SCHOOL OF BUSINESS WORKING PAPER NO. 293

On Transforming Belief Function Models to Probability Models

Barry R. Cobb and Prakash P. Shenoy

July 2003[†]
Revised February 2004

University of Kansas School of Business
1300 Sunnyside Ave, Summerfield Hall
Lawrence, KS 66045-7585 USA
{brcobb, pshenoy}@ku.edu

[†] Comments and suggestions for improvement are welcome and will be gratefully appreciated.

TABLE OF CONTENTS

ON TRANSFORMING BELIEF FUNCTION MODELS TO PROBABILITY MODELS

Barry R. Cobb and Prakash P. Shenoy

University of Kansas School of Business, 1300 Sunnyside Ave, Summerfield Hall,
Lawrence, KS 66045-7585, USA.

Abstract

In this paper, we explore methods for transforming a belief function model to an equivalent probability model. We propose and define the properties of a method called the plausibility transformation method. We compare the plausibility transformation method with the pignistic transformation method. These two methods yield qualitatively different probability models. We argue that the plausibility transformation method is the correct method that maintains belief function semantics.

Key Words: Bayesian networks, Dempster-Shafer belief functions, valuation-based systems, pignistic transformation, plausibility transformation

---

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 [Cobb and Shenoy 2003], we have argued that these two calculi have roughly the same expressive power.

In this paper, we examine techniques for transforming a belief function model to an equivalent Bayesian probability model. By an equivalent model, we mean a model that results in the same qualitative conclusions and is consistent with D-S belief function semantics. 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 [Jaffray 1989, Strat 1990], 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 an equivalent probability model, and then use the Bayesian decision theory to make decisions. Smets [1990] has suggested this strategy be used by applying the so-called “pignistic” transformation method. We will later argue that the pignistic transformation method is not consistent with Dempster’s rule of combination.

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 an equivalent probability 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 an equivalent probability model and use Bayes rule to compute the relevant marginals of the joint probability distribution.

Fifth, a correct transformation method will lead to an increased understanding of the theory of belief functions by providing probabilistic semantics for belief functions. The literature on belief functions is replete with examples where it is suggested that belief function theory is superior to probability theory since a “corresponding” probability model using the pignistic transformation leads to non-intuitive results [Bogler 1987, Delmotte and Smets 2001]. In all of these examples, if we use the plausibility transformation method, the two models—belief function model and the corresponding probability model using the plausibility transformation—give the same qualitative results.

Sixth, a correct transformation method that is consistent with belief function 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) to build belief function models and then use the plausibility transformation method to convert it to an equivalent probability model.

A popular method for transforming a belief function to a probability function is the so-called “pignistic” transformation method [Smets 1990, Smets and Kennes 1994, Smets 2002]. Another method is the “plausibility” transformation method [Voorbraak 1989, Shenoy and Shenoy 2002]. In this paper, we study the properties of these two methods and compare them. In many cases, these two methods lead to radically different probability models starting from the same belief function model. We argue that if one is interested in an equivalent model, the plausibility transformation method is the correct method, and that it results in a probability model that is consistent with Dempster’s rule of combination.

There are many different semantics of D-S belief functions, including multivalued mapping [Dempster 1966], random codes [Shafer 1987], transferable beliefs [Smets and Kennes 1994], and hints [Kohlas and Monney 1995], 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 [Walley 1987]. Also, Smets [2002] gives betting rates semantics for belief functions *assuming that* the pignistic transformation is the correct transformation. Since the pignistic transformation is inconsistent with Dempster’s rule, these betting rates semantics are not 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 or with Smets’ transferable belief model with the pignistic rule. One benefit of studying probability functions derived from D-S belief functions is a more clear understanding of D-S belief function semantics.

The main contributions of this paper are five theorems and three corollaries that describe some key properties of the plausibility transformation method (that are not satisfied by the pignistic method). These properties allow an integration of Bayesian and D-S reasoning that takes advantage of the flexibility in computation and decision-making provided by Bayesian calculus while retaining the superiority 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.

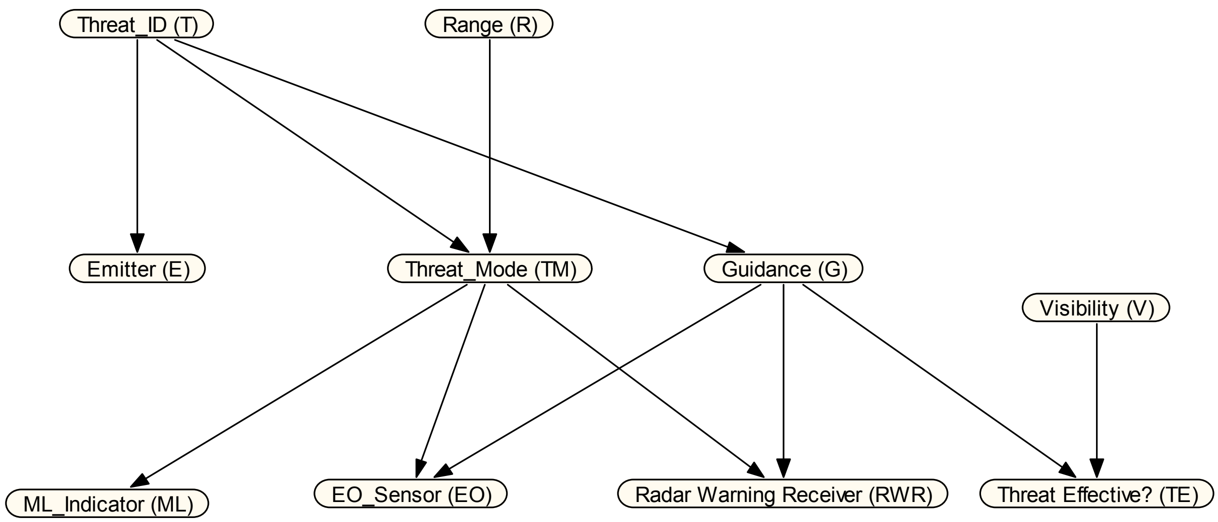
The remainder of this paper is organized as follows. Section 2 contains notation and definitions. Section 3 describes the pignistic and plausibility methods of transforming belief functions to probability functions. Section 4 studies four examples in detail to emphasize the differences between the pignistic transformation and the plausibility transformation methods. Section 5 contains the main results of the paper. In Section 6, we summarize and conclude. Proofs of all theorems are found in the Appendix.

2. Notation and Definitions

Probability Theory. Bayesian networks model knowledge about propositions in uncertain domains using graphical and numerical representations [Spiegelhalter *et al.* 1993]. At the qualitative level, a Bayesian network is a directed acyclic graph where nodes represent variables and the graph represents conditional independence relations among the variables. At the numerical level, a Bayesian network consists of a factorization of a joint probability distribution into a set of conditional distributions, one for each variable in the network. Additional knowledge in the form of likelihood functions can be used to update the joint probability distribution.

Figure 2.1 shows a Bayesian network for a hypothetical anti-air threat identification problem.

Figure 2.1. A Bayesian Network for an Anti-Air Threat Identification Problem



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

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.

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. The un-normalized combination will be denoted by $\otimes'$, i.e.,

$$(P_s \otimes' P_t)\,(x) = P_s(x^{\downarrow s})P_t(x^{\downarrow t}) \tag{2.2}$$

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:

$$P_s^{\downarrow(s\backslash\{X\})}(y) = \Sigma\{P_s(y, x) \mid x \in \Omega_X\} \tag{2.3}$$

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

Inference. The conditionals specified in the construction of a Bayesian network 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. 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.

Dempster-Shafer Theory of Belief Functions. Dempster-Shafer (D-S) belief networks are an alternative to Bayesian networks 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

$$m(\varnothing) = 0, \text{ and } \Sigma\{m(\boldsymbol{a}) \mid \boldsymbol{a} \in 2^{\Omega_s}\} = 1. \tag{2.4}$$

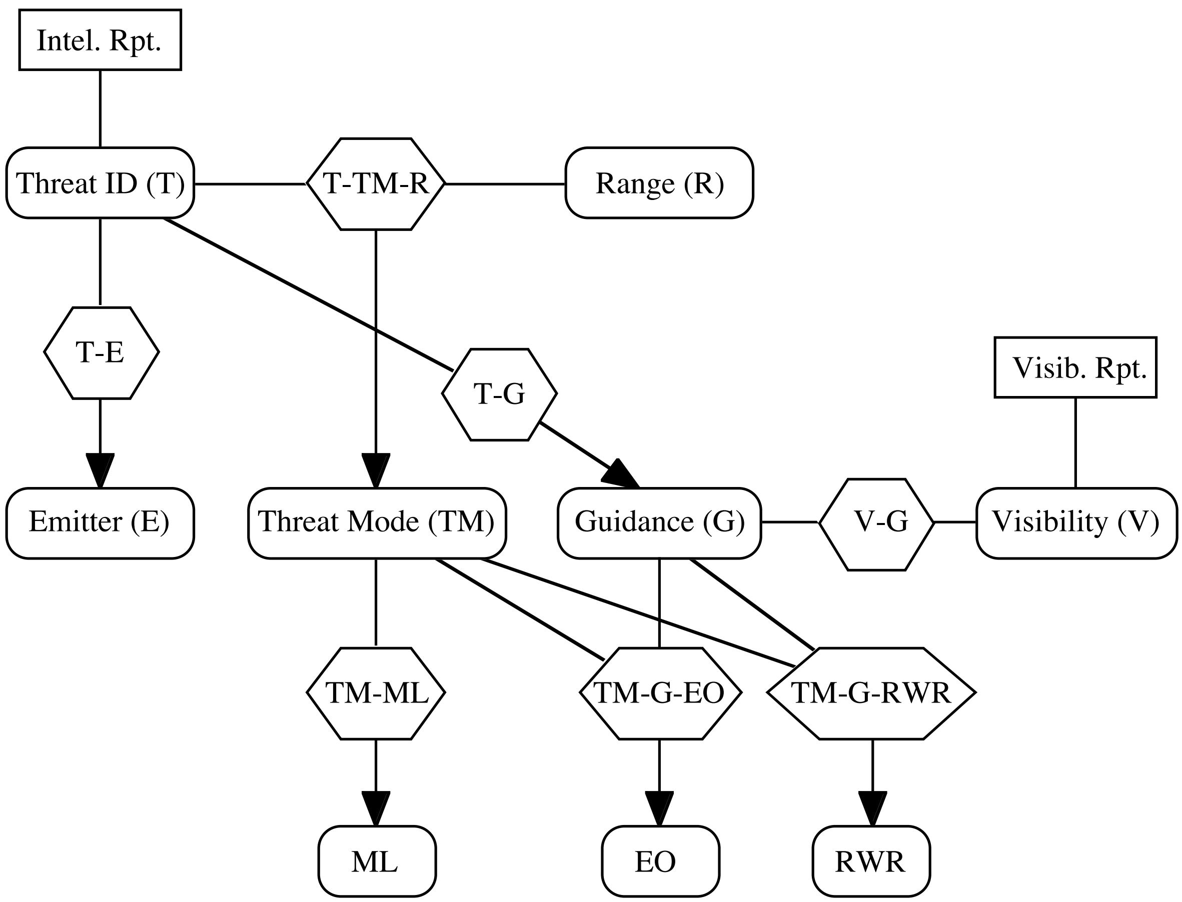
A bpa can also be stated in terms of a corresponding plausibility function or a belief function. The plausibility function Pl corresponding to a bpa m for s is defined as Pl: $2^{\Omega_s} \to [0,1]$ such that for all $\boldsymbol{a} \in 2^{\Omega_s}$,

$$Pl(\boldsymbol{a}) = \Sigma\{m(\boldsymbol{b}) \mid \boldsymbol{b} \cap \boldsymbol{a} \neq \varnothing\}. \tag{2.5}$$

