A COMPARISON OF BAYESIAN AND BELIEF FUNCTION REASONING

Barry R. Cobb and Prakash P. Shenoy

University of Kansas School of Business, 1300 Sunnyside Ave, Summerfield Hall, Lawrence, KS 66045-2003, USA.

Abstract

The goal of this paper is to compare the similarities and differences between Bayesian and belief function reasoning. Our main conclusion is that although there are obvious differences in semantics, representations, the rules for combining and marginalizing representations, there are many similarities. We claim that the two calculi have roughly the same expressive power. Each calculus has its own semantics that allow us to construct models suited for these semantics. Once we have a model in either calculus, one can transform it to the other by means of a suitable 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 this paper, we investigate the similarities and the differences between the two calculi.

A Bayesian network is a probability model consisting of a directed acyclic graph representing conditional independence assumptions in the joint distribution [Spiegelhalter *et al.*

1993]. A D-S belief network graphically describes knowledge and the relationships among variables using the so-called Dempster-Shafer theory of belief functions. Numerically, 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. A D-S belief network assigns D-S belief functions or basic probability assignments (bpa's) to subsets of the variables in the domain of each relation. Likelihood functions can be used to update a Bayesian network, while additional evidence entered as bpa's is used to update the D-S belief network.

The graphical representations in each type of network use variables and their respective state spaces. The relationships between these variables encode qualitative conditional independence assumptions of the uncertain domain. Bayesian and D-S networks model similar sets of conditional independence assumptions.

Similarities of Bayesian and D-S belief networks and their underlying calculi have been noted previously. The theory of belief functions captures Bayesian probability models as a special case, so any Bayesian network model can be replicated exactly in a D-S belief network model [Zarley *et al*. 1988]. Similarly, any D-S belief network model can be approximated by a corresponding Bayesian network model [Shafer 1986].

Shafer and Srivastava [1990] argue that the belief-function calculus is a generalization of the Bayesian probability calculus and show that any Bayesian model of uncertainty is also a belief function model. They make a case for using the belief function calculus in the context of auditing due to its greater flexibility, but emphasize that using belief functions does not eliminate the possibility of later using the advantages of propagation associated with Bayesian reasoning. According to their conclusions, belief functions allow non-statistical evidence to be modeled in a way that legitimately represents the underlying knowledge base.

Our basic thesis is that Bayesian and D-S reasoning have "roughly" the same expressive power. We say roughly since we do not have an exact metric to measure the expressiveness of an uncertainty calculus. The two calculi have different semantics. However, given a model in one calculus, it is possible to transform the model to the other and achieve the same qualitative

results. Confirming this thesis will allow these two calculi to be further integrated in decision-making applications. In a related paper [Cobb and Shenoy 2003], we examine methods for transforming a belief function model to an equivalent probability model.

The remainder of this paper is organized as follows. Section 2 describes the representation, semantics and process of making inferences in a Bayesian network. In Section 3, D-S belief network representations of uncertainty and their semantics are reviewed. Section 4 compares and contrasts important facets of Bayesian and D-S belief networks. Finally, Section 5 contains a summary and some conclusions.

2.   Bayesian Networks

Representation. 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, in a sense to be described shortly. 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.

Each variable in the network has a set of mutually exclusive and exhaustive possible values that comprise its state space. If there is a directed arc from variable X to variable Y, we say X is a *parent* of Y and Y is a *child* of X. Based on expert knowledge or empirical observations, a set of conditional probability distributions is specified for each variable, one for each configuration of states of the variable's parents.

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

identification problem. The graph representing probabilistic relationships among the nodes

reveals the conditional independence relations assumed in the network. Each variable in the

network has a set of parents. The parents of each node are listed in Table 2.1 for each variable in

the network of Figure 2.1.

Figure 2.1. A Bayesian Network for the Anti-Air Threat Identification Problem
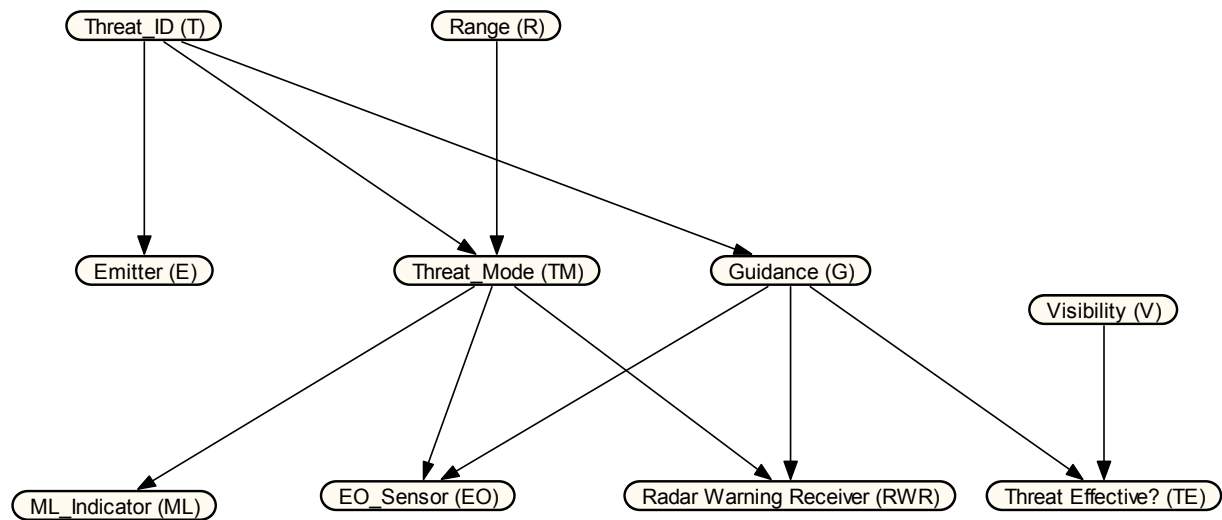
Consider an ordering of the variables such that the variables at the tail of directed arcs

precede variables at the heads of the directed arcs. Since the directed graph is acyclic, such an

ordering always exists. One such ordering in the example network of Figure 2.1 is *T R E TM G V*

ML EO RWR TE. Each variable in a Bayesian network is implicitly assumed to be conditionally

independent of its predecessors in the ordering, given its parents. Table 2.2 summarizes a

minimal set of conditional independence relations assumed in the Bayesian network of Figure

2.1. In summary, missing arcs from a variable to its successors in the ordering imply conditional

independence assumptions in the joint probability distribution represented by the Bayesian

network.

