insufficiency. In:Bergadano F., DeRaed L., eds, *Machine Learning ECML-94.* Lecture Notes in Artificial Intelligence 784, Springer-Verlag (1994), 379-382.
25. M.A. Kłopotek: Handling Causal Insufficiency in Dempster-Shafer Belief Networks. *Posters of Information Processing and Management of Uncertainty, Paris, 4-8 July 1994.* 41-42.
26. M.A. Kłopotek: Restricted Causal Inference Algorithm, In:Pehrson B., Simon I., eds, *Proc. World Computer Congress of IFIP, Hamburg 28 August - 2 September 1994, Vol.1* . Elsevier Scientific Publishers (North-Holland), Amsterdam (1994), 342-347
27. M.A. Kłopotek: On a Deficiency of the FCI Algorithm Learning Bayesian Networks from Data. to appear in *Demonstratio Mathematica.* Vol. XXXIII, 2000.
28. M.A. Kłopotek: Reasoning from Data in the Mathematical Theory of Evidence, In:*Proc. Eighth International Symposium On Methodologies For Intelligent Systems (ISMIS'94), Charlotte, North Carolina, USA, October 16-19, 1994.* Oak National Laboratory Publishers (1994), 71-84.
29. M.A. Kłopotek: Testumgebung für Entwicklung eines Beratungssystems auf der Basis der Mathematischen Theorie der Evidenz, *Öster. Zeitschrift für Statistik und Informatik, 1994*
30. M.A. Kłopotek: Interpretation of belief function in Dempster-Shafer Theory. *Foundations of Computing & Decision Sciences* **20**(1995)4, pp. 287-306.
31. M.A. Kłopotek: Identification of belief structure in Dempster-Shafer Theory. *Foundations of Computing & Decision Sciences* **21**(1996)1, pp. 35-54.
32. M.A. Kłopotek: *Methods of Identification and Interpretations of Belief Distributions in the Dempster-Shafer Theory* (in Polish). Publisher: Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, 1998, ISBN 83-900820-8-x.
33. M.A. Kłopotek, S.T. Wierzchoń: An Interpretation for the Conditional Belief Function in The Theory of Evidence. In:: Z.Ras, A. Skowron eds: *Foundations of Intelligent Systems.* Lecture Notes in Artificial Intelligence 1609, Springer-Verlag, Berlin, 1999, 494-502.
34. H.E. Kyburg Jr: Bayesian and non-Bayesian evidential updating. *Artificial Intelligence* **31**(1987), 271-293.
35. Y. Ma, D.C. Wilkins: Induction of uncertain rules and the sociopathicity property in DST . In:Kruse R., Siegel P.,eds, *Symbolic and Quantitative Approaches to Uncertainty.* Lecture Notes In Computer Science 548, Springer-Verlag (1991), 238-245.
36. Michalewicz M., Wierzchoń S.T., Kłopotek M.A.: Knowledge Acquisition, Representation & Manipulation in Decision Support Systems. In:, M. Dąbrowski Michalewicz M., Raś Z., eds, *Intelligent Information Systems.* Proceedings of a Workshop held in Augustów, Poland, 7-11 June, ICS PAS (1993), 210- 238, .
37. Nguyen H.T.: On random sets and belief functions, *J. Math. Anal. Appl.* **65**(1978), 539-542.

38. , D Pagac,E.M. Nebot,H Durrant-Whyte: An evidential approach to map-building for autonomous vehicles. IEEE Transactions on Robotics & Automation, vol.14, no. 4, 1998, pp.623-629.
39. J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Influence.* Morgan and Kaufmann, 1988
40. J. Pearl: Bayesian and Belief-Function formalisms for evidential reasoning:A conceptual analysis. In: Shafer G., Pearl J.,eds, *Readings in Uncertain Reasoning.* Morgan Kaufmann Pub.Inc., San Mateo CA, (1990), 540-569.
41. J. Pearl, T. Verma: A theory of inferred causation. In:Allen J., Fikes R., Sandewell E., eds, *Principles of Knowledge Representation and Reasoning.* Proc. of the Second IOnternational Conference, Cambridge, Massachusetts, April 22-25, 1991, San Mateo CA:Morgen Kaufmann (1991), 441-452.
42. G.M. Provan: A logic-based analysis of Dempster-Shafer Theory, *International Journal of Approximate Reasoning* **4**(1990), 451-495.
43. Provan G.M.: The validity of Dempster-Shafer belief functions, *International Journal of Approximate Reasoning*, 1992:6, 389-399.
44. Z.W. Ras: Query processing in distributed information systems, *Fundamenta Informaticae Journal*, Special Issue on Logics for Artificial Intelligence, IOS Press, Vol. XV, No. 3/4, 1991, 381-397
45. Ruspini E.H., Lowrance D.J., Strat T.M.: Understanding evidential reasoning, *International Journal of Approximate Reasoning*, 1992:6, 401-424.
46. E.H. Ruspini: The logical foundation of evidential reasoning, Tech. Note 408, SRI International, Menlo Park, Calif. USA, 1986.