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

$$Bel(\boldsymbol{a}) = \Sigma\{m(\boldsymbol{b}) \mid \boldsymbol{b} \subseteq \boldsymbol{a}\} \tag{2.6}$$

Figure 2.2. A Dempster-Shafer Belief Network for the Anti-Air Threat Identification Problem



The valuation network (VN) graph defined by Shenoy [1992] can be used to graphically represent the qualitative features of a D-S belief network. This is done for the hypothetical anti-air threat identification problem in Figure 2.2. 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 Figure 2.2, 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., $m^{\downarrow t}(\Omega_t) = 1$. Since the D-S network of Figure 2.2 models the same knowledge as described in the Bayesian network of Figure 2.1, most of the valuations representing the knowledge of the domain are conditionals. An exception is the bpa for $\{V, G\}$, which is not a conditional.

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.

If *r* and *s* are sets of variables, $r \subseteq s$, and $\boldsymbol{a}$ is a nonempty 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 nonempty 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 nonempty 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 [Dempster 1966]. 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

$$(m_A \oplus m_B)(\boldsymbol{c}) = K^{-1} \sum \{m_A(\boldsymbol{x}) m_B(\boldsymbol{y}) \mid (\boldsymbol{x}^{\uparrow (a \cup b)}) \cap (\boldsymbol{y}^{\uparrow (a \cup b)}) = \boldsymbol{c}\} \tag{2.7}$$

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

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

The un-normalized Dempster's rule of combination is denoted by $\oplus'$, i.e.,

$$(m_A \oplus' m_B)(\boldsymbol{c}) = \sum \{m_A(\boldsymbol{x}) m_B(\boldsymbol{y}) \mid (\boldsymbol{x}^{\uparrow (a \cup b)}) \cap (\boldsymbol{y}^{\uparrow (a \cup b)}) = \boldsymbol{c}\}$$

for all non-empty $\boldsymbol{c} \subseteq \Omega_{a \cup b}$.

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 [1976] shows that Dempster's rule is commutative and associative, so the bpa's representing the evidence in the network of Figure 2.2, for instance, could be combined in any order to yield the joint bpa.

A useful way to summarize the information contained in the resulting bpa is to calculate the corresponding plausibility function for singleton subsets. In the model of Figure 2.2, it may be useful to focus on the singleton elements of Threat (T) to determine which are now considered most likely. In the next section, we will suggest building a probability function based on the plausibility function to summarize the information in a belief function.

3. Transformation of Belief Function Models to Probability Models

Our main goal in this section is to describe a method for translating a belief function to an equivalent probability function. We believe that these two uncertain reasoning calculi are equally expressive and therefore such a transformation should exist. The process of transforming a bpa to an equivalent probability function is important for several reasons, as outlined in the Section 1.

We will examine two distinct methods for transforming a belief function to an equivalent probability function. The first method is called a pignistic transformation and is due to Philippe Smets [Smets 1990, Smets and Kennes 1994]. The second method is called the plausibility transformation and has been suggested by Voorbraak [1989] and used by Shenoy and Shenoy [2002].

Pignistic Transformation. Suppose m is a bpa for subset s. Let $BetP_m$ denote the corresponding probability function obtained using the pignistic transformation [Smets 1990, Smets and Kennes 1994, Smets 2002]. $BetP_m$ is defined as follows:

$$BetP_m(x) = \Sigma\left\{ \frac{m(\boldsymbol{a})}{|\boldsymbol{a}|} \mid \boldsymbol{a} \in 2^{\Omega_s} \text{ such that } x \in \boldsymbol{a} \right\} \tag{3.1}$$

for each $x \in \Omega_s$. We will refer to $BetP_m$ as the *pignistic probability function* (corresponding to bpa m).

Smets [1990, 2002] claims that beliefs are held at the credal level and are represented by belief functions, whereas the pignistic probability transformation is used to produce a probability function only to make decisions and not to represent beliefs. The pignistic transformation is justified based on a so-called "rationality" requirement, which implies a mathematical requirement of linearity. Baroni and Vicig [2003] shows that the pignistic transformation always provides values that lie between the belief and plausibility values. More precisely, if m is a bpa for s, and Pl_m and Bel_m are corresponding plausibility and belief functions, then

$$Bel_m(\{x\}) \leq BetP_m(x) \leq Pl_m(\{x\}) \text{ for all } x \in \Omega_s. \tag{3.2}$$

Belief functions are often interpreted as upper and lower bounds on some true but unknown probabilities. In this case, the inequality in (3.2) is compelling. But these semantics are inconsistent with Dempster's rule of combination, and since we are concerned with the D-S theory of belief functions, an inequality of the type in (3.2) is not compelling.

Plausibility Transformation. 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:

$$Pl\_P_m(x) = K^{-1}\, Pl_m(\{x\}) \tag{3.3}$$

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

Belief Transformation. A belief transformation method, denoted by Bel_P_m, which normalizes the belief function values of the singleton subsets of the state space, is another potential transformation method. However, this method would disregard all information in the non-singleton focal elements and would create an undefined probability function if no singleton focal elements exist. Additionally, changes could be made to non-singleton focal elements that would not be reflected in the transformed probability distribution. For these reasons, we will not study the belief transformation method any further. Daniel [2003] has defined a host of other transformation methods, none of which are compelling for the case of D-S theory of belief functions.

Is Any Useful Information Lost in the Transformation? Any method of transforming belief functions to probability functions may involve a potential loss of information. Consider three bpa's for X with state space $\Omega_X = \{x_1, x_2, x_3\}$:

$m_1(\Omega_X) = 1$

$m_2(\Omega_X) = 1/2$, $m_2(\{x_1\}) = 1/6$, $m_2(\{x_2\}) = 1/6$, $m_2(\{x_3\}) = 1/6$

$m_3(\{x_1\}) = 1/3$, $m_3(\{x_2\}) = 1/3$, $m_3(\{x_3\}) = 1/3$

The plausibility probability function and pignistic probability function for each of these bpa's are exactly the same, with probabilities 1/3 for each element in the state space. Although these bpa's produce identical probability functions, they represent varying degrees of knowledge about the state space: m_1 represents total ignorance about the true state, m_3 represents knowledge that all three states are equally likely, and m_2 represents a mid-point on the continuum between total ignorance and knowledge that the states are equally likely. The ability to model ignorance is commonly cited as an advantage of belief functions. However, a decision theory that takes advantage of this expressiveness of belief functions has yet to be formulated[1].

4. Four Examples

The previous example illustrates three cases where the pignistic transformation and the plausibility transformation will yield the same results. In general, the two transformations yield different results. To highlight the differences, we will examine four examples in considerable detail.

Example 1: **Peter, Paul, and Mary** [Smets and Kennes 1994]. A mafia don, the Godfather, has three assassins, Peter, Paul, and Mary. Needing to assassinate an informant, Mr. Jones, the

[1] Giang and Shenoy [2003] have described a decision theory for the class of partially consonant belief functions resulting from statistical evidence that is closed under Walley's rule of combination [Walley 1987]. This decision theory can explain ambiguity aversion as demonstrated by Ellsberg paradox [Ellsberg 1961].

Godfather decides to first toss a fair coin to decide the sex of the assassin. If the toss results in heads, he will pick Mary for the job. If the toss results in tails, he will ask either Peter or Paul to do the job. In the case of tails, we have no knowledge of how the Godfather will select between Peter and Paul. Now suppose we find Mr. Jones assassinated. An informant in the mafia organization has informed the district attorney (DA) about the Godfather's incomplete mechanism for choosing among Peter, Paul, and Mary. The DA would like to indict Peter, Paul or Mary (in addition to the Godfather). Who should the DA indict?

Let A denote the assassin variable. A has three states: Peter, Paul, and Mary. Given our knowledge of the incomplete protocol of how the assassin was selected, we can represent it by the bpa m_1 for A as follows: $m_1(\{\text{Mary}\}) = 0.5$, $m_1(\{\text{Peter, Paul}\}) = 0.5$. The pignistic probability function corresponding to m_1 is as follows: $BetP_{m_1}(\text{Mary}) = 0.5$, $BetP_{m_1}(\text{Peter}) = 0.25$, $BetP_{m_1}(\text{Paul}) = 0.25$. The plausibility probability function corresponding to m_1 is as follows: $Pl\_P_{m_1}(\text{Mary}) = Pl\_P_{m_1}(\text{Peter}) = Pl\_P_{m_1}(\text{Paul}) = 1/3$. The difference between the two probability functions in this example can be understood as follows. The lack of knowledge of how the Godfather will select between Peter and Paul in the case where the toss results in tails is one reason why we represent our knowledge using bpa m_1. If we knew the complete protocol, we would have a Bayesian belief function for A. Given that we want to transform m_1 to a probability function, the pignistic transformation completes the protocol by dividing probabilities equally between Peter and Paul. We refer to this assignment of equal probabilities as a *random choice* protocol. The plausibility transformation on the other hand is more cautious. We have no reason to believe a random mechanism will be used to decide between Peter and Paul. The mafia don may always prefer Peter to Paul, or perhaps Paul to Peter. As we said before, we know *nothing* about the mechanism. We don't even know whether it is deterministic, random, or something else. As per the plausibility function, there is a 0.5 chance that Mary is not the assassin, a 0.5 chance that Peter is not the assassin, and a 0.5 chance that Paul is not the assassin. This explains the plausibility probability function $Pl\_P_{m_1}$.

