Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Approximate Reasoning
41 (2006) 314–330
INTERNATIONAL JOURNAL OF APPROXIMATE REASONING
www.elsevier.com/locate/ijar

On the plausibility transformation method for translating belief function models to probability models

Barry R. Cobb [a], Prakash P. Shenoy [b,*]

[a] *Department of Economics and Business, Virginia Military Institute, Lexington, VA 24450, United States*
[b] *School of Business, University of Kansas, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7585, United States*

Received 1 January 2004; received in revised form 1 May 2005; accepted 1 June 2005
Available online 30 August 2005

Abstract

In this paper, we propose the plausibility transformation method for translating Dempster–Shafer (D–S) belief function models to probability models, and describe some of its properties. There are many other transformation methods used in the literature for translating belief function models to probability models. We argue that the plausibility transformation method produces probability models that are consistent with D–S semantics of belief function models, and that, in some examples, the pignistic transformation method produces results that appear to be inconsistent with Dempster's rule of combination.

Keywords: Dempster–Shafer belief functions; Plausibility transformation; Dempster's rule of combination; Pignistic transformation method

* Corresponding author. Tel.: +1 785 864 7551; fax: +1 785 864 5328.
E-mail addresses: cobbbr@vmi.edu (B.R. Cobb), pshenoy@ku.edu (P.P. Shenoy).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2005.06.008

1. Introduction

Bayesian probability theory and the Dempster–Shafer (D–S) theory of belief functions are two distinct calculi for modeling and reasoning with knowledge about propositions in uncertain domains. Bayesian networks and Dempster–Shafer belief networks both provide graphical and numerical representations of uncertainty. While these calculi have important differences, their underlying structures have many significant similarities. In a recent paper [3], we argue that these two calculi have roughly the same expressive power. We say roughly since we do not have a metric to measure expressiveness exactly.

There are many different semantics of D–S belief functions, including multivalued mapping [7], random codes [19], transferable beliefs [26], probability of provability [16], and hints [15], which are compatible with Dempster's rule of combination. However, the semantics of belief functions as upper and lower probability bounds on some true but unknown probability function are incompatible with Dempster's rule [29]. Also, Smets [24] gives betting rates semantics for belief functions *assuming that* the pignistic transformation is the correct transformation. Since the pignistic transformation does not appear to be consistent with Dempster's rule,[1] these betting rates semantics may not be valid for D–S belief functions. In this paper, we are concerned with the D–S theory of belief functions with Dempster's rule of combination as the updating rule, and not with theories of upper and lower probabilities, nor with Smets' transferable belief model with the pignistic rule. One benefit of studying probability functions derived from D–S belief functions is a clearer understanding of D–S belief function semantics.

In this paper, we propose a new method for translating a D–S belief function model to a Bayesian probability model. This is useful for several reasons. First, a large model of an uncertain domain may have some knowledge represented by belief functions, and some represented by probability functions. To reason with the entire model, one needs to either translate the belief functions to probability functions, or vice-versa.

Second, although there are several proposals for decision-making using belief functions (e.g., [17,14,27,30]), the theory of belief functions lacks a coherent decision theory to guide the choices of lotteries in which uncertainty is described by belief functions. One solution to this situation is to translate a belief function model to a probability model, and then use the Bayesian decision theory to make decisions. Smets [23] has suggested this strategy be used by applying the so-called "pignistic" transformation method. We are concerned that the pignistic transformation method may not be consistent with Dempster's rule of combination. One alternative is to use the plausibility transformation (in place of the pignistic transformation) for making decisions with belief functions.[2]

[1] In Section 4, we give an example and some arguments as to why the pignistic transformation is inconsistent with Dempster's rule of combination.

[2] In some situations such as the Ellsberg paradox [10], decision making using the probability function derived using the plausibility transformation leads to outcomes that are at variance with empirical findings. The topic of normative or descriptive decision making with D–S belief functions is beyond the scope of this paper.

Third, the marginal of a joint belief function for a variable with many states can have an exponential number of focal elements and may be too complex to comprehend. One method to summarize a complex belief function is to translate it to a probability mass function.

Fourth, given the computational complexity of Dempster's rule, it is easy to build belief function models where the marginals of the joint belief function for variables of interest are computationally intractable to calculate. In such cases, one can translate the belief function model to a probability model and use Bayes rule to compute the relevant marginals of the joint probability distribution.

Fifth, a transformation method that is consistent with Dempster's rule will lead to an increased understanding of the D–S theory of belief functions by providing probabilistic semantics for belief functions. For example, consider the basic probability assignment m for a variable H whose state space is $\Omega_H = \{h_1, \ldots, h_{70}\}$: $m(\{h_1\}) = 0.3$, $m(\{h_2\}) = 0.01$, $m(\{h_2, \ldots, h_{70}\}) = 0.69$. One can ask: What does this basic probability assignment mean in the context of, e.g., betting for or against h_1 versus h_2? A transformation method that is consistent with D–S theory semantics could provide a probability function that can be construed as a "meaning" of the basic probability assignment.

Sixth, the literature on belief functions is replete with examples where it is suggested that belief function theory is more expressive than probability theory since a "corresponding" probability model using the pignistic transformation leads to non-intuitive results (see, e.g., [2]). In these examples, if we use the plausibility transformation method to translate the belief function models, the two models—a belief function model and the corresponding probability model using the plausibility transformation—give the same qualitative results.

Seventh, a transformation method that is consistent with D–S belief function theory semantics will lead to a new method for building probabilistic models. One can use belief function semantics of distinct evidence (or no double-counting of uncertain knowledge [21]) to build belief function models and then use the transformation method to convert it to a probability model.

The main contributions of this paper are five theorems and three corollaries that describe some key properties of the plausibility transformation method. These properties allow an integration of Bayesian and D–S reasoning that takes advantage of the efficiency in computation and decision-making provided by Bayesian calculus while retaining the flexibility in modeling evidence that underlies D–S reasoning. These conclusions will lead to a greater understanding of the similarities between the two methods and allow belief function techniques to be used in probabilistic reasoning, and vice versa. We also discuss an example that questions the compatibility of the pignistic transformation method with Dempster's rule of combination.

The remainder of this paper is organized as follows. Section 2 contains notation and definitions. Section 3 describes the plausibility transformation method for translating belief functions to probability functions. Section 4 contains the main results of the paper including one example that is used to raise the issue whether the pignistic transformation method is compatible with Dempster's rule of combination. In Section 5, we summarize and conclude. Proofs of all theorems are found in Appendix A.