Table 2.1. Network Ordering Relationships for Threat ID Bayesian Network

Variable	Description	# States	Parents
T	Threat ID	7	{Ø}
R	Range	7	{Ø}
E	Emitter	7	{T}
TM	Threat Mode	7	{T, R}
G	Guidance	2	{T}
V	Visibility	3	{Ø}
ML	Missile Launch Indicator	2	{TM}
EO	Electro-Optical Sensor	3	{TM, G}
RWR	Radar Warning Receiver	2	{TM, G}
TE	Threat Effectiveness	2	{G, V}

Table 2.2. One Set of Conditional Independence Assumptions for the Threat ID Bayes Net

R ⊥ T

E ⊥ R | T

TM ⊥ E | {T, R}

G ⊥ {R, E, TM} | T

V ⊥ {T, R, E, TM, G}

ML ⊥ {T, R, E, G, V} | TM

EO ⊥ {T, R, E, V, ML} | {TM, G}

RWR ⊥ {T, R, E, V} | {TM, G}

TE ⊥ {T, R, E, TM, ML, EO, RWR} | {G, V}

As an illustration, if the true state of the Threat ID (*T*) is known, information about the true state of Range (*R*), Threat Mode (*TM*), or Emitter (*E*) gives no additional information regarding the probability distribution of the Guidance (*G*) variable; therefore, *G* is conditionally independent of *R*, *E* and *TM* given *T*. Similarly, Threat Effectiveness (*TE*) is conditionally independent of all variables in the network except its parents: Guidance (*G*) and Visibility (*V*). If there is more than one sequence of variables that is consistent with the directed arcs, then the set of conditional independence relations associated with each sequence can be shown to be

equivalent using the properties of conditional independence relations [Dawid 1979, Pearl 1988, Lauritzen *et al.* 1990]. While the lack of an arc between two nodes represents a conditional independence assumption, the presence of an arc does not necessarily imply dependence.

Conditional probability functions in a Bayesian network can be combined to produce the joint distribution of all variables by applying the multiplicative law of probability. The calculation of the joint distribution in the Bayesian network of Figure 2.1 can be accomplished as follows (where $\otimes$ denotes pointwise multiplication of functions):

$$P(T,R,E,TM,G,V,ML,EO,RWR,TE) =$$
$$P(T) \otimes P(R|T) \otimes P(E|T,R) \otimes P(TM|T,R,E) \otimes P(G|T,R,E,TM\} \otimes P(V|T,R,E,TM,G) \otimes$$
$$P(ML|T,R,E,TM,G,V) \otimes P(EO|T,R,E,TM,G,V,ML) \otimes$$
$$P(RWR|T,R,E,TM,G,V,ML,EO) \otimes P(TE|T,R,E,TM,G,V,ML,RWR)$$

However, by taking advantage of the conditional independence assumptions (listed in Table 2.2), the calculation can be reduced to:

$$P(T,R,E,TM,G,V,ML,EO,RWR,TE) =$$
$$P(T) \otimes P(R) \otimes P(E|T) \otimes P(TM|T,R) \otimes P(G|T) \otimes P(V) \otimes P(ML|TM) \otimes$$
$$P(EO|TM,G) \otimes P(RWR|TM,G) \otimes P(TE|G,V)$$

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 in a Bayesian network involves "pointwise" multiplication of probability functions. Suppose P_s is a probability function for s and P_t is a probability function for t. Then $P_s \otimes P_t$ is a probability function 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 $K = \Sigma \{P_s(x^{\downarrow s})P_t(x^{\downarrow t}) \mid x \in \Omega_{s \cup t}\}$ is a normalization constant. The un-normalized combination will be denoted by $\otimes'$, i.e.,

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

Marginalization in a Bayesian network involves addition over the state space of the variables being eliminated. Suppose P_s is a probability function for s, and suppose $A \in s$. The marginal of P_s for $s \backslash \{A\}$, denoted by $P_s^{\downarrow s \backslash \{A\}}$, is a probability function for $s \backslash \{A\}$ defined as follows:

$$P_s^{\downarrow s \backslash \{A\}}(x) = \Sigma \{P_s(x, a) \mid a \in \Omega_A\} \text{ for all } x \in \Omega_{s \backslash \{A\}}. \qquad (2.3)$$

Here, the symbol '\' denotes set-theoretic subtraction, i.e., $s \backslash r$ denote the set of all elements of s that are not in r.

Inference. The conditional probability functions (or conditionals, in short) 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 variables are instantiated into the network, combination of probabilities proceeds as pointwise multiplication of likelihoods and conditionals. This combination results in an *un-normalized* posterior joint distribution. The process of computing posterior marginal probabilities given observations or likelihoods is referred to as *evidence propagation*.

In the example of Figure 2.1, given the number of states in the state space of the variables, the joint distribution will have $7^4 \cdot 2^4 \cdot 3^2 = 345{,}744$ states. Determining the marginal probability of each variable from the joint distribution is conceptually simple, but computationally expensive. Fortunately, methods for calculating the marginal probabilities of variables of interest using local computation—without explicitly computing the joint distribution—are available [Pearl 1986, Lauritzen and Spiegelhalter 1988, Jensen *et al.* 1990, Shenoy and Shafer 1990]. Software is readily available for automating the process of inference in Bayesian networks; the examples presented here use Netica [www.norsys.com] to calculate the prior and posterior marginal distributions.