Clearly, the two transformation methods yield qualitatively different results starting from the same bpa m_1. Which probability distribution can be considered as equivalent to m_1? In the following paragraphs, we describe two arguments (flawed, in our opinion) in favor of the pignistic transformation method and two arguments (compelling, in our opinion) in favor of the plausibility transformation method.

Consider the following argument in favor of the pignistic transformation method[2]. From an argumentative point of view [Haenni and Lehmann 2002], there is exactly one "argument" for Mary and one "counter-argument" each for Mary, Peter and Paul, respectively, as follows:

	Arguments	*Counter-arguments*		*Bel*	*Pl*
Mary	Heads	Tails	⇒	0.5	0.5
Peter	–	Heads	⇒	0	0.5
Paul	–	Heads	⇒	0	0.5

From the argumentative point of view, a good transformation method should take both arguments and counter-arguments into account. The pignistic transformation method considers both in this example by averaging the weights of arguments and counter-arguments[3]. On the other hand, the plausibility transformation method takes only counter-arguments into account (ignoring arguments). What this argument fails to notice is that the counter-arguments for Peter and Paul are exactly the same as the argument for Mary. A belief function has exactly the same information as in a corresponding plausibility function, $Pl(\boldsymbol{a}) = 1 - Bel(\Omega_A \setminus \boldsymbol{a})$ for all $\boldsymbol{a} \subseteq \Omega_A$. Thus, in averaging the weights of arguments and counter-arguments, we are selectively double-counting some information and violating a fundamental tenet of uncertain reasoning. The plausibility transformation method is based only on the plausibility function with no risk of double counting of uncertain information.

Another argument against the plausibility transformation is as follows. The plausibility transformation assigns equal probabilities for Peter, Paul and Mary. If one were to use this

[2] This argument was provided by Rolf Haenni [private communication].

[3] Although in this example, $BetP_{m_1}(x) = [Bel_{m_1}(\{x\}) + Pl_{m_1}(\{x\})]/2$ for all $x \in \Omega_A$, this relation doesn't hold true in general.

plausibility probability function for betting purposes repeatedly, e.g., for and against Mary with odds 1:2, then one could set up a so-called Dutch book against such an user. However, such an argument can be used against any transform including the pignistic transform. If one uses the pignistic probability function to bet on Mary against Peter with odds 2:1, then one is susceptible to a Dutch book with a godfather who may always prefer Peter over Paul.

One way to resolve the conflict between *BetP* and *Pl_P* in this example is to appeal to a qualitative property of uncertain knowledge. Suppose we have two pieces of identical, independent evidence about the assassin, both equal to the bpa m_1. If we use Dempster's rule to combine these two independent pieces of evidence, we observe that $m_1 \oplus m_1 = m_1$, i.e., m_1 is idempotent. Idempotency is an important qualitative property of uncertain knowledge since double counting of idempotent knowledge is harmless. We notice that $Pl\_P_{m_1}$ is idempotent, i.e., $Pl\_P_{m_1} \otimes Pl\_P_{m_1} = Pl\_P_{m_1}$. Thus qualitatively, m_1 and $Pl\_P_{m_1}$ share the same property of idempotency. However notice that $BetP_{m_1}$ is not idempotent. Denoting $BetP_{m_1} \otimes BetP_{m_1}$ by $BetP_m$, we have $BetP_m$(Mary) = 2/3 and $BetP_m$(Peter) = $BetP_m$(Paul) = 1/6. In the next section, we will show that the plausibility transformation method always satisfies this important property of idempotency.

Continuing the Peter, Paul or Mary saga, suppose we subsequently learn that Peter has a cast-iron alibi during the time Mr. Jones was assassinated. This piece of evidence can be represented by the bpa m_2 for A as follows: m_2({Paul, Mary}) = 1. If we combine the two independent bpa's m_1 and m_2, we get $(m_1 \oplus m_2)$({Paul}) = $(m_1 \oplus m_2)$({Mary}) = 0.5. Since the joint bpa has only singleton focal subsets, both the pignistic and plausibility probability functions corresponding to $m_1 \oplus m_2$ agree: $BetP_{m_1 \oplus m_2}$(Paul) = $Pl\_P_{m_1 \oplus m_2}$(Paul) = $BetP_{m_1 \oplus m_2}$(Mary) = $Pl\_P_{m_1 \oplus m_2}$(Mary) = 0.5. However, if we were using the pignistic probability distribution $BetP_{m_1}$, and we update this probability distribution (using Bayes rule) with the evidence of Peter's alibi (represented with a likelihood vector that has 0 for Peter and 1's for Paul and Mary), we end with a probability distribution for A that has probability 2/3 for Mary and 1/3 for Paul, a result that does not coincide with $BetP_{m_1 \oplus m_2}$. On the other hand, if we were using the plausibility probability distribution $Pl\_P_{m_1}$, and we update this distribution with the evidence of Peter's alibi, the result is a probability distribution for *A* that has probability 1/2 for Paul and 1/2 for Mary, exactly the same probability distribution as $Pl\_P_{m_1 \oplus m_2}$. We will show in the next section that this equivalence for the plausibility transformation method is no coincidence. This is yet another argument for why the plausibility transformation method yields an equivalent probability model. ■

Next, we discuss an example from Smets [2002] that is alleged to be a "counter-example" to using the plausibility transformation method. On closer scrutiny, we observe that this example demonstrates that the pignistic transformation method is inconsistent with Dempster's rule of combination.

Example 2: Counter-Example [Smets 2002]**.** 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 pignistic probability function $BetP_m$ is as follows: $BetP_m(h_1) = 0.30$, $BetP_m(h_2) = 0.02$, $BetP_m(h_3) = \ldots = BetP_m(h_{70}) = 0.01$. The un-normalized plausibility probability function $Pl\_P'_m$ is as follows: $Pl\_P'_m(h_1) = 0.30$, $Pl\_P'_m(h_2) = 0.70$, $Pl\_P'_m(h_3) = \ldots = Pl\_P'_m(h_{70}) = 0.69$.

Clearly, the two probability functions are very different. The pignistic probability function has h_1 15 times more likely than h_2 whereas the plausibility probability function has h_2 2.33 times more likely than h_1. Smets [2002] says that the pignistic probability function is more appropriate than the plausibility probability function. We disagree. Our interpretation (supported by many authors, e.g., Baroni and Vicig [2003]) is that the pignistic transformation uses a random protocol where the probability of 0.69 is divided equally amongst the 69 states $h_2, \ldots, h_{70}$. But the random protocol is not part of the belief function semantics. If it were, there would be no need for belief functions.

Shafer [1976] states that $m(\boldsymbol{a})$ should be interpreted as the probability mass that is "confined to $\boldsymbol{a}$ but can move freely to every point of $\boldsymbol{a}$" (p. 40). Thus, we have belief of 0.70 *against* h_1, a belief of 0.30 *against* h_2, and a belief of 0.31 *against* $h_3, \ldots, h_{70}$. Rather than use a random choice protocol, the plausibility transformation assumes that all mass can move freely to any state in the focal element of the belief function. This is one argument for the plausibility probability function.

A more compelling argument for the plausibility transformation method involving Dempster's rule of combination is as follows. Consider a hypothetical situation where we have n independent pieces of evidence all of which are exactly equal to m. Combining these n pieces of evidence by Dempster's rule yields $m\oplus\ldots\oplus m$ (n times), which is denoted by m^n. For large n, for example, if $n \geq 500$, we observe that $m^n(\{h_2\}) \approx 1$, so the result is more consistent with Pl_P_m (that has h_2 as the most probable state) than with $BetP_m$ (that has h_1 as the most probable state). Notice that if we combine Pl_P_m n times using Bayes rule (or pointwise multiplication) and denote the result by $(Pl\_P_m)^n$, for large n we get the result that $(Pl\_P_m)^n(h_2) \approx 1$. On the other hand, we have $(BetP_m)^n(h_1) \approx 1$ for large n. Therefore, we conclude that Pl_P_m is equivalent to m and that $BetP_m$ isn't. This example demonstrates unequivocally that the pignistic transformation is inconsistent with Dempster's rule of combination. ■

Example 3: Non-Unique Most Plausible States. In the previous example, we had a unique most plausible state. Now consider a bpa m for a variable H with state space $\Omega_H = \{h_1, h_2, h_3, h_4, h_5, h_6\}$ as follows: $m(\{h_1, h_5\}) = 0.10$, $m(\{h_1, h_6\}) = 0.20$, $m(\{h_2, h_3, h_4\}) = 0.30$, $m(\{h_5, h_6\}) = 0.04$, $m(\Omega_H) = 0.36$. The pignistic probability function $BetP_m$ is as follows: $BetP_m(h_1) = 0.21$, $BetP_m(h_2) = 0.16$, $BetP_m(h_3) = 0.16$, $BetP_m(h_4) = 0.16$, $BetP_m(h_5) = 0.13$, $BetP_m(h_6) = 0.18$. The un-normalized plausibility probability function $Pl\_P'_m$ is as follows: $Pl\_P'_m(h_1) = 0.66$, $Pl\_P'_m(h_2) = 0.66$, $Pl\_P'_m(h_3) = 0.66$, $Pl\_P'_m(h_4) = 0.66$, $Pl\_P'_m(h_5) = 0.50$, $Pl\_P'_m(h_6) = 0.60$. Notice that h_1, h_2, h_3, and h_4 are the most plausible elements of the state space.