2. Notation and definitions

2.1. Probability theory

We will use upper-case Roman alphabets, such as X, Y, Z, etc., to denote variables, and lower-case Roman alphabets, such as r, s, t, etc., to denote sets of variables. Associated with each variable X, is a set of mutually exclusive and exhaustive set of possible states, which is denoted by Ω_X. If s is a set of variables, then its state space is given by $\Omega_s = \times\{\Omega_X | X \in s\}$.

A *probability potential* P_s for s is a function $P_s : \Omega_s \to [0, 1]$. We express our knowledge by probability potentials, which are combined to form the joint probability distribution, which is then marginalized to the variables of interest.

In order to define combination of probability functions, we first need a notation for the projection of states of a set of variables to a smaller set of variables. Here projection simply means dropping extra coordinates; if (w, x, y, z) is a state of $\{W, X, Y, Z\}$, for example, then the projection of (w, x, y, z) to $\{W, X\}$ is simply (w, x), which is a state of $\{W, X\}$. If s and t are sets of variables, $s \subseteq t$, and x is a state of t, then $x^{\downarrow s}$ denotes the projection of x to s.

2.1.1. Combination

Combination in a Bayesian network involves "pointwise" multiplication of functions. Suppose P_s is a probability potential for s and P_t is a probability potential for t. Then $P_s \otimes P_t$ is a probability potential for $s \cup t$ defined as follows:

$$(P_s \otimes P_t)(x) = K^{-1} P_s(x^{\downarrow s}) P_t(x^{\downarrow t}) \tag{2.1}$$

for each $x \in \Omega_{s\cup t}$, where $K = \sum\{P_s(x^{\downarrow s}) P_t(x^{\downarrow t}) | x \in \Omega_{s\cup t}\}$ is the normalization constant.

2.1.2. Marginalization

Let $s\backslash\{X\}$ denote the set-theoretic subtraction of the variable X from set s. Marginalization in a Bayesian network involves addition over the state space of the variables being eliminated. Suppose P_s is a probability potential for s, and suppose $X \in s$. The *marginal of P_s for $s\backslash\{X\}$*, denoted by $P_s^{\downarrow(s\backslash\{X\})}$, is the probability potential for $s\backslash\{X\}$ defined as follows:

$$P_s^{\downarrow(s\backslash\{X\})}(y) = \sum\{P_s(y, x) \mid x \in \Omega_X\} \tag{2.2}$$

for all $y \in \Omega_{s\backslash\{X\}}$.

2.1.3. Inference

The probability potentials specified in a probability model can be used to calculate the prior joint distribution of the variables in the model. Inference in a Bayesian network involves updating the prior joint distribution with observations of actual states of certain variables or likelihoods of occurrence of variables based on new information. The observations and likelihoods are modeled as probability potentials. Once the likelihoods or observations are included in the model, the combination of all potentials is called the joint posterior distribution. Usually, one is interested in the marginals of the joint posterior function for some variables of interest.

2.2. Dempster–Shafer theory of belief functions

Dempster–Shafer (D–S) belief networks are an alternative to probability for modeling knowledge about propositions in uncertain domains graphically and numerically. At the qualitative level, a D–S belief network provides a graphical description of the knowledge base by modeling variables and their relations. At the numerical level, a D–S belief network assigns a D–S belief function or basic probability assignment (bpa) to subsets of the variables in the domain of each relation. Additional knowledge entered as evidence is used to update the D–S belief network.

If Ω_s is the state space of a set of variables s, a function $m : 2^{\Omega_s} \to [0, 1]$ is a bpa for s whenever

$$m(\emptyset) = 0 \quad \text{and} \quad \sum\{m(\boldsymbol{a}) \mid \boldsymbol{a} \in 2^{\Omega_s}\} = 1. \tag{2.3}$$

A bpa can also be stated in terms of a corresponding plausibility function or a belief function. The plausibility function Pl corresponding to a bpa m for s is defined as $Pl : 2^{\Omega_s} \to [0, 1]$ such that for all $\boldsymbol{a} \in 2^{\Omega_s}$,

$$Pl(\boldsymbol{a}) = \sum\{m(\boldsymbol{b}) \mid \boldsymbol{b} \cap \boldsymbol{a} \neq \emptyset\}. \tag{2.4}$$

The belief function Bel corresponding to a bpa m for s is defined as $Bel : 2^{\Omega_s} \to [0, 1]$ such that for all $\boldsymbol{a} \in 2^{\Omega s}$,

$$Bel(\boldsymbol{a}) = \sum\{m(\boldsymbol{b}) \mid \boldsymbol{b} \subseteq \boldsymbol{a}\}. \tag{2.5}$$

The valuation network (VN) graph defined by Shenoy [20] can be used to graphically represent the qualitative features of a D–S belief network. An example of a valuation network is shown in Fig. 1. The rounded rectangles represent variables and the hexagons represent valuations, which are functions representing knowledge about relations between the variables. Each valuation is connected by an edge to each variable in its domain to create a bipartite graph. Rectangles represent evidence. In Fig. 1, evidence is available for variables T and V. The arcs connecting valuations to variables are typically undirected; however if a bpa m for a set of variables, say $h \cup t$, is a "conditional" for some, say h, given the rest t, then this is indicated by making the edges between m and the variables in h directed. Suppose m is a bpa for $h \cup t$. We say m is a *conditional* for h given t if $m^{\downarrow t}$ is a vacuous bpa, i.e., $m^{\downarrow t}(\Omega_t) = 1$. Most of the valuations in Fig. 1 are conditionals. An exception is the bpa for $\{V, G\}$, which is not a conditional.

2.2.1. Projection and extension of subsets

Before we can define combination and marginalization for bpa's, we need the concepts of projection and extension of subsets of a state space.