The network from Figure 2.1 is shown again in Figure 2.2 with the prior marginal probabilities calculated based on the assigned conditional distributions.

Figure 2.2. Numerical Representation of the Threat ID Bayesian Network
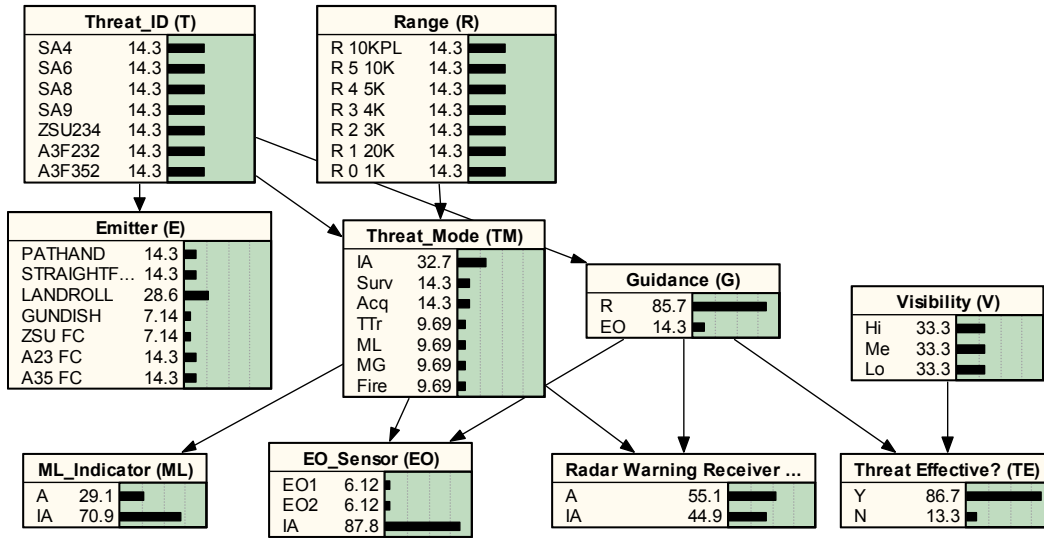
Suppose the following information becomes available:

(1) The true state of Threat Effectiveness (*TE*) = Y.

(2) Intelligence reports the presence of threats SA-4, SA-6, ZSU-23/4, A3F-23/2 and A3F-35/2 with low confidence (3:2 odds)[1]

| L(T)	SA-4	SA-6	SA-8	SA-9	ZSU-23/4	A3F-23/2	A3F-35/2
	3	3	2	2	3	3	3

---

[1] We have in mind an intelligence source with reliability 1/3 who tells us that the enemy only has threats SA-4, SA-6, ZSU-23/4, A3F-23/2 and A3F-35/2. Thus, using the language of belief function, we can model this as a basic probability assignment: m({SA-4, SA-6, ZSU-23/4, A3F-23/2 and A3F-35/2}) = 1/3, m(Ω_T) = 2/3. If we convert this bpa to a probability function using the plausibility transformation method [Cobb and Shenoy 2003, see also the discussion in Section 4 of this paper], we get the un-normalized function as shown, i.e., if e denotes the evidence, then, e.g., P(e|SA-4)/P(e|SA-8) = 3/2, etc. Notice that since Bayesian combination operation defined in (2.1) involves normalization, we have the flexibility of expressing a likelihood vector in relative magnitudes. Thus the likelihood vector (1, 1, 2/3. 2/3. 1, 1, 1) is equivalent to the likelihood vector (3, 3, 2, 2, 3, 3, 3) since the relative magnitudes are the same.

(3) Visibility is bright but hazy (1:2:0 for Hi:Me:Lo)

L(V)	Hi	Me	Lo
	1	2	0

Adding this information as likelihoods in the Bayesian network yields the marginal probabilities of the posterior distribution, as shown in Figure 2.3.

Figure 2.3. Bayesian Network after Calculation of the Marginals of the Posterior Distribution
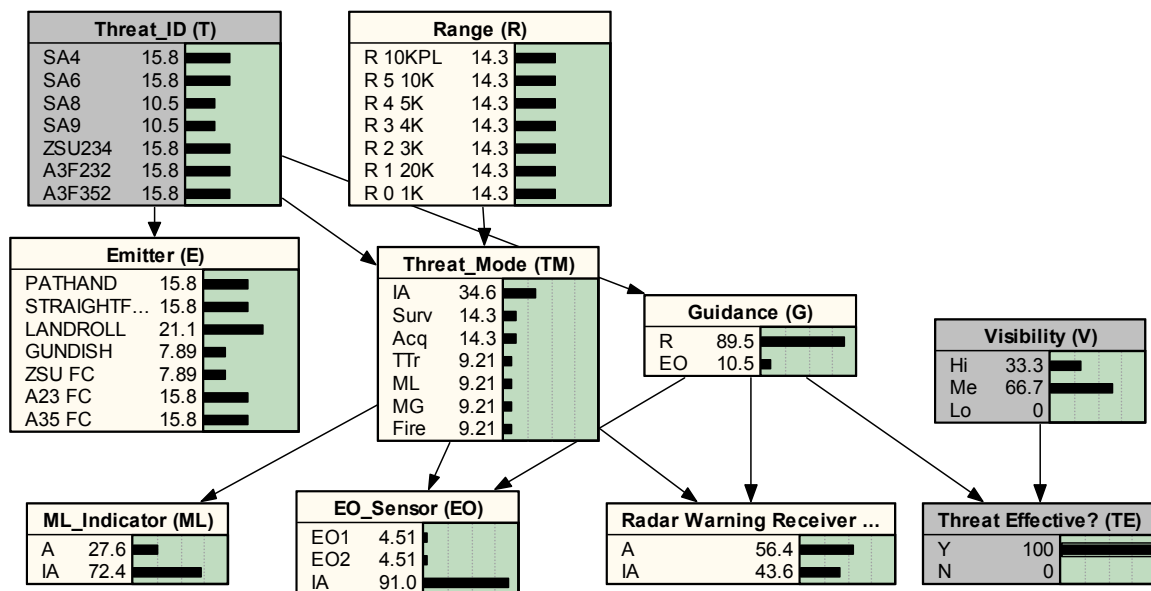
The posterior distribution provides updated information about the probabilities of each state of the Threat (*T*) variable. Prior to adding the new information to the network, each state of *T* had a probability of occurrence of 1/7. The posterior distribution reveals that the states {SA-8} and {SA-9} are slightly less likely than the other five states.