Let m^∞ denote $Lim_{n\to\infty}\, m^n$, let $(BetP_m)^\infty$ denote $Lim_{n\to\infty}\, (BetP_m)^n$, and let $(Pl\_P_m)^\infty$ denote $Lim_{n\to\infty}\, (Pl\_P_m)^n$. m^∞ is as follows: $m^\infty(\{h_1\}) = m^\infty(\{h_2, h_3, h_4\}) = 0.5$. $(BetP_m)^\infty$ is as follows: $(BetP_m)^\infty(h_1) = 1$, $(BetP_m)^\infty(h_2) = (BetP_m)^\infty(h_3) = (BetP_m)^\infty(h_4) = (BetP_m)^\infty(h_5) = (BetP_m)^\infty(h_6) = 0$. $(Pl\_P_m)^\infty$ is as follows: $(Pl\_P_m)^\infty(h_1) = (Pl\_P_m)^\infty(h_2) = (Pl\_P_m)^\infty(h_3) = (Pl\_P_m)^\infty(h_4) = 0.25$, $(Pl\_P_m)^\infty(h_5) = (Pl\_P_m)^\infty(h_6) = 0$. Thus we conclude that Pl_P_m is equivalent to m and that $BetP_m$ isn't. This example is another illustration of the inconsistency of the pignistic transformation with Dempster's rule of combination. ■

Example 4: **Target Identification Problem** [Delmotte and Smets 2001]. A target identification system is composed of 30 sensors, S_i, i = 1, …, 30. Each sensor S_i is in one of two states x_i or y_i. The state of the sensors depends on an unknown target that is assumed to be in one of two states: t_1 denoting friend, or t_2 denoting foe. The state of each sensor also depends on whether it is working or not. When in working condition, a sensor reading of x_i correctly identifies a target of type t_1 and a sensor reading of y_i correctly identifies a target of type t_2. When a sensor is not in working condition, nothing is known about the relationship between the sensor reading and the actual target type. The 30 sensors are of two types, high quality and low quality. A high quality sensor has a 99% probability of being in working condition whereas a low quality sensor has only a 90% probability of being in working condition. Also, the first 11 sensors S_1, …, S_{11} are high quality sensors, and the remaining 19 sensors S_{12}, …, S_{30} are low quality sensors. Data in the form of sensor readings is collected as follows: x_1, …, x_{10}, y_{11}, x_{12}, y_{13}, …, y_{30}. What conclusions can we draw about the actual target type?

Figure 4.1 depicts the problem as a valuation network consisting of one variable (Target) and 30 independent pieces of evidence (the 30 sensors). First, we will represent the evidence from the 30 sensors by bpa's and compute the joint belief function for T. Next, we will represent the evidence by probability functions using the pignistic transformation and compute the joint probability function for T. Finally, we will represent the evidence by probability functions obtained using the plausibility transformation and compute the joint probability function for T.

Figure 4.1: A Valuation Network Model for the Target Identification Problem

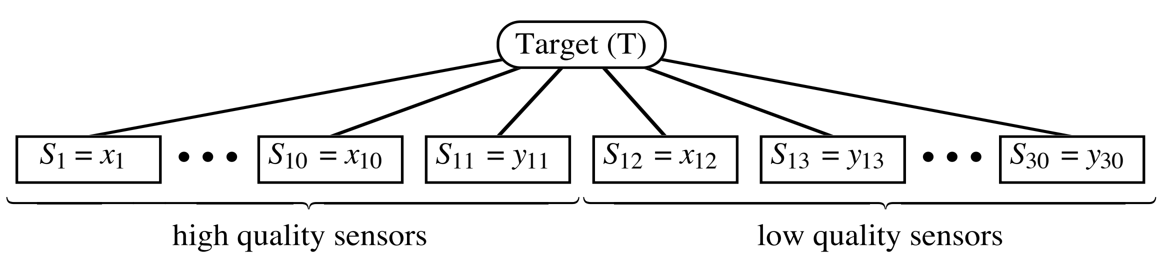


Table 4.1 shows the data collected from the sensors represented as evidence in bpa's. We can reach a conclusion about the target identity by calculating the joint bpa for the 30 sensors, which in this case amounts to a marginal bpa for the variable T. Using Dempster's rule, the joint bpa m is given by $m = m_1 \oplus \ldots \oplus m_{30}$.

Table 4.1: Bpa Encoding of Sensor Readings

Sensor $S_i = x_i$, $i = 1, \ldots, 10$

$a \subseteq \Omega_T$	$m_i(a)$
$\{t_1\}$	0.99
$\{t_1, t_2\}$	0.01

Sensor $S_{11} = y_{11}$

$a \subseteq \Omega_T$	$m_{11}(a)$
$\{t_2\}$	0.99
$\{t_1, t_2\}$	0.01

Sensor $S_{12} = x_{12}$

$a \subseteq \Omega_T$	$m_{12}(a)$
$\{t_1\}$	0.90
$\{t_1, t_2\}$	0.10

Sensor $S_i = y_i$, $i = 13, \ldots, 30$

$a \subseteq \Omega_T$	$m_i(a)$
$\{t_2\}$	0.90
$\{t_1, t_2\}$	0.10

This series of combinations leads to the joint bpa's (un-normalized and normalized) listed in Table 4.2 along with the corresponding plausibility function. Much of the evidence from the sensors conflicts, so the un-normalized bpa places very little mass on each non-empty subset.

Table 4.2: The Joint Bpa's and Plausibility Functions for 30 Sensors

$a \in 2^{\Omega_T}$	Un-normalized bpa	Normalized bpa (m)	Plausibility (Pl_m)
Ø	≈1	0	0
$\{t_1\}$	$\approx 1.00 \times 10^{-20}$	≈ 0.9090	≈ 0.9091
$\{t_2\}$	$\approx 1.00 \times 10^{-21}$	≈ 0.0909	≈ 0.0909
$\{t_1, t_2\}$	$\approx 1.00 \times 10^{-41}$	≈ 0.0000	1

Thus as per the belief function model, the target is approximately 10 times more likely to be a friend than a foe. Next, we will model this problem with probabilities using the pignistic transformations of the 30 belief functions. The probability functions are shown in Table 4.3.

If we combine the 30 probability functions using pointwise multiplication and normalize the resulting probability function, we obtain the result shown in Table 4.4.

Table 4.3: Pignistic Probability Function Encoding of Sensor Readings

Sensor $S_i = x_i$, $i = 1, \ldots, 10$

$x \in \Omega_T$	$BetP_{m_i}(x)$
t_1	0.995
t_2	0.005

Sensor $S_{11} = y_{11}$

$x \in \Omega_T$	$BetP_{m_{11}}(x)$
t_1	0.005
t_2	0.995

Sensor $S_{12} = x_{12}$

$x \in \Omega_T$	$BetP_{m_{12}}(x)$
t_1	0.95
t_2	0.05

Sensor $S_i = y_i$, $i = 13, \ldots, 30$

$x \in \Omega_T$	$BetP_{m_i}(x)$
t_1	0.05
t_2	0.95

Table 4.4. The Joint Pignistic Probability Model for the Target Identification Problem

$x \in \Omega_T$	Un-normalized Probability	Normalized Probability
t_1	≈ 1.723E–26	≈ 0.0820
t_2	≈ 1.930E–25	≈ 0.9180
Sum	≈ 2.102E–25	1

Notice that the pignistic probability model of the target identification problem is qualitatively different from the belief function model. As per the pignistic probability model, the probability that the target is a foe is approximately 11 times more likely than the probability that the target is a friend. In general, if m_1 and m_2 are two bpa's, then $BetP_{m_1} \otimes BetP_{m_2} \neq BetP_{m_1 \oplus m_2}$.

Next, consider the probability model for the target identification problem obtained from the belief function model using the plausibility transformation. This model is shown in Table 4.5.

If we combine the 30 plausibility probability functions using pointwise multiplication and normalize the resulting probability function, we obtain the result shown in Table 4.6 below.

Notice that the result is similar to the belief function model obtained in Table 4.2. In the next section, we will prove that this equivalence between the belief function model conclusion and plausibility probability function is always true. ■

Table 4.5: Plausibility Probability Function Encoding of Sensor Readings

Sensor $S_i = x_i$, $i = 1, \ldots, 10$

$x \in \Omega_T$	$Pl\_P_{m_i}(x)$
t_1	0.9901
t_2	0.0099

Sensor $S_{11} = y_{11}$

$x \in \Omega_T$	$Pl\_P_{m_{11}}(x)$
t_1	0.0099
t_2	0.9901

Sensor $S_{12} = x_{12}$

$x \in \Omega_T$	$Pl\_P_{m_{12}}(x)$
t_1	0.9091
t_2	0.0909

Sensor $S_i = y_i$, $i = 13, \ldots, 30$

$x \in \Omega_T$	$Pl\_P_{m_i}(x)$
t_1	0.0909
t_2	0.9091

Table 4.6. The Joint Plausibility Probability Model

$x \in \Omega_T$	Un-normalized Probability	Normalized Probability
t_1	≈ 1.4656E–21	≈ 0.9091
t_2	≈ 1.4656E–22	≈ 0.0909
Sum	≈ 1.6121E–21	1

5. Justification and Properties of the Plausibility Transformation

In all four examples described in the previous section, there is a discrepancy between the pignistic probability function obtained after combining all evidence with Dempster's rule and the pignistic probabilities obtained by using Bayes rule to combine the pignistic probability derived from each individual piece of evidence. Smets [2002] resolves this apparent discrepancy of the pignistic transformation by stating that beliefs are held at the credal level and one only descends to the probability space for decision making at the time a decision has to be made. This is not satisfactory to us. Decision-making is a not a static activity performed at fixed points in time. The decisions we make create new chance variables or influence the distribution of existing chance variables and the distribution of chance variables influences the decisions we make. The process of decision-making is interleaved with the process of uncertain inference—one cannot separate the two activities.

As we stated before, the pignistic transformation satisfies condition (3.2) that is motivated by the upper and lower probability semantics of belief functions. Our position is that the pignistic transformation is the proper transformation method for belief functions interpreted as upper and lower probabilities. In this case, one should not use Dempster's rule of combination, as this rule is incompatible with upper and lower probability semantics of belief functions.

We view probability and belief functions as two uncertainty calculi with the roughly the same expressive power. It shouldn't make a difference which calculus one is using to represent knowledge. One should get roughly the same results regardless of the calculi one is using if the models built using the calculi are equivalent. With this criterion, the plausibility transformation passes the test while the pignistic transformation fails.

Some justifications for the pignistic transformation are given in [Smets and Kennes 1994, Smets 2002]. Here we will give some intuitive justifications for the plausibility transformation.

Haspert [2001] 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 [1982] was among the earliest to discuss decision making with plausibility functions. Appriou [1991] suggests selecting the hypothesis with the maximum plausibility in a decision-making context.