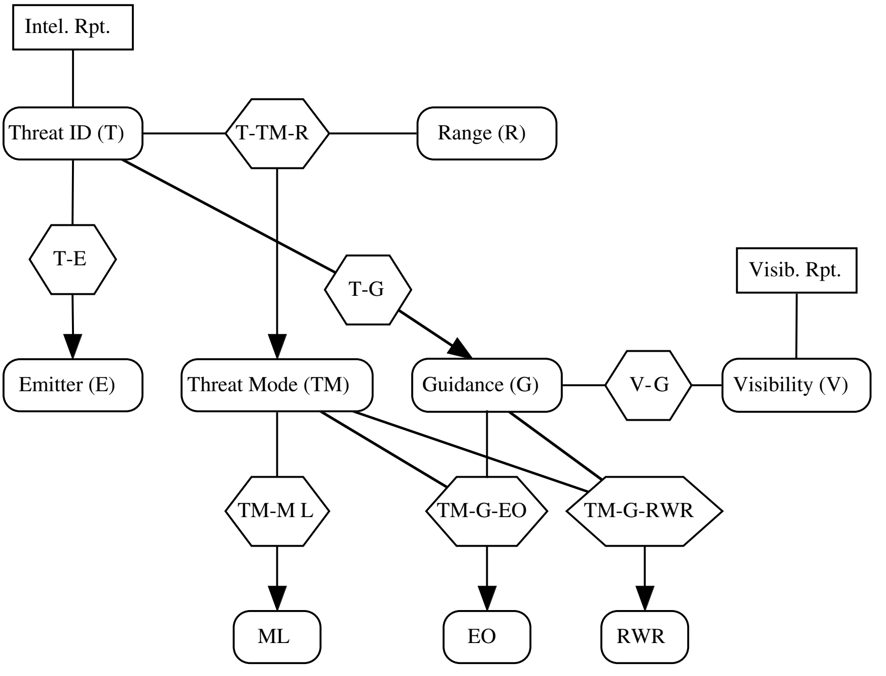

Fig. 1. A Dempster–Shafer belief network for an anti-air threat identification problem.

If r and s are sets of variables, $r \subseteq s$, and $\boldsymbol{a}$ is a non-empty subset of Ω_s, then the *projection of* $\boldsymbol{a}$ *to* r, denoted by $\boldsymbol{a}^{\downarrow r}$, is the subset of Ω_r given by $\boldsymbol{a}^{\downarrow r} = \{x^{\downarrow r} \mid x \in \boldsymbol{a}\}$.

By extension of a subset of a state space to a subset of a larger state space, we mean a cylinder set extension. If r and s are sets of variables, $r \subset s$, and $\boldsymbol{a}$ is a non-empty subset of Ω_r, then the *extension of* $\boldsymbol{a}$ *to* s is $\boldsymbol{a} \times \Omega_{s \setminus r}$. Let $\boldsymbol{a}^{\uparrow s}$ denote the extension of $\boldsymbol{a}$ to s. For example, if $\boldsymbol{a}$ is a non-empty subset of $\Omega_{\{W,X\}}$, then $\boldsymbol{a}^{\uparrow\{W,X,Y,Z\}} = \boldsymbol{a} \times \Omega_{\{Y,Z\}}$.

Calculation of the joint bpa in a D–S belief network is accomplished by combination using Dempster's rule [7]. Consider two bpa's m_A and m_B for a and b, respectively. The combination of m_A and m_B, denoted by $m_A \oplus m_B$, is a bpa for $a \cup b$ given by

$$(m_A \oplus m_B)(\boldsymbol{z}) = K^{-1} \sum \left\{ m_A(\boldsymbol{x}) m_B(\boldsymbol{y}) \mid (\boldsymbol{x}^{\uparrow(a \cup b)}) \cap (\boldsymbol{y}^{\uparrow(a \cup b)}) = \boldsymbol{z} \right\} \tag{2.6}$$

for all non-empty $\boldsymbol{z} \subseteq \Omega_{a \cup b}$, where K is a normalization constant given by

$$K = \sum \left\{ m_A(\boldsymbol{x}) m_B(\boldsymbol{y}) \mid (\boldsymbol{x}^{\uparrow(a \cup b)}) \cap (\boldsymbol{y}^{\uparrow(a \cup b)}) \neq \emptyset \right\}. \tag{2.7}$$

Clearly, if the normalization constant is equal to zero, the combination is not defined, so the two bpa's are said to be *not combinable*. If the bpa's m_A and m_B are based on independent bodies of evidence, then $m_A \oplus m_B$ represents the result of pooling these bodies of evidence. Shafer [18] shows that Dempster's rule is commutative and associative, so the bpa's representing the evidence in the network of Fig. 1, for instance, could be combined in any order to yield the joint bpa.

2.2.2. Marginalization

Suppose m is a bpa for s, and suppose $t \subset s$. The marginal of m for t, denoted by $m^{\downarrow t}$, is the bpa for t defined as follows:

$$m^{\downarrow t}(\boldsymbol{a}) = \sum \{m(\boldsymbol{b}) \mid \boldsymbol{b}^{\downarrow t} = \boldsymbol{a}\} \tag{2.8}$$

for each $\boldsymbol{a} \subseteq \Omega_t$, where $\boldsymbol{b}^{\downarrow t}$ denotes the subset of Ω_t obtained by projecting each element of $\boldsymbol{b}$ to t. Intuitively, marginalization corresponds to coarsening of knowledge.

Similar to the probabilistic case, we make inferences from a belief function model by computing the marginal of the joint belief function for variables of interest. All belief functions that constitute the belief function model must be independent.

3. The plausibility transformation method

Our main goal in this section is to describe a new method for translating a belief function model to a corresponding probability function model. One method of achieving this is to translate each independent belief function in the belief function model to a corresponding probability function. The collection of probability functions then constitutes a corresponding probability model.

Suppose m is a bpa for subset s. Let Pl_m denote the plausibility function for s corresponding to bpa m. Let Pl_P_m denote the probability function that is obtained from m using the plausibility transformation method. Pl_P_m is defined as follows:

$$Pl\_P_m(x) = K^{-1} Pl_m(\{x\}) \tag{3.1}$$

for all $x \in \Omega_s$, where $K = \sum\{Pl_m(\{x\}) \,|\, x \in \Omega_s\}$ is a normalization constant. We will refer to Pl_P_m as the *plausibility probability function* corresponding to bpa m.

3.1. Other transformation methods

The most commonly used transformation method is the pignistic transformation method[3] defined as follows. Let $BetP_m$ denote the pignistic probability function for s corresponding to bpa m. Then,

$$BetP_m(x) = \sum \left\{ \frac{m(a)}{|a|} \mid \boldsymbol{a} \in 2^{\Omega_s} \text{ such that } x \in \boldsymbol{a} \right\} \tag{3.2}$$

for all $x \in \Omega_s$. Daniel [6] has defined a host of other transformation methods.

[3] The name of the transformation is due to Smets [23], but the transformation has been used in the D–S belief function literature earlier (see, e.g., [9]).