The type of inference where marginal probabilities for variables of interest are updated is referred to as *sum propagation*. Another type of inference—*max propagation*—finds the configuration of states of variables that has the maximum probability, i.e., a mode of the joint distribution.

Bayesian networks use conditional probability functions as their numerical inputs. Dempster-Shafer (D-S) belief networks are graphical representations that use D-S belief functions or basic probability assignments as their numerical inputs. These are described in the next section.

3. Dempster-Shafer Belief Networks

Representation. 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 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.

The valuation network (VN) graph defined by Shenoy [1992] can be used to construct a D-S belief network. This is done for the hypothetical anti-air threat identification problem in Figure 3.1. 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 Figure 3.1, evidence is available regarding variables R 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., $m^{\downarrow t}(\Omega_t) = 1$. In the D-S network shown in Figure 3.1, since this network models the same knowledge as described in the Bayes net of Figure 2.1, most of the valuations representing the knowledge of the domain are conditionals. Thus, e.g., the bpa $m_{T\text{-}E}$ is a conditional for E given T. In the valuation network

representation in Figure 3.1, we have chosen to omit variable *TE* and have represented the relationship between visibility and guidance directly as bpa $m_{V\text{-}G}$. See Section 4 for more details about the bpa $m_{V\text{-}G}$.

      Semantics. D-S belief networks are constructed with the fundamental assumption that combining all bpa's yields the correct joint bpa. Let m_i be a D-S bpa for valuation node i in a D-S belief network. Then $m = \oplus \{m_i \mid i = 1, \ldots, n\}$ is the joint bpa which is obtained by combining all bpa's in a D-S belief network. In the example of Figure 3.1:

$$m = m_{T-TM-R} \oplus m_{T-E} \oplus m_{T-G} \oplus m_{V-G} \oplus m_{TM-ML} \oplus m_{G-M-EO} \oplus m_{TM-G-RWR} \oplus m_T \oplus m_V$$

      Assuming the joint bpa can be determined in this way equates to assuming a set of conditional independence relations in the joint bpa [Shenoy 1994]. These conditional independence assumptions are encoded in the structure of the valuation network graph, as opposed to the numerical details of the bpa's. Variable X is a *neighbor* of variable Y if there exists a non-vacuous bpa whose domain includes $\{X, Y\}$. Suppose r, s, and t are three disjoint subsets of variables. Then r and s are *conditionally independent given t* if every path from a variable in r to a variable in s contains a variable in t. For example, consider three disjoint subsets of variables in the network of Figure 3.1: $r = \{T, R, E\}$, $s = \{V\}$, and $t = \{G\}$. All paths from a node in r to a node in s includes variable G. Therefore, r and s are conditionally independent given t. Notice that each variable, given its neighbors, is independent of all remaining variables. Table 3.1 all such conditional independence assumptions in the network of Figure 3.1.

12

Figure 3.1. A Dempster-Shafer Belief Network for the Anti-Air Threat Identification Problem
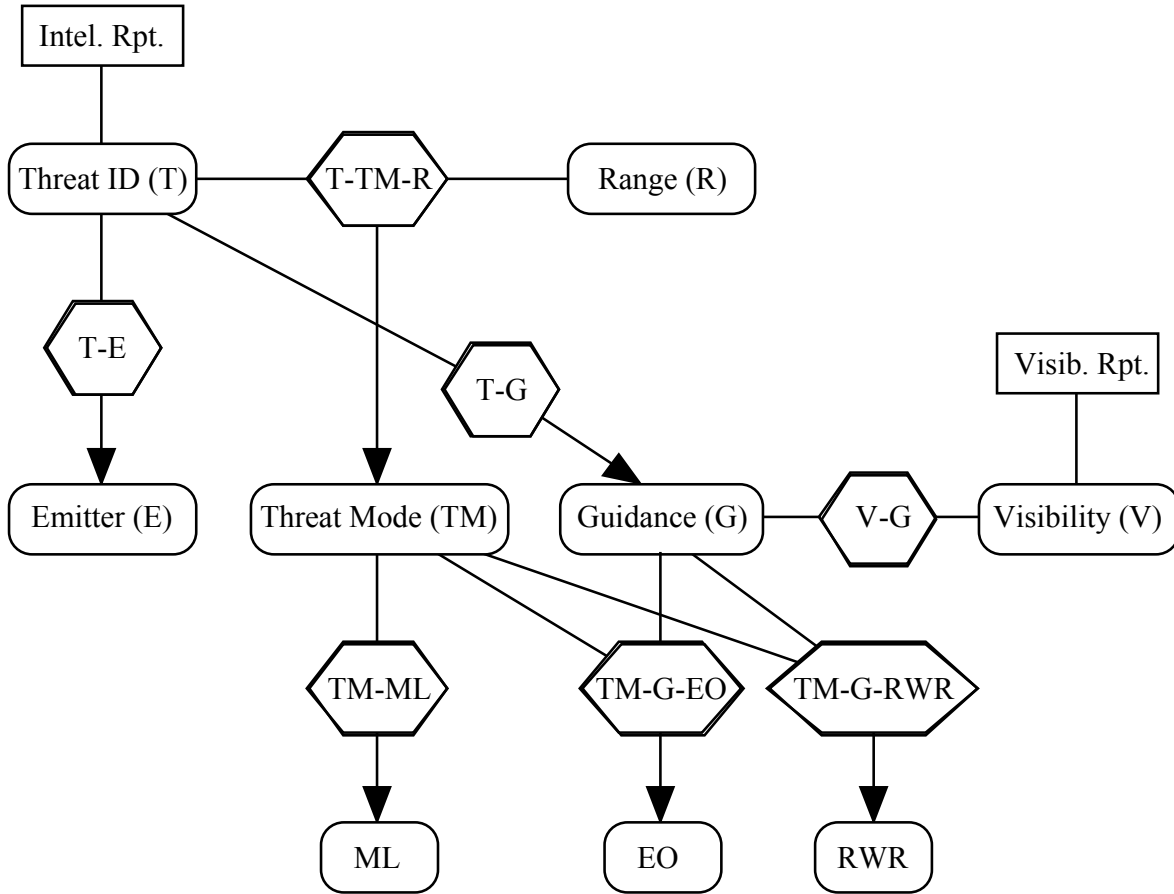
The conditional independence condition implicit in a D-S belief network ignores the presence of conditional bpa's. Shenoy [1994b] describes how the presence of conditionals can be used to detect additional independence conditions. If one takes into account these additional conditions then conditional independence assumptions in the D-S belief network of Figure 3.1 are exactly the same as those in the Bayesian network of Figure 2.1.