Dempster [1968] states that the upper probability bound (or plausibility) associated with a belief function is the appropriate likelihood function which contains all sample information. Similarly, Halpern and Fagin [1992] 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, $Pl_{Ob}(H_i) = 1 - Bel_{Ob}(H_i^{\mathrm{c}}) = Pr_i(Ob)/c$, 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: $P_B(b) = \Sigma\{P_{A,B}(a, b) \mid a \in \Omega_A\}$ 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 $P_{A \mid b}$ is given by:

$$P_{A \mid b}(a) = P_{A,B}(a, b) / P_B(b) \tag{5.1}$$

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 $m_{A \mid b}$, is given by $(m_{A,B} \oplus m_b)^{\downarrow A}$. Let $Pl_{A \mid b}$ denote the corresponding plausibility function for A. It can be shown [Shafer 1976] that $Pl_{A \mid b}$ is given by:

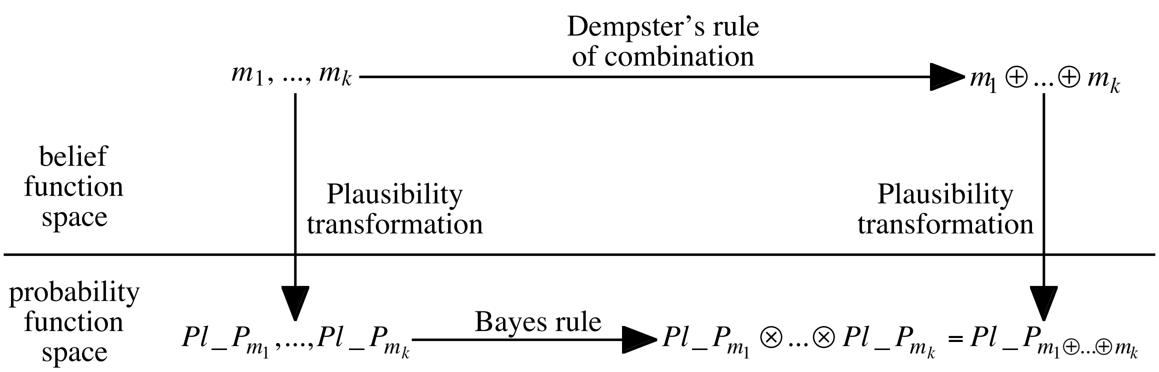
$$Pl_{A \mid b}(\{a\}) = Pl_{A,B}(\{(a, b)\})/Pl_B(\{b\}) \tag{5.2}$$

for all $a \in \Omega_B$. Comparing (5.1) and (5.2) 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 (5.2) could be restated in terms of the Bel function. To provide further justification for the plausibility transformation, we will state the following general theorem from Voorbraak [1989].

Theorem 5.1. 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 $m = m_1 \oplus \ldots \oplus m_k$ is the joint bpa, Pl_m is the associated plausibility function and Pl_P_m is the corresponding plausibility probability function, then $Pl\_P_{m_1} \otimes \ldots \otimes Pl\_P_{m_k} = Pl\_P_m$.

The statement of the theorem is depicted pictorially in Figure 5.1. Voorbraak [1989], who refers to the plausibility transformation as a Bayesian approximation of a belief function, states that combining Bayesian approximations is computationally less involved than combining belief functions. Notice that from a computational perspective, it is much faster to compute $Pl\_P_{m_1} \otimes \ldots \otimes Pl\_P_{m_k}$ than it is to compute Pl_P_m (since the latter involves Dempster's rule of combination and the former involves Bayes rule).

Figure 5.1. A Pictorial Depiction of the Statement of Theorem 5.1.



Theorem 5.1 is significant for several reasons. First, we often create a belief function model and eventually reduce the findings to an equivalent probability function. Assuming the transformation used is the plausibility transformation, Theorem 5.1 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 5.1 tells us that we can create an equivalent probability model and achieve a more tractable result by using Bayes rule. Third, there are many ways to transform a belief function to an equivalent probability function. Given bpa m, unlike Voorbraak [1989], we don't view Pl_P_m as an *approximation* of m. Instead, we view Pl_P_m as an *equivalent probability* encoding of the information in m. Thus if we have a belief function model consisting of $\{m_1, \ldots, m_k\}$, then we view $\{Pl\_P_{m_1}, \ldots, Pl\_P_{m_k}\}$ as an equivalent probability model. Theorem 5.1 can be seen as a regularity condition for any transformation method. Since this condition is not satisfied by the pignistic transformation, one can question the appropriateness of the pignistic transformation as a candidate for finding an equivalent probability model. A corollary of Theorem 5.1 is that Pl_P_m is idempotent if m is idempotent.

Corollary 5.2. 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., $Pl\_P_m \otimes Pl\_P_m = Pl\_P_m$.

To demonstrate that the plausibility transformation is consistent with Dempster's rule of combination, we consider another property of belief functions. 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 $y \in \Omega_s \backslash \{x\}$, where P^n denotes $P \otimes \ldots \otimes P$ (n times). Belief functions have a similar property, as stated in the following theorem.

Theorem 5.3. 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 $y \in \Omega_s \backslash \{x\}$. Let m^n denote $m \oplus \ldots \oplus m$ (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, an equivalent probability function should have x as its most probable state. This property is of course satisfied for the plausibility transformation, as stated in the following corollary.

Corollary 5.4. 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 $y \in \Omega_s \setminus \{x\}$. 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 $y \in \Omega_s \setminus \{x\}$.

In Example 2 presented in Section 4, $m^{500}(\{h_2\}) \approx 1$, so the most plausible hypothesis in m is h_2, consistent with Pl_P_m and not $BetP_m$.

In Example 3 described earlier, the belief function m has no unique most plausible state x. Instead, we have four most plausible states h_1, h_2, h_3, and h_4. In probability theory, if P is such that $\boldsymbol{t} \subseteq \Omega_s$ 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 $x, y \in \boldsymbol{t}$, and $P^\infty(z) = 0$ for all $z \in \Omega_s \setminus \boldsymbol{t}$. Belief functions have a similar property, as stated in the following theorem.

Theorem 5.5. Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $\boldsymbol{t} \subseteq \Omega_s$ is a subset of most plausible states, i.e., $Pl_m(\{\mathrm{x}\}) = Pl_m(\{y\})$ for all $x, y \in \boldsymbol{t}$, and $Pl_m(\{x\}) > Pl_m(\{z\})$ for all $x \in \boldsymbol{t}$, and $z \in \Omega_s \setminus \boldsymbol{t}$. Let m^∞ denote $Lim_{n\to\infty} m^n$, and let Pl_{m^∞} be the corresponding plausibility function. Then there exists a partition $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ of $\boldsymbol{t}$ such that $m^\infty(\boldsymbol{a}_i) = 1/k$ for $i = 1, \ldots, k$, i.e., $Pl_{m^\infty}(x) = Pl_{m^\infty}(y) = 1/k$ for all $x, y \in \boldsymbol{t}$, and $Pl_{m^\infty}(z) = 0$ for all $z \in \Omega_s \setminus \boldsymbol{t}$.

Theorem 5.5 is a generalization of Theorem 5.3 in the sense that if $|\boldsymbol{t}| = 1$, then Theorem 5.5 reduces to Theorem 5.3. The following corollary generalizes the result in Corollary 5.4 for the case of non-unique most plausible states.

Corollary 5.6. Consider a bpa m for s (with corresponding plausibility function Pl_m) such that $\boldsymbol{t} \subseteq \Omega_s$ is a subset of most plausible states, i.e., $Pl_m(\{x\}) = Pl_m(\{y\})$ for all $x, y \in \boldsymbol{t}$, and $Pl_m(\{x\}) > Pl_m(\{z\})$ for all $x \in \boldsymbol{t}$ and $z \in \Omega_s \setminus \boldsymbol{t}$. 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 $(Pl\_P_m)^{\infty}(x) = (Pl\_P_m)^{\infty}(y) = 1/|\,\boldsymbol{t}\,|$ for all $x, y \in \boldsymbol{t}$, and $(Pl\_P_m)^{\infty}(z) = 0$ for all $z \in \Omega_s \setminus \boldsymbol{t}$.

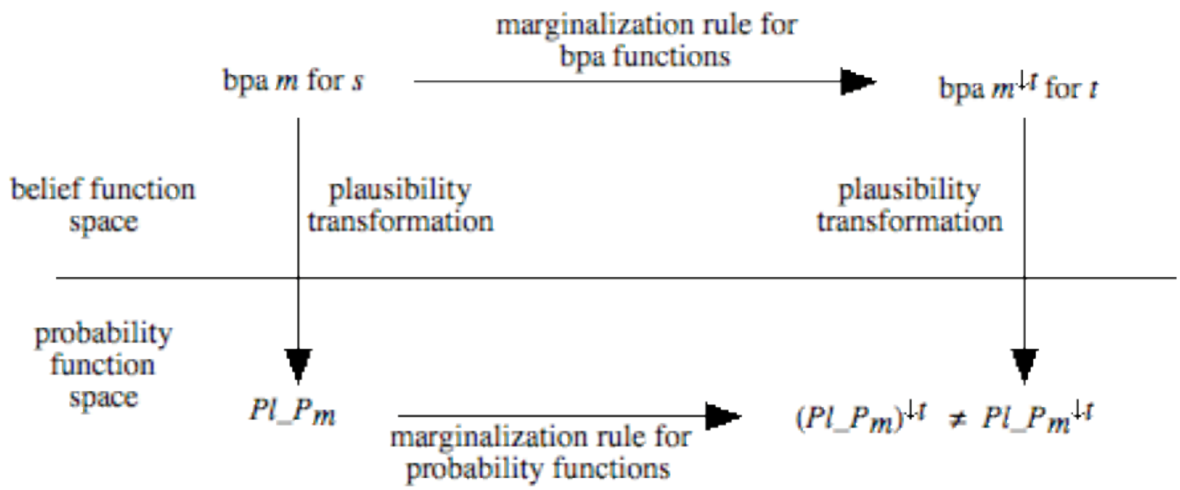
A useful method of transforming a belief function to an equivalent probability function should have properties that allow the transformed model to produce consistent and efficient results when variables in a network are evaluated. In general, combination of marginals in a D-S belief network is accomplished with local computation using two operations: combination and marginalization [Shenoy and Shafer 1990, Shenoy 1997]. Combination corresponds to the aggregation of knowledge, whereas marginalization corresponds to the focusing of knowledge. Suppose m is a bpa for s, and suppose $t \subset s$. The marginal of m for t, denoted $m^{\downarrow t}$, is the bpa for t defined as follows:

$$m^{\downarrow t}(a) = \sum \{m(b) \mid b^{\downarrow t} = a\} \tag{5.3}$$

for each $a \in \Omega_t$, where $b^{\downarrow t}$ denotes the subset of Ω_t obtained by projecting each element of b to t.