4. Properties of the plausibility transformation

Haspert [13] identifies the significance of the relationship between the D–S plausibility function and probability functions, noting that when multiple belief functions on the same domain are combined using Dempster's rule, the masses in the resulting bpa migrate to the outcome for which the product of the plausibility terms is the greatest. He presents heuristic arguments that indicate that the plausibility function can be used to link Bayesian and D–S reasoning. Giles [11] was among the earliest to discuss decision making with plausibility functions. Appriou [1] suggests selecting the hypothesis with the maximum plausibility in a decision-making context.

Dempster [8] states that the upper probability bound (or plausibility) associated with a belief function is the appropriate likelihood function that contains all sample information. Similarly, Halpern and Fagin [12] observe that the plausibility function calculated from a given belief function behaves similarly to a likelihood function and can be used to update beliefs. Given a set H consisting of basic hypotheses—one of which is true—and another set Ob consisting of basic observations, $Pl_{Ob}(H_i) = 1 - Bel_{Ob}(H_i^c) = Pr_i(Ob)/c$, where $c = \max_{j=1,\ldots,m} Pr_j(Ob)$, the plausibility function representing the observations appropriately captures the evidence of the observations.

Additionally, one form of Bayes rule has an analogous rule in terms of plausibility functions. Suppose $P_{A,B}$ is a prior joint probability distribution function for two variables A and B. The marginal distribution for B, denoted by P_B, can be computed from $P_{A,B}$ as follows: $P_B(b) = \sum\{P_{A,B}(a,b) \mid a \in \Omega_A\}$ for all $a \in \Omega_A$. Now suppose we observe $B = b$ where $P_B(b) > 0$. Then, the posterior marginal probability function for A, denoted by $P_{A|b}$ is given by

$$P_{A|b}(a) = P_{A,B}(a,b)/P_B(b) \tag{4.1}$$

for all $a \in \Omega_A$. Now consider the same situation in belief function calculus. Suppose $m_{A,B}$ and $Pl_{A,B}$ represent a prior bpa and the corresponding plausibility function for $\{A, B\}$. Let Pl_B denote the marginal plausibility function for B. Now suppose we observe $B = b$ such that $Pl_B(\{b\}) > 0$. This can be represented by the bpa m_b for B where $m_b(\{b\}) = 1$. The posterior marginal bpa for A, denoted by $m_{A|b}$, is given by $(m_{A,B} \oplus m_b)^{\downarrow A}$. Let $Pl_{A|b}$ denote the corresponding plausibility function for A. It can be shown [18] that $Pl_{A|b}$ is given by

$$Pl_{A|b}(\{a\}) = Pl_{A,B}(\{(a,b)\})/Pl_B(\{b\}) \tag{4.2}$$

for all $a \in \Omega_B$. Comparing (4.1) and (4.2) suggests that the correspondence between a belief function and probability function is via the plausibility function. This correspondence alone does not justify the plausibility transformation, because (4.2) could be restated in terms of the *Bel* function. To provide further justification for the plausibility transformation, we will state the following theorem from Voorbraak [28].

Theorem 4.1. *Suppose $m_1, \ldots, m_k$ are k bpa's. Suppose $Pl_{m_1}, \ldots, Pl_{m_k}$ are the associated plausibility functions, and suppose $Pl\_P_{m_1}, \ldots, Pl\_P_{m_k}$ are the corresponding probability functions obtained using the plausibility transformation. If $m = m_1 \oplus \cdots \oplus$*

m_k is the joint bpa, Pl_m is the associated plausibility function and Pl_P_m is the corresponding plausibility probability function, then $Pl\_P_{m_1} \otimes \cdots \otimes Pl\_P_{m_k} = Pl\_P_m$.

The statement of the theorem is depicted pictorially in Fig. 2. Voorbraak's motivation in stating the result was to efficiently compute a Bayesian approximation of the joint belief function. Notice that from a computational perspective, it is much faster to compute $Pl\_P_{m_1} \otimes \cdots \otimes Pl\_P_{m_k}$ than it is to compute Pl_P_m (since the latter involves Dempster's rule of combination and the former involves Bayes rule). We regard the plausibility probability function Pl_P_m as a translation of m from D–S belief function theory to probability theory, and not necessarily as an approximation.

Theorem 4.1 is significant for several reasons. First, we often create a belief function model, compute the joint belief function, and then translate the joint belief function to a probability function for the reasons described in Section 1. If the transformation used is the plausibility transformation, Theorem 4.1 tells us that we can escape the computational complexity of Dempster's rule and use Bayes rule instead to obtain the same result.

Second, it is often easy to construct belief function models where it is intractable to compute the joint belief function using Dempster's rule. Theorem 4.1 tells us that we can translate the belief function model to a probability model and achieve a more tractable result in probability theory by using Bayes rule.

Third, a qualitative aspect of uncertain knowledge is idempotency. Generally, most uncertain knowledge is non-idempotent. However, some knowledge is idempotent. Examples are observations of values of variables, vacuous knowledge, etc. It is natural to expect the idempotent knowledge be represented by idempotent representations in any calculi that are used to represent the knowledge. A corollary of Theorem 4.1 is that Pl_P_m is idempotent with respect to Bayes rule if m is idempotent with respect to Dempster's rule.

Corollary 4.2. *If m is idempotent with respect to Dempster's rule, i.e., $m \oplus m = m$, then Pl_P_m is idempotent with respect to Bayes rule, i.e., $Pl\_P_m \otimes Pl\_P_m = Pl\_P_m$.*

Fourth, if we use the plausibility transformation for decision making purposes, then Theorem 4.1 tells us that a two-level decision-making scheme such as the one proposed by Smets [23] (with the pignistic transformation) is unnecessary. Since

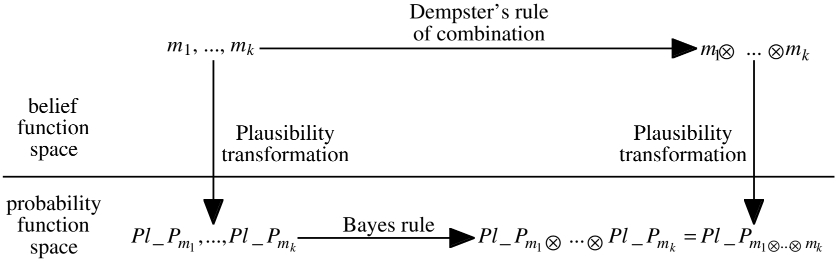

Fig. 2. A pictorial depiction of the statement of Theorem 4.1.

we get the same results whether we use Dempster's rule with belief functions or Bayes rule with probability functions that are translations of the belief functions, we might as well work with the probability functions (from a computational perspective). This does not mean we do not need the D–S theory of belief functions as the different semantics of this theory provides several methods for building models that are otherwise not available in other calculi.