D-S belief network models are created using two-part construction. First, the structure of the belief network is formed using one of two methods. An expert's causal knowledge can be used to assess conditional independence relations by determining "direct causes"; this method is

similar to that used in constructing a Bayesian network. Alternatively, conditional independence

relations can be directly assessed using the no double-counting interpretation.

Table 3.1. Conditional Independence Assumptions for the Threat ID D-S Belief Network

Variable	Neighbors	Conditional Independence Assumption
T	R, E, TM, G	$T \perp \{V, ML, EO, RWR\} \mid \{R, E, TM, G\}$
TM	T, G, R, ML, EO, RWR	$TM \perp \{E, V\} \mid \{T, G, R, ML, EO, RWR\}$
G	T, TM, EO, RWR, V	$G \perp \{E, R, ML\} \mid \{T, TM, EO, RWR, V\}$
E	T	$E \perp \{R, TM, G, V, ML, EO, RWR\} \mid T$
R	T, TM	$R \perp \{E, G, V, ML, EO, RWR\} \mid \{T, TM\}$
V	G	$V \perp \{T, TM, G, R, ML, EO, RWR\} \mid G$
RWR	TM, G	$RWR \perp \{T, E, R, V, ML, EO\} \mid \{TM, G\}$
EO	TM, G	$EO \perp \{T, E, R, V, ML, RWR\} \mid \{TM, G\}$
ML	TM	$ML \perp \{T, R, G, V, ML, EO, RWR\} \mid TM$

To illustrate the no double-counting interpretation of conditional independence, consider

two variables Threat (T) and Emitter (E) in the D-S belief network described previously with the

joint distribution $P_{T, E}$. Probability theory defines the factorization of the joint distribution as:

$P_{T, E} = P_T \otimes P_{E|T}$ where $P_T = P_{T, E}^{\downarrow T}$ and $P_{E|T} = P_{T, E}/P_T$. Thus, it is okay to combine P_T and $P_{E|T}$

since it will always give the correct joint function $P_{T, E}$. In the illustration, P_T encodes

information about T, but $P_{E|T}$ encodes nothing about T because $P_{E|T}$ has the property that $P_{E|T}^{\downarrow T}$

is vacuous. Thus, there is no double counting of information about T in the combination of P_T

and $P_{E|T}$.

Suppose now that $P_{T, E}$ is constructed by combining P_T with P_E. Consider P_E that is

computed from $P_{T, E}$ as follows: $P_E = P_{T, E}^{\downarrow E}$. In general, $P_{T, E} \neq P_T \otimes P_E$. Thus P_T and P_E are not

independent with respect to $P_{T, E}$. Notice that $P_E = P_{T, E}^{\downarrow E} = [P_T \otimes P_{E|T}]^{\downarrow E}$. Since P_E contains

information about T, in combining P_T and P_E, information about P_T is double-counted and

therefore this product will clearly not yield the correct joint distribution $P_{T, E}$. There are two

cases where P_T and P_E are independent with respect to $P_{T, E}$. First, if $P_T = P_T \otimes P_T$, then $P_{T, E} =$

$P_T \otimes P_E$. In this case, although there is double counting of P_T, this is of no consequence because P_T is idempotent. Second, suppose that the values of $P_{E|T}$ don't depend on T, i.e., $P_{E|T}(e_i, t_j) = P_{E|T}(e_i, t_k)$ for all $e_i \in \Omega_E$, and all $t_j, t_k \in \Omega_T$. In this case, it is easy to show that $P_{E|T}(e_i, t_j) = P_E(e_i)$ for all $t_j \in \Omega_T$, and consequently, $P_E = P_{T, E}^{\downarrow E} = [P_T \otimes P_{E|T}]^{\downarrow E} = [P_T \otimes P_E]^{\downarrow E}$. In the process of marginalizing $P_T \otimes P_E$ to E, information in P_T is completely removed. Therefore $P_E \otimes P_T = P_{T, E}$ and there is no double counting.