Pointwise multiplication of probabilities (created using the plausibility transformation) provides an alternative method of combination to Dempster's rule. However, this operation cannot be used, in general, when a solution algorithm involves marginalization (see Figure 5.2). In some cases, the plausibility transformation must be applied after the local computation solution algorithm to obtain an equivalent marginal probability function.

Figure 5.2. Plausibility Probability Transformation is not Preserved under Marginalization



As an example of the inconsistency depicted in Figure 5.2, consider the following bpa on the domain $\{V, G\}$:

$$m_{V\text{-}G}(\{(v_1, g_1), (v_1, g_2)\}) = 0.6$$
$$m_{V\text{-}G}(\{(v_1, g_1), (v_2, g_1)\}) = 0.3$$
$$m_{V\text{-}G}(\{(v_1, g_1), (v_1, g_2), (v_2, g_1), (v_2, g_2), (v_3, g_1), (v_3, g_2)\}) = 0.1$$

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

$$m_{V-G}{}^{\downarrow\{G\}}(\{g_1\}) = 0.3 \qquad Pl_{m_{V-G}{}^{\downarrow G}}(\{g_1\}) = 1.0 \qquad Pl\_P_{m_{V-G}{}^{\downarrow G}}(g_1) = 1.0/1.7 = 0.588$$
$$m_{V-G}{}^{\downarrow\{G\}}(\{g_1, g_2\}) = 0.7 \qquad Pl_{m_{V-G}{}^{\downarrow G}}(\{g_2\}) = 0.7 \qquad Pl\_P_{m_{V-G}{}^{\downarrow G}}(g_2) = 0.7/1.7 = 0.412$$

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

$$Pl_{m_{V-G}}(\{(v_1, g_1)\}) = 1.0 \qquad Pl_{m_{V-G}}(\{(v_2, g_1)\}) = 0.4 \qquad Pl_{m_{V-G}}(\{(v_3, g_1)\}) = 0.1$$
$$Pl_{m_{V-G}}(\{(v_1, g_2)\}) = 0.7 \qquad Pl_{m_{V-G}}(\{(v_2, g_2)\}) = 0.1 \qquad Pl_{m_{V-G}}(\{(v_3, g_2)\}) = 0.1$$

$$Pl\_P_{m_{V-G}}(v_1, g_1) = 0.417 \qquad Pl\_P_{m_{V-G}}(v_2, g_1) = 0.167 \qquad Pl\_P_{m_{V-G}}(v_3, g_1) = 0.042$$
$$Pl\_P_{m_{V-G}}(v_1, g_2) = 0.292 \qquad Pl\_P_{m_{V-G}}(v_2, g_2) = 0.042 \qquad Pl\_P_{m_{V-G}}(v_3, g_2) = 0.042$$

Marginalizing this probability function to G gives:

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

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.

Theorem 5.7. 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. $r_i \cap r_j = \varnothing$ for all $i \neq j$. Let m denote $m_1 \oplus \ldots \oplus m_k$. Then, $Pl\_P_{m^{\downarrow t}} = Pl\_P_{m_1^{\downarrow t}} \otimes \ldots \otimes Pl\_P_{m_k^{\downarrow t}}$.

The following theorem allows us to find the plausibility function for a marginal bpa without calculating the marginal bpa.

Theorem 5.8. Suppose m is a bpa for s and $t \subseteq s$. Then,

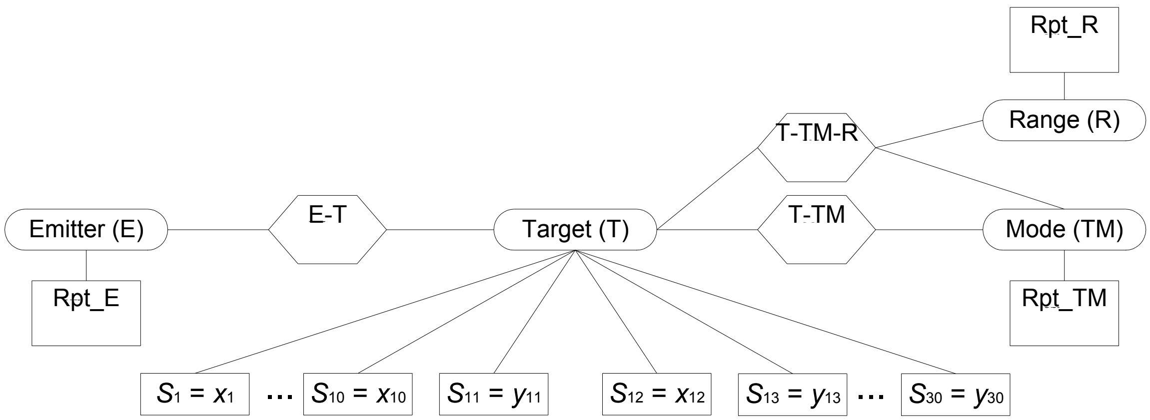
$$Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a} \cap \boldsymbol{c}^{\downarrow t} \neq \varnothing} m(\boldsymbol{c})$$

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

Theorem 5.1 shows that combination of bpa's can be accomplished using Bayes rule when the bpa's are transformed to equivalent probability functions. In a D-S belief network we often perform marginalization after combination. When using the plausibility transformation, some of the combinations can be made after conversion to probabilities and calculation of some marginal bpa's can be avoided because of Theorem 5.7 and 5.8. Next, we solve an example problem using the results of Theorems 5.1, 5.7, and 5.8 to show the computational efficiency of the plausibility probability function.

We will illustrate combination of plausibility probabilities and construction of a marginal probability function from a D-S belief network by considering the example in Figure 5.3. This is a combination of the D-S belief network in Figure 2.2 and the Valuation Network in Figure 4.1 (the 30 sensors problem). The network has 4 variables and 36 valuations. Each variable has two possible states, defined as follows: $\Omega_E = \{e_1, e_2\}$, $\Omega_T = \{t_1, t_2\}$, $\Omega_R = \{r_1, r_2\}$, $\Omega_{TM} = \{tm_1, tm_2\}$. The goal of the example is to efficiently create a marginal probability mass function for T.

Figure 5.3. The D-S Belief Network for the Modified Anti-Air Threat Identification Problem



The set of valuations—$\{m_1, \ldots, m_{36}\}$— in the network is defined in Table 5.1 (see Table 4.1 for the definitions corresponding to $m_1, \ldots, m_{30}$).

A probability distribution for T can be found by computing the marginal of the joint for T, i.e. $(m_1 \oplus \ldots \oplus m_{36})^{\downarrow T}$ using Dempster's rule, then using the plausibility transformation to create plausibility probabilities. The marginals of this joint bpa can be found using local computation axioms [Shenoy and Shafer 1990, Shenoy 1997] and selecting an appropriate deletion sequence using one of several available heuristics [Kong 1986]. One appropriate deletion sequence is R, E, TM. As a result of Theorem 5.7, however, these same axioms can be used to simplify the computation, given that the desired outcome is a probability mass function for T.

Table 5.1. Definition of Bpa's in the D-S Belief Network of Figure 5.3

$a \in 2^{\Omega_E}$	$m_{31}(a)$
$\{e_1\}$	0.2
$\{e_2\}$	0.7
Ω_E	0.1

$a \in 2^{\Omega_{\{E, T\}}}$	$m_{32}(a)$
$\{(e_1, t_1)\}$	0.05
$\{(e_1, t_2), (e_2, t_1)\}$	0.05
$\{(e_1, t_1), (e_1, t_2), (e_2, t_1)\}$	0.80
$\Omega_{\{E, T\}}$	0.10

$a \in 2^{\Omega_{\{T, TM\}}}$	$m_{33}(a)$
$\{(t_1, tm_1)\}$	0.2
$\{(t_1, tm_1), (t_2, tm_2)\}$	0.2
$\{(t_1, tm_1), (t_2, tm_1), (t_2, tm_2)\}$	0.2
$\Omega_{\{T, TM\}}$	0.4

$a \in 2^{\Omega_{\{T, TM, R\}}}$	$m_{34}(a)$
$\{(t_1, tm_1, r_1)\}$	0.4
$\{(t_2, tm_2, r_2)\}$	0.1
$\Omega_{\{T, TM, R\}}$	0.5

$a \in 2^{\Omega_{TM}}$	$m_{35}(a)$
$\{tm_1\}$	0.01
$\{tm_2\}$	0.70
Ω_{TM}	0.29

$a \in 2^{\Omega_R}$	$m_{36}(a)$
$\{r_1\}$	0.1
$\{r_2\}$	0.6
Ω_R	0.3

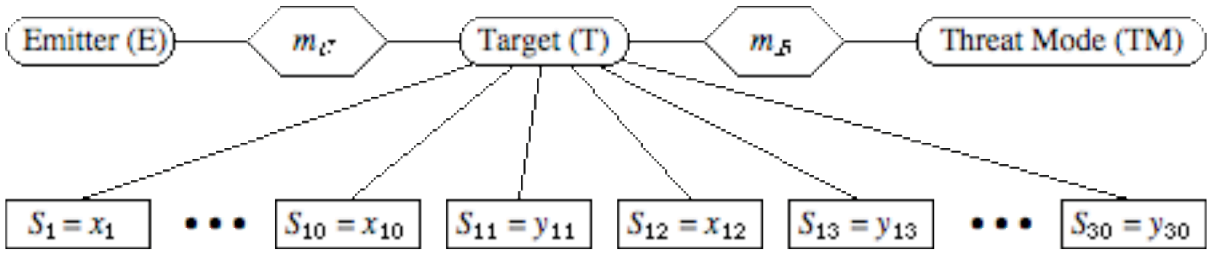
The solution proceeds by first calculating $m_A = (m_{34} \oplus m_{36})^{\downarrow\{T, TM\}}$. Next, we calculate $m_B = m_A \oplus m_{33} \oplus m_{35}$. Subsequently, we calculate $m_C = m_{31} \oplus m_{32}$. The results of these calculations are shown in Table 5.2 (ignoring normalization). The D-S belief network appears as in Figure 5.4 after these operations.