Fifth, one can ask: Why is the property stated in Theorem 4.1 compelling as a generic property for a transformation method? In the following example (adapted from Smets [24]), we demonstrate why non-compliance with this property leads to results that are incompatible with Dempster's rule of combination, the primary updating rule of D–S belief function theory.

Example 4.1. Consider a bpa m for a variable H with state space $\Omega_H = \{h_1, \ldots, h_{70}\}$ as follows: $m(\{h_1\}) = 0.30$, $m(\{h_2\}) = 0.01$, $m(\{h_2, h_3, \ldots, h_{70}\}) = 0.69$. For this bpa m, the plausibility probability function Pl_P_m is as follows: $Pl\_P_m(h_1) = 0.30/49.72 \approx 0.0063$, $Pl\_P_m(h_2) = 0.70/49.72 \approx 0.0146$, $Pl\_P_m(h_3) = \cdots = Pl\_P_m(h_{70}) = 0.69/49.72 \approx 0.0144$, and the pignistic probability function $BetP_m$ is as follows: $BetP_m(h_1) = 0.30$, $BetP_m(h_2) = 0.02$, $BetP_m(h_3) = \cdots = BetP_m(h_{70}) = 0.01$. Notice that the two probability functions are quite different. According to Pl_P_m, h_2 is 2.33 times more probable than h_1. According to $BetP_m$, h_1 is 15 times more probable than h_2. Clearly, the two probability models are incompatible with each other. Which of these two models corresponds to the knowledge in bpa m? The answer depends, of course, on the semantics of D–S belief function theory, which is intrinsically tied to Dempster's rule of combination.

To answer this question, consider the following hypothetical scenario consisting only of bpa m and Dempster's rule of combination. We are interested in the true state of variable H. We start with complete ignorance. Starting from day 1, each day we receive an independent piece of evidence that is represented by bpa m described above. Thus, e.g., on day 2, our total belief is described by $m^2 = m \oplus m$ which is as follows: $m^2(\{h_1\}) = 0.09/0.58 \approx 0.15517$, $m^2(\{h_2\}) = 0.0139/0.58 \approx 0.02397$, and $m^2(\{h_2, \ldots, h_{70}\}) = 0.4761/0.58 \approx 0.82086$. On day 3, our total belief for H is given by m^3, and so on. Table 1 gives the details of some of these functions.

Suppose one subscribes to Smets's decision theory based on the pignistic transformation. On day 1, our belief for H is given by m, and as per $BetP_m$, we are willing to bet for h_1 against h_2 with odds 15:1. On day 2, our total belief for H is given by $m^2 = m \oplus m$, and as per $BetP_{m^2}$ we are willing to bet for h_1 against h_2 with odds 4.33:1. One can ask: Why did the odds for h_1 against h_2 diminish on day 2 from 15:1 to 4.33:1? If the evidence on day 1 supported h_1 against h_2, and a similar evidence was received on day 2, the odds for h_1 against h_2 should have increased and not decreased. On day 3, as per $BetP_{m^3}$, we are willing to bet on h_1 against h_2 with odds ≈1.4:1. On day 4, as per $BetP_{m^4}$, we are now willing to bet for h_2 against h_1 with odds ≈2.06:1, and so on. Dempster's rule of combination tells us that each successive evidence supports h_2 against h_1. This is inconsistent with $BetP_m$. Thus, we question whether the pignistic transformation method is compatible with Dempster's rule.

Table 1
Values of basic probability assignments in Example 4.1 (values are specified up to 5 decimal places)

BPA	Focal sets		
	$\{h_1\}$	$\{h_2\}$	$\{h_2, \ldots, h_{70}\}$
m	0.30	0.01	0.69
m^2	0.15517	0.02397	0.82086
m^3	0.07297	0.03916	0.88786
m^4	0.03263	0.05410	0.91326
m^5	0.01425	0.06843	0.91732
m^{10}	0.00021	0.13399	0.86580

On the other hand, suppose we subscribe to decision making with the plausibility transformation method. On day 1, our belief for H is given by m, and as per Pl_P_m, we are willing to bet for h_2 against h_1 with odds 2.33:1. On day 2, our total belief for H is given by $m^2 = m \oplus m$, and as per $Pl\_P_{m^2}$ we are willing to bet for h_2 against h_1 with odds 2.33^2:1. Notice that this result is a consequence of Theorem 4.1 since $Pl\_P_m \otimes Pl\_P_m = Pl\_P_{m^2}$. On day k, as per $Pl\_P_{m^k}$, we are willing to bet for h_2 against h_1 with odds 2.33^k:1. Thus, the plausibility transformation method appears to be consistent with Dempster's rule.

Smets [24,25] provides a justification for the pignistic transformation by showing that it is invariant with respect to a linear additive updating rule (in the same sense as Theorem 4.1, but with Dempster rule replaced by the linear additive updating rule). We do not find this justification convincing since the linear additive updating rule is not central to D–S belief function theory. The main updating rule in D–S belief function theory is Dempster's rule, and if one substitutes Dempster's rule for the linear additive updating rule in Smets's justification, it results in the condition stated in Theorem 4.1, which is satisfied by the plausibility transformation and not by the pignistic transformation.

This example also demonstrates why the result stated in Theorem 4.1 is fundamental for any method that proposes to translate a D–S belief function model to a corresponding probability model for any of the reasons given in Section 1 including decision-making.

To further demonstrate that the plausibility transformation is consistent with Dempster's rule of combination, we consider another asymptotic property of this transformation. In probability theory, assuming there is a unique state x that is most probable according to a probability function P, x has the property that $\lim_{n\to\infty} P^n(x) = 1$, and $\lim_{n\to\infty} P^n(y) = 0$ for all $y \in \Omega_s \setminus \{x\}$, where P^n denotes $P \otimes \cdots \otimes P$ (n times). Belief functions have a similar property, as stated in the following theorem.