The model developed in Figure 3.1 for the hypothetical anti-air threat identification problem will be correct if there is no double-counting of uncertain information in the joint bpa:

$$m = m_{T-TM-R} \oplus m_{T-E} \oplus m_{T-G} \oplus m_{V-G} \oplus m_{TM-ML} \oplus m_{G-M-EO} \oplus m_{TM-G-RWR} \oplus m_T \oplus m_V.$$

Projection and Extension of Subsets. Before we can define combination and marginalization for bpa potentials, 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 is a proper subset of s, and $\boldsymbol{a}$ is a nonempty subset of Ω_r, then the *extension of $\boldsymbol{a}$ to s* is $\boldsymbol{a} \times \Omega_{s\backslash r}$. 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 the extension of $\boldsymbol{a}$ to $\{W, X, Y, Z\}$ is $\boldsymbol{a} \times \Omega_{\{Y, Z\}}$.

Dempster's Rule of Combination. 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_X and m_Y for x and y, respectively. The combination of m_X and m_Y, denoted by $m_X \oplus m_Y$, is a bpa for $x \cup y$ given by

$$(m_X \oplus m_Y)(\boldsymbol{c}) = K^{-1} \Sigma \{m_X(\boldsymbol{a}) m_Y(\boldsymbol{b}) \mid (\boldsymbol{a}^{\uparrow(x \cup y)}) \cap (\boldsymbol{b}^{\uparrow(x \cup y)}) = \boldsymbol{c}\}$$

for all nonempty $\boldsymbol{c} \subseteq \Omega_{x \cup y}$, where K is a normalization constant given by

$$K = \Sigma \{m_X(\boldsymbol{a}) m_Y(\boldsymbol{b}) \mid (\boldsymbol{a}^{\uparrow(x \cup y)}) \cap (\boldsymbol{b}^{\uparrow(x \cup y)}) \neq \varnothing\}.$$

The un-normalized joint bpa for $x \cup y$ is given by

$$(m_X \oplus' m_Y)(c) = \Sigma\{m_X(a)m_Y(b) \mid (a^{\uparrow(x \cup y)}) \cap (b^{\uparrow(x \cup y)}) = c\}$$

for all nonempty $c \subseteq \Omega_{x \cup y}$.

Clearly, if the normalization constant is equal to zero, the combination is not defined, so the two bpa's are termed *not combinable*. If the bpa's m_X and m_Y are based on independent bodies of evidence, then $m_X \oplus m_Y$ 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 3.1, for instance, could be combined in any order to yield the joint bpa.

Marginalization. Like marginalization for probability functions, marginalization for bpa is obtained by addition. Suppose m is a bpa for s, and suppose $A \in s$. The *marginal* of m for $s \backslash \{A\}$, denoted by $m^{\downarrow(s \backslash \{A\})}$, is the bpa for $s \backslash \{A\}$ defined as follows:

$$m^{\downarrow(s \backslash \{A\})}(a) = \Sigma\{m(b) \mid b \subseteq \Omega_s \text{ such that } b^{\downarrow(s \backslash \{A\})} = a\}$$

for all nonempty subsets a of $\Omega_{s \backslash \{A\}}$.

To illustrate Dempster's rule, consider m_{T-E}, , one of the valuations in the D-S belief network of Figure 3.1. Define the state spaces for the variables Threat (T) and Emitter (E) as follows:

Ω_T	Description	Ω_E	Description
t_1	SA-4	e_1	Pathhand
t_2	SA-6	e_2	Straightflush
t_3	SA-8	e_3	Landroll
t_4	SA-9	e_4	Gundish
t_5	ZSU-23/4	e_5	ZSU-FC
t_6	A3F-23/2	e_6	A23-FC
t_7	A3F-35/2	e_7	A35-FC

Configurations of the two variables are assigned mass in the bpa m_{T-E} as follows:

$$m_{T-E}(\{(t_1,e_1), (t_2,e_2), (t_3,e_3), (t_4,e_3), (t_5,e_4), (t_5,e_5), (t_6,e_6), (t_7,e_7)\}) = 1.00$$

Suppose weak evidence exists that the Threat (T) is in the subset, $a = \{t_1, t_2, t_5, t_6, t_7\}$. This evidence about the Threat (T) is introduced into the network through another bpa:

$m_T(\{t_1, t_2, t_5, t_6, t_7\}) = 0.3$, $m_T(\Omega_T) = 0.7$

These two bpa's are combined using Dempster's Rule resulting in the joint bpa $m_{T\text{-}E} \oplus m_T$ described as follows:

$$(m_{T-E} \oplus m_T)(\{(t_1, e_1), (t_2, e_2), (t_5, e_4), (t_5, e_5), (t_6, e_6), (t_7, e_7)\}) = 0.3$$

$$(m_{T-E} \oplus m_T)(\{(t_1, e_1), (t_2, e_2), (t_3, e_3), (t_4, e_3), (t_5, e_4), (t_5, e_5), (t_6, e_6), (t_7, e_7)\}) = 0.7$$

The resulting combination can be marginalized to the Threat (T) variable as follows:

$$(m_{T-E} \oplus m_T)^{\downarrow T}(\{t_1, t_2, t_5, t_6, t_7\}) = 0.3 \text{, } (m_{T-E} \oplus m_T)^{\downarrow T}(\Omega_T) = 0.7 \text{.}$$

A useful way to summarize the information contained in the resulting bpa is to calculate the corresponding plausibility function. The plausibility function corresponding to a bpa m_T is defined as a function $Pl: 2^{\Omega_s} \rightarrow [0,1]$ such that: $Pl(\{a\}) = \Sigma\{m_T(b) \mid b \cap a \neq \varnothing\}$.