47. G. Rebane, J. Pearl: The recovery of causal poly-trees from statistical data. In:Kanal L.N., Levit T.S., Lemmer J.F., eds, *Uncertainty in Artificial Intelligence 3.* Elsevier Science Publishers B.V., (North Holland) (1989), 175-182
48. M. Singh, M. Valtorta: Construction of bayesian network structures from data: A brief survey and an efficient algorithm. *International Journal of Approximate Reasoning* , **12**(1995), 111-131
49. Spirtes P., Glymour C., Scheines R.: *Causation, Prediction and Search.* Lecture Notes in Statistics 81, Springer-Verlag, 1993.
50. G. Shafer: *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, 1976.
51. G. Shafer: Belief functions: An introduction. In: Shafer G., Pearl J.,eds, *Readings in Uncertain Reasoning.* Morgan Kaufmann Pub.Inc., San Mateo CA, (1990), 473-482.
52. G. Shafer, R. Srivastava: The Bayesian and Belief-Function Formalisms. A General Prospective for Auditing. In: Shafer G., Pearl J.,eds, *Readings in Uncertain Reasoning.* Morgan Kaufmann Pub.Inc., San Mateo CA, (1990), 482-521.
53. G. Shafer: Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning* **4**(1990), 323-362.
54. P.P. Shenoy: Conditional independence in valuation based systems. *International Journal of Approximate Reasoning* **109**(1994),
55. Shenoy P., Shafer G.: Axioms for probability and belief-function propagation, in: Shachter R.D., Levitt T.S., Kanal L.N., Lemmer J.F. (eds): *Uncertainty in Artificial Intelligence 4*, Elsevier Science Publishers B.V. (North Holland), 1990,
56. A. Skowron: Boolean reasoning for decision rules generation, J. Komorowski, Z.W.Raś (Eds): *Methodologies for Intelligent Systems*,

7^th International Symposium, ISMIS'93, Torndheim, Norway, June 1993, Proccedings, Lecture Notes in Artif.Intel.689, Springer-Verlag, 1993, 295-305
57. A. Skowron: A synthesis of decision rules: applications of discernibility matrix properties, In:M. Dąbrowski, M. Michalewicz, Z. Raś eds: *Intelligent Information Systems*, Proceedings of a Workshop held in Augustów, Poland, 7-11 June, 1993, ISBN 83-900820-2-0, 30-46
58. A. Skowron, J.W. Grzymała-Busse: From rough set theory to evidence theory. In:Yager R.R., Kasprzyk J. and Fedrizzi M., eds, *Advances in the Dempster-Shafer Theory of Evidence.* J. Wiley, New York (1994), 193-236.
59. Ph. Smets: Resolving misunderstandings about belief functions. *International Journal of Approximate Reasoning* **6**(1992), 321-344.
60. Ph. Smets, R. Kennes: The tranferable belief model. *Artif.Intel.* **66**(1994), 191-234.
61. Smets Ph.: Numerical represenation of uncertainty. In: Gabbay DF.M., Smets Ph. Eds.: Handbook of defeasible reasoning and uncertainty management systems. Vol. 3. Kluwer, Doordrecht, 1998, 265-309.
62. Srivastava, R. P., "Decision Making Under Ambiguity: A Belief-Function Perspective," Archives of Control Sciences, Vol. 6 (XLII), 1997, No. 1-2, pp. 5-27.
63. M. Studeny: Formal properties of conditional independence in different calculi of AI, Sonderforschungsbericht 343, Diskrete Strukturen in der Mathematik, Universität Bielefeld
64. A. Vang: *SQL and Relational Databases.* Microtrend Books, Slawson Communications Inc., 1991
65. L. Wasserman: Comments on Shafer's "Perspectives on the theory and practice of belief functions", *International Journal of Approximate Reasoning* 1992:6:367-375.
66. S.T. Wierzchoń: On plausible reasoning, in: M.M.Gupta, T.Yamakawa eds.: *Fuzzy Logic in Knowledge Based Systems, Decision and Control* , North-Holland Amsterdam, 1988, 133-152.
67. K. Weichselberger, S. Poehlmann, *A Methodology for Uncertainty in Knowledge-Based Systems.* . Lecture Notes in AI 419, Springer Verlag, 1990.
68. S.T. Wierzchoń, M.A. Kłopotek, Modified Component Valuations in Valuation Based Systems As A Way to Optimize Query Processing *Journal of Intelligent Information Systems* 9, 157-180 (1997).
69. S.T. Wierzchoń: *Metody repreyentacji i pryetwaryania informacji niepewnej w ramach teorii Dempstera-Shafera*, monography, Instytut Podstaw Informatyki PAN, 1996.
70. Zarley D.K., Hsia Y, Shafer G.: Evidential reasoning using DELIEF. In:*Proc. of the Seventh National conference on Artificial Intelligence (AAAI-88), 1.* Minneapolis MN (1988), 205-209.
71. Q. Zhu, E.S. Lee: Dempster-Shafer Approach in Propositional Logic. *International Journal of Intelligent Systems* **8**(1993), 341-349.