Theorem 4.3. *Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $x \in \Omega_s$ is the most plausible state, i.e., $Pl_m(\{x\}) > Pl_m(\{y\})$, for all $y \in \Omega_s \setminus \{x\}$. Let m^n denote $m \oplus \cdots \oplus m$ (n times), let m^∞ denote $\lim_{n\to\infty} m^n$, and let*

Pl_{m^∞} denote the plausibility function corresponding to m^∞. Then $Pl_{m^\infty}(\{x\}) = 1$, and $Pl_{m^\infty}(\{y\}) = 0$ for all $y \in \Omega_s \backslash \{x\}$.

If a unique most plausible state x exists in a bpa m, a corresponding probability function should have x as its most probable state. This property is satisfied for the plausibility transformation, as stated in the following corollary.

Corollary 4.4. *Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $x \in \Omega_s$ is the most plausible state, i.e., $Pl_m(\{x\}) > Pl_m(\{y\})$, for all $y \in \Omega_s \backslash \{x\}$. Let Pl_P_m denote the plausibility probability function corresponding to m, and let $(Pl\_P_m)^\infty$ denote $\lim_{n\to\infty} (Pl\_P_m)^n$. Then $(Pl\_P_m)^\infty(x) = 1$, and $(Pl\_P_m)^\infty(y) = 0$ for all $y \in \Omega_s \backslash \{x\}$.*

In Theorem 4.3 stated earlier, the belief function m was assumed to have a unique most plausible state x. Now suppose we have a non-singleton subset of most plausible states. In probability theory, if P is such that $\mathbf{t} \subseteq \Omega_s$ is a subset of most probable states, and P^∞ denotes $\lim_{n\to\infty} P^n$, then $P^\infty(x) = P^\infty(y)$ for all $x, y \in \mathbf{t}$, and $P^\infty(z) = 0$ for all $z \in \Omega_s \backslash \mathbf{t}$. Belief functions have a similar property, as stated in the following theorem.

Theorem 4.5. *Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $\mathbf{t} \subseteq \Omega_s$ is a subset of most plausible states, i.e., $Pl_m(\{x\}) = Pl_m(\{y\})$ for all $x, y \in \mathbf{t}$, and $Pl_m(\{x\}) > Pl_m(\{z\})$ for all $x \in \mathbf{t}$, and $z \in \Omega_s \backslash \mathbf{t}$. Let m^∞ denote $\lim_{n\to\infty} m^n$, and let Pl_{m^∞} be the corresponding plausibility function. Then there exists a partition $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ of $\mathbf{t}$ such that $m^\infty(\mathbf{a}_i) = 1/k$ for $i = 1, \ldots, k$, i.e., $Pl_{m^\infty}(\{x\}) = Pl_{m^\infty}(\{y\}) = 1/k$ for all $x, y \in \mathbf{t}$, and $Pl_{m^\infty}(\{z\}) = 0$ for all $z \in \Omega_s \backslash \mathbf{t}$.*

Theorem 4.5 is a generalization of Theorem 4.3 in the sense that if $|\mathbf{t}| = 1$, then Theorem 4.5 reduces to Theorem 4.3. The following corollary generalizes the result in Corollary 4.4 for the case of non-unique most plausible states.

Corollary 4.6. *Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $\mathbf{t} \subseteq \Omega_s$ is a subset of most plausible states, i.e., $Pl_m(\{x\}) = Pl_m(\{y\})$ for all $x, y \in \mathbf{t}$, and $Pl_m(\{x\}) > Pl_m(\{z\})$ for all $x \in \mathbf{t}$ and $z \in \Omega_s \backslash \mathbf{t}$. Let Pl_P_m denote the plausibility probability function corresponding to m, and let $(Pl\_P_m)^\infty$ denote $\lim_{n\to\infty} (Pl\_P_m)^n$. Then $(Pl\_P_m)^\infty(x) = (Pl\_P_m)^\infty(y) = 1/|\mathbf{t}|$ for all $x, y \in \mathbf{t}$, and $(Pl\_P_m)^\infty(z) = 0$ for all $z \in \Omega_s \backslash \mathbf{t}$.*

In general, computation of marginals in a D–S belief network is accomplished with local computation using two operations: combination and marginalization [22]. The plausibility transformation is not invariant with respect to marginalization. Formally, suppose m is a bpa for s, and suppose $t \subset s$. Then $(Pl\_P_m)^{\downarrow t}$ is not always equal to $Pl\_P_{m^{\downarrow t}}$. This is graphically shown in Fig. 3.

Example 4.2. As an example of the inconsistency depicted in Fig. 3, consider the following bpa on the domain $\{V, G\}$:

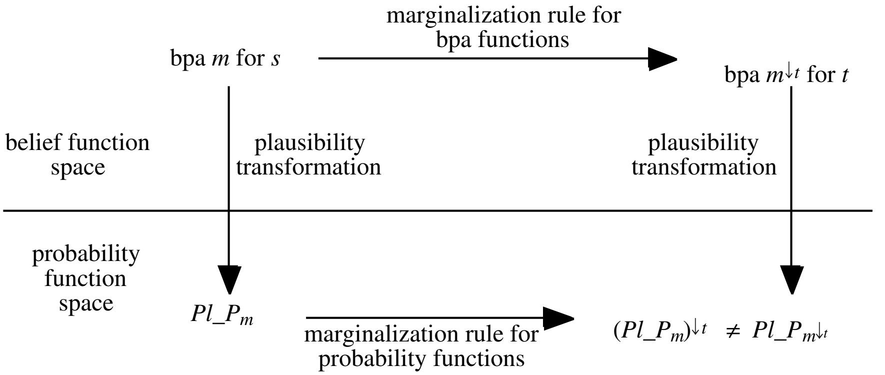

Fig. 3. Plausibility probability transformation is not invariant under marginalization.

$$m_{V\text{-}G}(\{(v_1,g_1),(v_1,g_2)\}) = 0.6,$$

$$m_{V\text{-}G}(\{(v_1,g_1),(v_2,g_1)\}) = 0.3,$$

$$m_{V\text{-}G}(\{(v_1,g_1),(v_1,g_2),(v_2,g_1),(v_2,g_2),(v_3,g_1),(v_3,g_2)\}) = 0.1.$$

Computing the marginal of the bpa for G, then using the plausibility transformation to calculate $Pl\_P_{m_{V\text{-}G}^{\downarrow G}}$ gives

$$m_{V\text{-}G}^{\downarrow\{G\}}(\{g_1\}) = 0.3, \quad Pl_{m_{V\text{-}G}^{\downarrow G}}(\{g_1\}) = 1.0, \quad Pl\_P_{m_{V\text{-}G}^{\downarrow G}}(g_1) = 1.0/1.7 = 0.588,$$

$$m_{V\text{-}G}^{\downarrow\{G\}}(\{g_1,g_2\}) = 0.7, \quad Pl_{m_{V\text{-}G}^{\downarrow G}}(\{g_2\}) = 0.7, \quad Pl\_P_{m_{V\text{-}G}^{\downarrow G}}(g_2) = 0.7/1.7 = 0.412.$$