In this case, it may be useful to focus on the singleton elements of Threat (T) to determine which are now considered most likely. Plausibilities are calculated as follows, with the rightmost column representing the normalized plausibility calculated by dividing the plausibility of each element by the sum of all the plausibilities:

Ω_T	Description	$Pl(T)$	"Normalized" $Pl(T)$
t_1	SA-4	0.3+0.7 = 1.0	0.15625
t_2	SA-6	0.3+0.7 = 1.0	0.15625
t_3	SA-8	0.7	0.10938
t_4	SA-9	0.7	0.10938
t_5	ZSU-23/4	0.3+0.7 = 1.0	0.15625
t_6	A3F-23/2	0.3+0.7 = 1.0	0.15625
t_7	A3F-35/2	0.3+0.7 = 1.0	0.15625

Based on the combined evidence, the SA-8 and SA-9 states are considered slightly weaker than the other potential threats.

Bayesian networks and D-S belief networks have been proposed as alternative methods of modeling knowledge about propositions in uncertain domains. Both systems are composed of

both graphical and numerical representations. The next section compares Bayesian networks and D-S belief networks and examines the similarities and differences.

4.   A Comparison

The previous two sections have described the representation and semantics of Bayesian networks and Dempster-Shafer (D-S) belief networks. Differences between the two types of models exist in the graphical representations, numerical details, semantics and methods of performing inference. However, the two types of models are also similar in important aspects.

The graphical representations in each type of network use variables and assign state spaces to each variable. The relationships between these variables encode qualitative conditional independence assumptions of the uncertain domain. The differences in the conditional independence assumptions in Bayesian and D-S networks are superficial, as most of the conditional independence assumptions encoded in a Bayesian network are also represented in the corresponding D-S belief network. Those conditional independence assumptions present in the Bayesian network that are not encoded in the D-S belief network are a consequence of the valuations not being conditional probabilities. These assumptions are not used for propagation by general-purpose algorithms for computing marginals.

Bayesian network model construction involves assessing conditional independence relations by considering "direct causes" and "irrelevance" as criteria, whereas construction of a D-S belief network model involves assessing conditional independence relations using semantics of "no double-counting." At the numerical level, a Bayesian network is composed of a factorization of a joint probability distribution for the variables in the network that can be used to specify prior marginal probabilities for each variable. In a D-S belief network, bpa's are specified for each valuation and can be used to calculate a joint bpa for the variables in the network, provided there is no double-counting of evidence.

The differing numerical representations in Bayesian networks and D-S belief networks each have relative advantages and weaknesses. Conditional probabilities are easy to use for representing causal knowledge, but can be difficult and non-intuitive to use for associational, logical, and other types of non-causal knowledge. On the other hand, bpa's are not intuitive to use for capturing causal knowledge, but are much easier to use for capturing evidence. Updating of knowledge in Bayesian networks is accomplished by using likelihoods, whereas updating of knowledge in D-S belief networks is performed by specifying evidence as bpa's.

As an example of the differences in numerical representation of causal knowledge using Bayesian and D-S belief networks, consider a domain with two variables: Cancer (C) and Smoking (S). The state space of S is comprised of two propositions: s = a person is a smoker and $\sim s$ = a person is a non-smoker; C is defined similarly for cancer. The knowledge that smoking causes cancer can be encoded in conditional probabilities by assigning a higher probability to the presence of cancer given that a person smokes and a relatively lower probability to cancer given that a person does not smoke. For instance, based on expert knowledge or historical data, a conditional probability representation, P, is specified as:

$$P(c \mid s) = 0.4, P(\sim c \mid s) = 0.6, P(c \mid \sim s) = 0.1, P(\sim c \mid \sim s) = 0.9$$

The probability of having cancer given that a person does not smoke is 0.10, but this probability increases to 0.40 given that a person smokes.

Smets [1978] (see also [Shafer 1982, Smets 1993a]) defines an operation, called the ballooning extension, for creating a bpa from a conditional probability representation for models consisting of a finite number of conditions, each based on independent empirical data. Using the above conditional representation P, the following bpa assignments are created:

$$m_1(\{(c, s), (c, \sim s), (\sim c, \sim s)\}) = 0.4 \quad m_2(\{(c, \sim s), (c, s), (\sim c, s)\}) = 0.1$$

$$m_1(\{(\sim c, s), (c, \sim s), (\sim c, \sim s)\}) = 0.6 \quad m_2(\{(\sim c, \sim s), (c, s), (\sim c, s)\}) = 0.9$$

In this example, let $\Theta = \{s, \sim s\}$ denote the set of possible values for the parameter, θ, $X = \{c, \sim c\}$ represent the set of outcomes and P_θ: $\theta \in \Theta$ denote the probability model. The corresponding bpa, m, is constructed by combining m_1 and m_2 using Dempster's Rule as follows:

$$m(\{(c,s),(c,\sim s)\}) = (0.4) \cdot (0.1) = 0.04$$
$$m(\{(c,s),(\sim c,\sim s)\}) = (0.4) \cdot (0.9) = 0.36$$
$$m(\{(\sim c,s),(c,\sim s)\}) = (0.6) \cdot (0.1) = 0.06$$
$$m(\{(\sim c,s),(\sim c,\sim s)\}) = (0.6) \cdot (0.9) = 0.54$$

The bpa, m, has the following properties:

(1) $m^{\downarrow s}$ is vacuous.

(2) If $m_s(s) = 1$, then $(m \oplus m_s)^{\downarrow c}(c) = 0.4$, $((m \oplus m_s)^{\downarrow c}(\sim c) = 0.6$

(3) If $m_{\sim s}(\sim s) = 1$, then $(m \oplus m_{\sim s})^{\downarrow c}(c) = 0.1$, $((m \oplus m_{\sim s})^{\downarrow c}(\sim c) = 0.9$

The ballooning extension representation of conditional probability distribution described above is only one method of several others that also satisfy the three conditions stated above. Other methods for constructing belief functions from conditional probability distributions have been described by Black and Laskey [1990], Dubois and Prade [1986], and Srivastava [1997].