Table 5.2. Result of First Two Calculations in the Plausibility Probability Solution

$a \in 2^{\Omega_{\{E,T\}}}$	$m_C(a)$
$\{(e_1, t_1)\}$	0.015
$\{(e_1, t_2)\}$	0.010
$\{(e_2, t_1)\}$	0.595
$\{(e_1, t_1), (e_1, t_2)\}$	0.180
$\{(e_1, t_2), (e_2, t_1)\}$	0.005
$\{(e_2, t_1), (e_2, t_2)\}$	0.070
$\{(e_1, t_1), (e_1, t_2), (e_2, t_1)\}$	0.080
$\Omega_{\{E, T\}}$	0.010

$a \in 2^{\Omega_{\{T,TM\}}}$	$m_B(a)$
$\{(t_1, tm_1)\}$	0.0790
$\{(t_2, tm_2)\}$	0.2113
$\{(t_1, tm_1), (t_2, tm_1)\}$	0.0030
$\{(t_1, tm_1), (t_2, tm_2)\}$	0.0290
$\{(t_1, tm_2), (t_2, tm_2)\}$	0.1400
$\{(t_1, tm_1), (t_2, tm_1), (t_2, tm_2)\}$	0.0290
$\Omega_{\{T, TM\}}$	0.0580

Figure 5.4. D-S Belief Network after Deletion of Variables *E* and *R*



The set of bpa's—all of which are defined on T—remaining after previous combinations is $\sigma = \{m_B, m_C, m_1, \ldots, m_{30}\}$. Define s_i as the domain of bpa m_i. In this instance, $s_B = \{T, TM\}$, $s_C = \{E, T\}$, $s_1 \ldots s_{30} = \{T\}$; thus, $t_i = (s_i - \{T\})$ is pairwise disjoint for all $i = B, C, 1, \ldots, 30$ so Theorem 5.7 can be used to calculate the marginal probability function for T. To compute the marginal for T, we calculate the following plausibility values: $Pl_{m_B^{\downarrow\{T\}}}, Pl_{m_C^{\downarrow\{T\}}}, Pl_{m_1}, \ldots, Pl_{m_{30}}$. Theorem 5.8 can be used to calculate the plausibilities of $m_B{}^{\downarrow T}$ and $m_C{}^{\downarrow T}$. The plausibility values are combined using pointwise multiplication, then normalized to create the plausibility probability function. Results are shown in Table 5.3. The plausibility probabilities developed using combinations made possible by Theorems 5.1 and 5.7 are the same as those determined by calculating the marginal belief function for T with Dempster's rule. The probability of the target being type t_1 is 0.951, whereas the probability of the target being type t_2 is 0.049.

Table 5.3. Final Calculation of the Plausibility Probability Function

$x \in \Omega_T$	$Pl_{m_B{}^{\downarrow T}}(\{x\})$	$Pl_{m_C{}^{\downarrow T}}(\{x\})$	$Pl_{m_1} \otimes \ldots \otimes Pl_{m_{30}}$	Un-normalized Product	$Pl\_P_m(x)$
t_1	0.955	0.33800	≈ 1.00 x 10^{-20}	≈ 3.228E-21	0.951
t_2	0.355	0.47028	≈ 1.00 x 10^{-21}	≈ 1.669E-22	0.049
Sum				≈ 3.395E-21	1.000

These results demonstrate the computational efficiency that characterizes the plausibility transformation. The answers presented are the same as those found by calculating the joint bpa for all variables in the network, marginalizing the joint bpa, then using the plausibility transformation. In this sense, the answers to these examples are exact.

6. Conclusions and Summary

In summary, if T transforms a bpa m in a belief function model to an equivalent probability function $T(m)$, T should satisfy four basic properties:

1) *Invariance with respect to combination*: $T(m_1 \oplus \ldots \oplus m_n) = T(m_1) \otimes \ldots \otimes T(m_n)$, which is satisfied for the plausibility transformation, according to Theorem 5.1;
2) *Idempotency*: If m is idempotent, then $T(m)$ is idempotent; which is satisfied by the plausibility probability transformation according to Corollary 5.2;
3) *Unique most plausible state*: If $Lim_{n \to \infty}\, m^n(h_i) = 1$, then $Lim_{n \to \infty}\, T^n(m)(h_i) = 1$; which is satisfied for the plausibility transformation according to Corollary 5.4; and
4) *Non-unique most plausible states*: If $Lim_{n \to \infty}\, Pl_{m^n}(x) = Lim_{n \to \infty}\, Pl_{m^n}(y)$ for all $x, y \in \boldsymbol{t} \subseteq \Omega_s$, and $Lim_{n \to \infty}\, Pl_m{}^n(z) = 0$,for all $z \in \Omega_s \setminus \boldsymbol{t}$, then $Lim_{n \to \infty}\, T^n(m)(x) = Lim_{n \to \infty}\, T^n(m)(y)$ for all $x, y \in \boldsymbol{t}$, and $Lim_{n \to \infty}\, T^n(m)(z) = 0$ for all $z \in \Omega_s \setminus \boldsymbol{t}$; this property is satisfied for the plausibility transformation according to Corollary 5.6.

We notice that the pignistic transformation does not satisfy any of these properties. By satisfying these four properties, the plausibility transformation method results in an equivalent probability function that is consistent with Dempster-Shafer belief function theory that has Dempster's rule of combination as a central concept.

Until now, most of the literature on belief functions have used the so-called pignistic method for transforming belief function models to equivalent probability models. We believe this method is inappropriate, as it is inconsistent with Dempster's rule of combination. We conjecture that the plausibility transformation method is the only method that satisfies these axioms, but we don't have a proof of this claim.

Acknowledgements

The research was partly funded by a contract from Sparta, Inc., to the second author. We are grateful for extensive comments from Philippe Smets, Rolf Haenni, Art Dempster, and Yang

Chen on earlier drafts of this paper. We thank the participants of the ECSQARU 2003 conference for valuable comments, suggestions, and discussions.

Appendix: Proofs

Proof of Theorem 5.1. The proof of Theorem 5.1 follows directly from the proof of Proposition 2 in [Voorbraak 1989]. 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 are equivalent for singleton subsets. ■

Proof of Corollary 5.2: Follows immediately from the statement of Theorem 5.1. ■

Proof of Theorem 5.3. If $Pl_m(\{x\}) > Pl_m(\{y\})$, for all $y \in \Omega_s \backslash \{x\}$, then $Pl_m(\{x\}) \cdot Pl_m(\{x\}) > Pl_m(\{y\}) \cdot Pl_m(\{y\})$, for all $y \in \Omega_s \backslash \{x\}$.

Using the definition of plausibility:

$$\left[\sum_{\substack{q \\ y_q \cap \{x\} \neq \varnothing}} m(y_q) \cdot \sum_{\substack{r \\ z_r \cap \{x\} \neq \varnothing}} m(z_r)\right] > \left[\sum_{\substack{q \\ y_q \cap \{y\} \neq \varnothing}} m(y_q) \cdot \sum_{\substack{r \\ z_r \cap \{y\} \neq \varnothing}} m(z_r)\right]$$ which can be re-written as:

$$\left[\sum_{\substack{q,r \\ y_q \cap \{x\} \neq \varnothing \\ z_r \cap \{x\} \neq \varnothing}} m(y_q) \cdot m(z_r)\right] > \left[\sum_{\substack{q,r \\ y_q \cap \{y\} \neq \varnothing \\ z_r \cap \{y\} \neq \varnothing}} m(y_q) \cdot m(z_r)\right]$$. Some of the intersections in these terms are singletons while some may contain subsets with more than one state. Thus, the above statement can be re-written as:

$$\left[\sum_{\substack{q,r \\ y_q \cap \{x\} = \{x\} \\ z_r \cap \{x\} = \{x\}}} m(y_q) \cdot m(z_r)\right] + \left[\sum_{\substack{q,r \\ \{x\} \subset y_q \cap \{x\} \\ \{x\} \subset z_r \cap \{x\}}} m(y_q) \cdot m(z_r)\right] > \left[\sum_{\substack{q,r \\ y_q \cap \{y\} = \{y\} \\ z_r \cap \{y\} = \{y\}}} m(y_q) \cdot m(z_r)\right] + \left[\sum_{\substack{q,r \\ \{y\} \subset y_q \cap \{y\} \\ \{y\} \subset z_r \cap \{y\}}} m(y_q) \cdot m(z_r)\right]$$

This statement can now be re-written with a normalization constant as:

$$K^{-1}\left(\left[\sum_{\substack{q,r\\ y_q\cap z_r=\{x\}}} m(y_q)\cdot m(z_r)\right]+\left[\sum_{\substack{q,r\\ \{x\}\subset y_q\cap z_r}} m(y_q)\cdot m(z_r)\right]\right)>K^{-1}\left(\left[\sum_{\substack{q,r\\ y_q\cap z_r=\{y\}}} m(y_q)\cdot m(z_r)\right]+\left[\sum_{\substack{q,r\\ \{y\}\subset y_q\cap z_r}} m(y_q)\cdot m(z_r)\right]\right)$$

Using Dempster's rule to simplify the left-hand term inside the outer parentheses yields:

$$(m\oplus m)(x)+\left[\sum_{\substack{q,r\\ \{x\}\subset y_q\cap z_r}} m(y_q)\cdot m(z_r)\right]>(m\oplus m)(y)+\left[\sum_{\substack{q,r\\ \{y\}\subset y_q\cap z_r}} m(y_q)\cdot m(z_r)\right]$$

which can be re-written as:

$$(m\oplus m)(x)+\left[\sum_{\{x\}\subset a}(m\oplus m)(a)\right]>(m\oplus m)(y)+\left[\sum_{\{y\}\subset a}(m\oplus m)(a)\right],$$

or $m^2(\{x\})+\left[\sum_{\{x\}\subset a} m^2(a)\right]>m^2(\{y\})+\left[\sum_{\{y\}\subset a} m^2(a)\right]$, for all $y\in\Omega_s\backslash\{x\}$.

Using the same process, it can be shown that if $Pl_m(\{x\}) > Pl_m(\{y\})$,

then $m^k(\{x\})+\left[\sum_{\{x\}\subset a} m^k(a)\right]>m^k(\{y\})+\left[\sum_{\{y\}\subset a} m^k(a)\right]$,

for all $y\in\Omega_s\backslash\{x\}$, and for any finite k. Each time Dempster's rule is performed and singleton subsets are included, the normalized mass moves from non-singleton subsets to singleton subsets; thus, as k becomes larger and larger, $\sum_{\{x\}\subset a} m^k(a)$ and $\sum_{\{y\}\subset a} m^k(a)$ become smaller and smaller.