Alternatively, calculating plausibilities and probabilities for the configurations of $\{V,G\}$ yields:

$$Pl_{m_{V\text{-}G}}(\{(v_1,g_1)\}) = 1.0, \quad Pl_{m_{V\text{-}G}}(\{(v_2,g_1)\}) = 0.4, \quad Pl_{m_{V\text{-}G}}(\{(v_3,g_1)\}) = 0.1,$$

$$Pl_{m_{V\text{-}G}}(\{(v_1,g_2)\}) = 0.7, \quad Pl_{m_{V\text{-}G}}(\{(v_2,g_2)\}) = 0.1, \quad Pl_{m_{V\text{-}G}}(\{(v_3,g_2)\}) = 0.1,$$

$$Pl\_P_{m_{V\text{-}G}}(v_1,g_1) = 0.417, \quad Pl\_P_{m_{V\text{-}G}}(v_2,g_1) = 0.167, \quad Pl\_P_{m_{V\text{-}G}}(v_3,g_1) = 0.042,$$

$$Pl\_P_{m_{V\text{-}G}}(v_1,g_2) = 0.292, \quad Pl\_P_{m_{V\text{-}G}}(v_2,g_2) = 0.042, \quad Pl\_P_{m_{V\text{-}G}}(v_3,g_2) = 0.042.$$

Marginalizing this probability function to G gives

$$(Pl\_P_{m_{V\text{-}G}})^{\downarrow G}(g_1) = 0.625, \quad (Pl\_P_{m_{V\text{-}G}})^{\downarrow G}(g_2) = 0.375.$$

Clearly, the probabilities using the plausibility transformation are not, in general, the same before and after marginalization. However, there are special cases where the plausibility transformation yields the same result before and after marginalization. One such special case is stated in the following theorem. An example illustrating the use of this theorem is given in [4].

Theorem 4.7. *Suppose m_i is bpa for s_i where $s_i = t \cup r_i$, for $i = 1,\ldots,k$. Suppose $r_1,\ldots,r_k$ are pairwise disjoint, i.e., $r_i \cap r_j = \emptyset$ for all $i \neq j$. Let m denote $m_1 \oplus \cdots \oplus m_k$. Then, $Pl\_P_{m^{\downarrow t}} = Pl\_P_{m_1^{\downarrow t}} \otimes \cdots \otimes Pl\_P_{m_k^{\downarrow t}}$.*

Finally, the following theorem allows us to find the plausibility function for a marginal bpa without having to calculate the marginal bpa.

Theorem 4.8. *Suppose m is a bpa for s and $t \subseteq s$. Then,*

$$Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a} \cap \boldsymbol{c}^{\downarrow t} \neq \emptyset} m(\boldsymbol{c})$$

for all $\boldsymbol{a} \subseteq \Omega_t$.

5. Conclusions and summary

The main goal of this paper has been to propose the plausibility transformation method for translating belief function models to probability models, and describe some of its properties. In particular, we have demonstrated that it results in probability models that are invariant with respect to combination, and consequently retains the D–S semantics of belief functions whose primary updating rule is Dempster's rule of combination. However, the plausibility transformation is not invariant with respect to marginalization.

There are a number of other transformation methods proposed in the literature. One of them is the pignistic transformation method. For some examples, the pignistic transformation method results in probability models that are qualitatively different from the probability models produced by the plausibility transformation method. We question whether the pignistic transformation is compatible with Dempster's rule of combination. A comparison of these two methods with several examples is described in [5].

Acknowledgements

We are grateful for extensive comments from Arthur P. Dempster, Yang Chen, Rolf Haenni, Philippe Smets, Jean-Yves Jaffray, Milan Daniel, Liping Liu, and James D. Park on earlier drafts of this paper. We thank the participants of the ECSQARU-2003 conference for valuable comments, suggestions, and discussions.

Appendix A. Proofs

Proof of Theorem 4.1. The proof follows directly from the proof of Proposition 2 in [28]. The proof also follows from the fact that Dempster's rule can be stated as the product of commonality functions and the plausibility and commonality functions have the same values for singleton subsets. □

Proof of Corollary 4.2. Follows immediately from the statement of Theorem 4.1. □

Proof of Theorem 4.3. It follows from Theorem 4.1 that $(Pl\_P_m)^n = Pl\_P_{m^n}$. Taking the limit as $n \to \infty$ on both sides we have $\lim_{n\to\infty}(Pl\_P_m)^n = \lim_{n\to\infty} Pl\_P_{m^n}$, i.e., $(Pl\_P_m)^\infty = Pl\_P_{m^\infty}$. Since $(Pl\_P_m)^n(x) = K_1(Pl_m(\{x\}))^n$ (K_1 is a constant independent of x), $Pl\_P_{m^n}(x) = K_2 Pl_{m^n}(\{x\})$ (K_2 is a constant independent of x), and x is the unique most plausible state, it follows that $(Pl\_P_m)^\infty(x) = 1$, $(Pl\_P_m)^\infty(y) = 0$ for all $y \in \Omega_s \backslash \{x\}$. Therefore, $Pl\_P_{m^\infty}(x) = 1$, and $Pl\_P_{m^\infty}(y) = 0$ for all $y \in \Omega_s \backslash \{x\}$. Therefore $Pl_{m^\infty}(\{x\}) = 1$, and $Pl_{m^\infty}(\{y\}) = 0$ for all $y \in \Omega_s \backslash \{x\}$. □

Proof of Corollary 4.4. Follows immediately from Theorem 4.3 and the definition of Pl_P_m in (3.2). □

Proof of Theorem 4.5. The proof of this theorem is similar to the Proof of Theorem 4.3 and is therefore omitted. □

Proof of Corollary 4.6. Follows immediately from Theorem 4.5 and the definition of Pl_P_m in (3.2). □

The following proposition is a simpler version of Theorem 4.7. We will use it to prove Theorem 4.7.

Proposition A.1. *Suppose m_1 and m_2 are bpa's for s_1 and s_2 where $s_1 = t \cup r_1$ and $s_2 = t \cup r_2$. Suppose r_1 and r_2 are disjoint, i.e., $r_1 \cap r_2 = \emptyset$. Then, $Pl\_P_{(m_1 \oplus m_2)^{\downarrow t}} = Pl\_P_{m_1^{\downarrow t}} \otimes Pl\_P_{m_2^{\downarrow t}}$.*

Proof. It follows from the axioms proposed by Shenoy and Shafer [22] that $(m_1 \oplus m_2)^{\downarrow t} = m_1^{\downarrow t} \oplus m_2^{\downarrow t}$. The proof of this proposition now follows directly from Proposition 4.1 by substituting $m_1^{\downarrow t}$ for m_1 and $m_2^{\downarrow t}$ for m_2. □

Proof of Theorem 4.7. The proof follows directly from the proof of Proposition A.1. □

Proof of Theorem 4.8. The marginal bpa of m for t is defined as $m^{\downarrow t}(\boldsymbol{a}) = \sum\{m(\boldsymbol{b}) \mid \boldsymbol{b}^{\downarrow t} = \boldsymbol{a}\}$ for all $\boldsymbol{a} \subseteq \Omega_t$. The plausibility function values of the marginal bpa of m for t are defined as $Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a}\cap\boldsymbol{d}\neq\emptyset} m^{\downarrow t}(\boldsymbol{d})$ for each $\boldsymbol{a} \subseteq \Omega_t$. This formula can be rewritten as $Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a}\cap\boldsymbol{c}^{\downarrow t}\neq\emptyset} m(\boldsymbol{c})$ for each $\boldsymbol{a} \subseteq \Omega_t$, which proves the theorem. □

References

[1] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, Revue Scientifique de la Défense 11 (1991) 27–40.

[2] P.L. Bogler, Shafer–Dempster reasoning with applications to multisensor target identification systems, IEEE Transactions on Systems, Man and Cybernetics 17 (6) (1987) 968–977.

[3] B.R. Cobb, P.P. Shenoy, A comparison of Bayesian and belief function reasoning, Information Systems Frontiers 5 (4) (2003) 345–358.

[4] B.R. Cobb, P.P. Shenoy, On transforming belief function models to probability models, Working Paper No. 293, School of Business, University of Kansas, Lawrence, KS, 2003. Available from: <http://hdl.handle.net/1808/156>.

[5] B.R. Cobb, P.P. Shenoy, A comparison of methods for transforming belief function models to probability models, in: T.D. Nielsen, N.L. Zhang (Eds.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, 2711, 2003, pp. 255–266.

[6] M. Daniel, Transformations of belief functions to probabilities, in: J. Vejnarová (Ed.), Proceedings of the 6th Workshop on Uncertainty Processing in Expert Systems, 2003, pp. 77–90.

[7] A.P. Dempster, New methods of reasoning toward posterior distributions based on sample data, Annals of Mathematical Statistics 37 (1966) 355–374.

[8] A.P. Dempster, A generalization of Bayesian inference, Journal of the Royal Statistical Society, Series B 30 (2) (1968) 205–247.

[9] D. Dubois, H. Prade, On several representations of an uncertain body of evidence, in: M.M. Gupta, E. Sanchez (Eds.), Fuzzy Information and Decision Processes, North-Holland, Amsterdam, 1982, pp. 167–181.

[10] D. Ellsberg, Risk, ambiguity and the Savage's axioms, Quarterly Journal of Economics 75 (4) (1961) 643–669.

[11] R. Giles, Foundations for a possibility theory, in: M.M. Gupta, E. Sanchez (Eds.), Fuzzy Information and Decision Processes, North-Holland, Amsterdam, 1982, pp. 83–195.

[12] J.Y. Halpern, R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, Artificial Intelligence 54 (1992) 275–317.

[13] J.K. Haspert, Relating Bayesian to Dempster–Shafer analyses, Working Paper, Institute for Defense Analyses, 2001.

[14] J.Y. Jaffray, Linear utility theory for belief functions, Operation Research Letters 8 (1989) 107–112.

[15] J. Kohlas, P.A. Monney, A Mathematical Theory of Hints: An Approach to Dempster–Shafer Theory of EvidenceLecture Notes in Mathematical Systems, 425, Springer-Verlag, Berlin, 1995.

[16] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, CA, 1988.

[17] D. Schmeidler, Subjective probability and expected utility without additivity, Econometrica 57 (1989) 571–587.

[18] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[19] G. Shafer, Belief functions and possibility measures, in: J. Bezdek (Ed.), The Analysis of Fuzzy Information, CRC Press, Boca Raton, FL, 1987, pp. 51–84.

[20] P.P. Shenoy, Valuation-based systems: a framework for managing uncertainty in expert systems, in: L.A. Zadeh, J. Kacprzyk (Eds.), Fuzzy Logic for the Management of Uncertainty, Wiley, New York, 1992, pp. 83–104.

[21] P.P. Shenoy, No double counting semantics for conditional independence, in: F.G. Cozman, R. Nau, T. Seidenfeld (Eds.), Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications, SIPTA, 2005, pp. 306–314.

[22] P.P. Shenoy, G. Shafer, Axioms for probability and belief function propagation, in: R.D. Shachter, T.S. Levitt, J.F. Lemmer, L.F. Kanal (Eds.), Uncertainty in Artificial Intelligence, vol. 4, 1990, pp. 169–198; Reprinted in: J. Pearl, G. Shafer (Eds.), Readings in Uncertain Reasoning, Morgan Kaufmann, San Mateo, CA, pp. 575–610.

[23] P. Smets, Constructing the pignistic probability function in a context of uncertainty, in: M. Henrion, R.D. Shachter, L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence, vol. 5, 1990, pp. 29–40.
[24] P. Smets, Decision making in a context where uncertainty is represented by belief functions, in: R.P. Srivastava, T.J. Mock (Eds.), Belief Functions in Business Decisions, Physica-Verlag, Heidelberg, 2002, pp. 316–332.
[25] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2) (2005) 133–147.
[26] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.
[27] T. Strat, Decision analysis using belief functions, International Journal of Approximate Reasoning 4 (1990) 391–417.
[28] F. Voorbraak, A computationally efficient approximation of Dempster–Shafer theory, International Journal of Man–Machine Studies 30 (1989) 525–536.
[29] P. Walley, Belief function representation of statistical evidence, The Annals of Statistics 15 (4) (1987) 1439–1465.
[30] R.R. Yager, Decision making under Dempster–Shafer uncertainties, International Journal of General Systems 20 (1982) 233–245.