Suppose $m^k = m\oplus\ldots\oplus m$ (k times). For some finite number k, if $Pl_m(\{x\}) > Pl_m(\{y\})$, then $m^k(\{x\}) > m^k(\{y\})$, $\sum_{\{x\}\subset a} m^k(a)\approx 0$, and $\sum_{\{y\}\subset a} m^k(a)\approx 0$, for all $y\in\Omega_s\backslash\{x\}$, thus $m^k(\{x_1\}) + m^k(\{x_2\})+\ldots+ m^k(\{x_t\}) \approx 1$, i.e. all mass in m^k moves to the singleton focal elements.

Thus, by continuing to use Dempster's rule with normalization to calculate $m^n = m^k\oplus m^{n-k}$, if $Pl_m(\{x\}) > Pl_m(\{y\})$, for all $y\in\Omega_s\backslash\{x\}$, then $m^\infty(\{x\}) = 1$. If $m^\infty(\{x\}) = 1$, it follows that $Pl_{m^\infty}(\{x\}) = 1$, which proves the theorem. ■

Proof of Corollary 5.4: Follows immediately from Theorem 5.3 and the definition of Pl_P_m in (3.2). ■

Proof of Theorem 5.5: If $Pl_m(\{x\}) = Pl_m(\{y\})$, for all $x, y \in \boldsymbol{t}$, and $Pl_m(\{x\}) > Pl_m(\{z\})$, for all $x \in \boldsymbol{t}$ and $z \in \Omega_s \setminus \boldsymbol{t}$, then recall from the proof of Theorem 5.3 that for some finite number k, if $Pl_m(\{x\}) > Pl_m(\{z\})$, then $m^k(\{x\}) > m^k(\{z\})$ and $Pl_m(\{x\}) \cdot Pl_m(\{x\}) = Pl_m(\{y\}) \cdot Pl_m(\{y\})$, for all $y \in \Omega_s \setminus \{x\}$.

Note that by replacing inequality with equaility in the proof of Theorem 5.3, it follows that if $Pl_m(\{x\}) \cdot Pl_m(\{x\}) = Pl_m(\{y\}) \cdot Pl_m(\{y\})$, then

$$m^k(\{x\}) + \left[\sum_{\{x\} \subset a} m^k(a)\right] = m^k(\{y\}) + \left[\sum_{\{y\} \subset a} m^k(a)\right],$$

for all for all $x, y \in \boldsymbol{t}$ and for any finite k. Define $\boldsymbol{b}_i \subseteq \Omega_s$ and $\boldsymbol{c}_i \subseteq \Omega_s$ as subsets of the state space of s containing i states where $x \subseteq \boldsymbol{b}_i$ and $y \subseteq \boldsymbol{c}_i$. The previous statement can now be re-written as

$m^k(b_1) + m^k(b_2) + \ldots + m^k(b_n) = m^k(c_1) + m^k(c_2) + \ldots + m^k(c_n)$. Each time Dempster's rule is performed the normalized mass moves to smaller and smaller subsets of the state space.

Thus, by continuing to use Dempster's rule with normalization to calculate $m^n = m^k \oplus m^{n-k}$, if $Pl_m(\{x\}) = Pl_m(\{y\})$, for all $x, y \in \boldsymbol{t}$, then $m^n(\boldsymbol{b}_j) = m^n(\boldsymbol{c}_k)$, where $\boldsymbol{b}_j$ and $\boldsymbol{c}_k$ are the subsets with the minimal number of states that contain x and y, respectively. It follows from the proof of Theorem 5.3 that for all $z \in \Omega_s \setminus \boldsymbol{t}$, $m^n(\boldsymbol{d}) = 0$ for all subsets such that $z \subseteq \boldsymbol{d}$. Thus, $\boldsymbol{b}_j$ is one of k members of a partition $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ of $\boldsymbol{t}$ such that $m^\infty(\boldsymbol{a}_i) = 1/k$, which proves the theorem. ■

Proof of Corollary 5.6: Follows immediately from Theorem 5.5 and the definition of Pl_P_m in (3.2). ■

The following proposition is a simpler version of Theorem 5.7. We will use it to prove Theorem 5.7.

Proposition 5.7. 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. $r_1 \cap r_2 = \varnothing$. Then, $Pl\_P_{(m_1 \oplus m_2)^{\downarrow t}} = Pl\_P_{m_1^{\downarrow t}} \otimes Pl\_P_{m_2^{\downarrow t}}$.

Proof of Proposition 5.7: It follows from the axioms proposed by Shenoy and Shafer [1986, 1990] 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 5.1 by substituting $m_1^{\downarrow t}$ for m_1 and $m_2^{\downarrow t}$ for m_2. ■

Proof of Theorem 5.7. The proof of Theorem 5.7 follows directly from the proof of Proposition 5.7. ■

Proof of Theorem 5.8: The marginal bpa of m for t is defined as $m^{\downarrow t}(a) = \sum\{m(b) \mid b^{\downarrow t} = a\}$ for all $a \subseteq \Omega_t$. The plausibility function values of the marginal bpa of m for t are defined as $Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a} \cap \boldsymbol{d} \neq \varnothing} m^{\downarrow t}(\boldsymbol{d})$ for each $\boldsymbol{a} \subseteq \Omega_t$. This formula can be rewritten as $Pl_{m^{\downarrow t}}(\boldsymbol{a}) = \sum_{\boldsymbol{a} \cap \boldsymbol{c}^{\downarrow t} \neq \varnothing} m(\boldsymbol{c})$ for each $a \subseteq \Omega_t$, which proves the theorem. ■

Bibliography

Appriou, A. 1991. Probabilitiés et incertitude en fusion de données multisenseurs. *Revue Scientifique de la Défense*, **11** 27–40.

Baroni, P., P. Vicig. 2003. Transformations from imprecise to precise probabilities. In: T. D. Nielsen, N. L. Zhang, eds. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty,* Lecture Notes in Artificial Intelligence, **2711** 37–49.

Black, P. K., K. B. Laskey. 1990. Hierarchical evidence and belief functions. In: R. D. Shachter, T. S. Levitt, J. F. Lemmer, L. F. Kanal, eds. *Uncertainty in Artificial Intelligence,* **4** 207–215.

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

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

Daniel, M. 2003. Transformations of belief functions to probabilities. In: J. Vejnarová, ed. *Proceedings of the 6th Workshop on Uncertainty Processing in Expert Systems*, 77–90.

Delmotte, F., P. Smets. 2001. Target identification based on the transferable belief model interpretation of the Dempster-Shafer model, Part II: Applications. Working Paper, LAMIH—Université de Valenciennes, IRIDIA—Université Libre de Bruxelles.

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

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

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

Giang, P. H., P. P. Shenoy. 2003. Making decisions with partially consonant belief functions. In U. Kjærulff and C. Meek (eds.), *Uncertainty in Artificial Intelligence: Proceedings of the Nineteenth Conference*, 272–280, Morgan Kaufmann, San Francisco, CA.

Giles R. 1982. Foundations for a possibility theory. In M. M. Gupta and E. Sanchez (eds.), *Fuzzy Information and Decision Processes*, 83–195, North-Holland, Amsterdam.

Haenni, R., N. Lehmann. 2003. Probabilistic argumentation systems: A new perspective on Dempster-Shafer theory. *International Journal of Intelligent Systems*, **18**(1), 93–106.

Halpern, J. Y., R. Fagin. 1992. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, **54** 275–317.

Haspert, J. K. 2001. Relating Bayesian to Dempster-Shafer analyses. Working Paper, Institute for Defense Analyses.

Jaffray, J. Y. 1989. Linear utility theory for belief functions. *Operation Research Letters*, **8** 107–112.

Kohlas, J., P.A. Monney. 1995. A mathematical theory of hints: An approach to Dempster-Shafer theory of evidence. In: *Lecture Notes in Mathematical Systems*, Springer-Verlag, Berlin, **425**.

Kong, A. 1986. Multivariate belief functions and graphical models. Doctoral dissertation, Department of Statistics, Harvard University.

Spiegelhalter, D. J., A. P. Dawid, S. L. Lauritzen, R. G. Cowell. 1993. Bayesian analysis in expert systems. *Statistical Science*, **8**(3) 219–283.

Shafer, G. 1976. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J.

Shafer, G. 1986. The combination of evidence. *International Journal of Intelligent Systems*, **1** 155–179.

Shafer, G. 1987. Belief functions and possibility measures. In: J. Bezdek, ed. *The Analysis of Fuzzy Information.* CRC Press, Boca Raton, FL*,* 51–84.

Shafer, G., R. Srivastava. 1990. The Bayesian and belief-function formalisms: A general perspective for auditing. In: *Auditing: A Journal of Practice and Theory* (1990). Reprinted in: J. Pearl, G. Shafer, eds. *Readings in Uncertain Reasoning*. Morgan Kaufmann, San Mateo, CA, 482–521.

Shenoy, P. P. 1992. 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, 83–104.

Shenoy, P. P. 1997. Binary join trees for computing marginals in the Shenoy-Shafer architecture. *International Journal of Approximate Reasoning*, **17**(2,3) 239–263.

Shenoy, C., P. P. Shenoy. 2002. Modeling financial portfolios using belief functions. In: R. P Srivastava, T. J. Mock, eds. *Belief Functions in Business Decisions*, Physica–Verlag, Heidelberg, 316–332.

Shenoy, P. P., G. Shafer. 1986. Propagating belief functions with local computations. *IEEE Expert*, **1**(3) 43–52.

Shenoy, P. P., G. Shafer. 1990. 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,* **4** 169-198. Reprinted in: J. Pearl, G. Shafer, eds. *Readings in Uncertain Reasoning*. Morgan Kaufmann, San Mateo, CA, 575–610.

Smets, P. 1990. 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*, **5** 29–40.

Smets, P. 1993b. No Dutch book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. In: R. Scozzafava, ed. *Workshop on Probabilistic Expert Systems*, Societa Italiana di Statistica, Rome, 181–204.

Smets, P. 2002. 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, 316–332.

Smets, P., R. Kennes. 1994. The transferable belief model. *Artificial Intelligence*, **66** 191–234.

Spiegelhalter, D. J., A. P. Dawid, S. L. Lauritzen, R. G. Cowell. 1993. Bayesian analysis in expert systems. *Statistical Science*, **8**(3) 219–283.

Strat, T. 1990. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, **4** 391–417.

Voorbraak, F. 1989. A computationally efficient approximation of Dempster-Shafer theory. *International Journal of Man-Machine Studies*, **30** 525–536.

Walley, P. 1987. Belief function representation of statistical evidence. *The Annals of Statistics*, **15**(4), 1439–1465.