To illustrate the representation of non-causal knowledge using both Bayesian and D-S numerical representations, consider the example of visibility (V) and guidance (G) of anti-air threats from the anti-air threat identification problem. A conditional representation requires creating a dummy variable (TE) whose parents are V and G, then instantiating $TE = y$, and expressing the constraint between V and G in $P(TE = y \mid V, G)$. The conditional representation for each combination of states of V and G is defined in Table 4.1.

Table 4.1. Conditional Representation for each Combination of States of V and G

P(TE|V, G)	y	n
Hi, R	1	0
Hi, EO	1	0
Me, R	1	0
Me, EO	0.4	0.6
Lo, R	1	0
Lo, EO	0.1	0.9

A bpa for the valuation V-G in the D-S belief network is as follows:

$$m_{V-G}(\{(Hi, R),(Hi, EO),(Me, R),(Lo, R)\}) = 0.6$$
$$m_{V-G}(\{(Hi, R),(Hi, EO),(Me, R),(Me, EO),(Lo, R)\}) = 0.3$$
$$m_{V-G}(\{(Hi, R),(Hi, EO),(Me, R),(Me, EO),(Lo, R),(Lo, EO)\}) = 0.1$$

To understand this bpa, suppose $V = Hi$ is represented as $m_{V=Hi}(\{Hi\}) = 1$. Then $(m_{V-G} \oplus m_{V=Hi})^{\downarrow G}$ is as follows: $(m_{V-G} \oplus m_{V=Hi})^{\downarrow G}(\{R, EO\}) = 1$. Similarly, if $V = Me$, then $(m_{V-G} \oplus m_{V=Me})^{\downarrow G}(\{R\}) = 0.6$, $(m_{V-G} \oplus m_{V=Me})^{\downarrow G}(\{R, EO\}) = 0.4$. And if $V = Lo$, then $(m_{V-G} \oplus m_{V=Lo})^{\downarrow G}(\{R\}) = 0.9$, $(m_{V-G} \oplus m_{V=Lo})^{\downarrow G}(\{R, EO\}) = 0.1$. Thus, m_{V-G} can be thought of as a "conditional" for G yielding the above bpa's for different observed values of V.

Belief functions are easier to use for modeling evidence than likelihoods. For example, in the anti-air threat identification problem, if weak evidence (reliability = 1/3) is available that the enemy has only SA-4, SA-6, ZSU-23/4, A3F-23/2 and A3F-35/2 threats, this evidence can be easily modeled with the following bpa:

$$m(\{SA\text{–}4, SA\text{–}6, ZSU\text{–}23/4, A3F\text{–}23/2, A3F\text{–}35/2\}) = 1/3$$

$$m(\Omega_T) = 2/3$$

To model this evidence as a likelihood function, we assign a likelihood for each state that is proportional to the plausibility of that state:

$$L(SA-4) = L(SA-6) = L(ZSU-23/4) = L(A3F-23/2) = L(A3F-35/2) = 1$$
$$L(SA-8) = L(SA-9) = 2/3$$

These likelihoods can then be entered into the Bayesian network in the usual inference process. The likelihoods should be based on evidence alone so that there is no double counting of knowledge already represented in the Bayesian network.

We conclude this section by discussing an example used by Bogler [1987] to claim that belief functions are superior to probability theory. However, we show that if the belief function model is transformed to a probability model correctly (by using the so-called plausibility transformation [Cobb and Shenoy 2003], then the two calculi give identical conclusions.

Bogler's Example. Consider a Dempster-Shafer model of two independent pieces of evidence in the context of the anti-air threat identification problem:

1. Weak evidence, i.e., reliability is only 0.3, that the threat is a SA-4 missile modeled using bpa m_1:
$$m_1(\{SA-4\}) = 0.3$$
$$m_1(\{SA-4, SA-6, SA-8, SA-9, ZSU-23/4, A3F-23/2, A3F-35/2\}) = 0.7$$

2. Evidence that the missile is definitely a SAM modeled using bpa m_2:
$$m_2(\{SA-4, SA-6, SA-8, SA-9\}) = 1$$

Using Dempster's rule, the evidence is combined into bpa $m_1 \oplus m_2$:
$$(m_1 \oplus m_2)(\{SA-4\}) = 0.3$$
$$(m_1 \oplus m_2)(\{SA-4, SA-6, SA-8, SA-9\}) = 0.7$$

The plausibilities of each singleton subsets in the combined bpa are:

$$Pl_{m_1 \oplus m_2}(\{SA-4\}) = 1$$
$$Pl_{m_1 \oplus m_2}(\{SA-6\}) = 0.7$$
$$Pl_{m_1 \oplus m_2}(\{SA-8\}) = 0.7$$
$$Pl_{m_1 \oplus m_2}(\{SA-9\}) = 0.7$$

The ratio of the plausibilities of states {SA-4} and {SA-6} are:

$$\frac{Pl_{m_1 \oplus m_2}(\{SA-4\})}{Pl_{m_1 \oplus m_2}(\{SA-6\})} = \frac{1}{0.7} = 1.43$$

Suppose a Bayesian probability model has an equally likely prior probability distribution, *P*, for the seven states of the Threat ID (*T*) variable:

$$P(\{SA-4\}) = P(\{SA-6\}) = P(\{SA-8\}) = P(\{SA-9\}) = P(\{SZSU-23/4\}) = P(\{A3F-23/2\})$$
$$= P(\{A3F-35/2\}) = \frac{1}{7}$$

The weak evidence that the threat is an SA-4 missile is converted from the bpa representation to an un-normalized probability function using the plausibility transformation method [Cobb and Shenoy 2003], which yields the un-normalized probability function as follows:

$$L_1(\{SA-4\}) = 1$$
$$L_1(\{SA-6\}) = L_1(\{SA-8\}) = L_1(\{SA-9\}) = L_1(\{ZSU-23/4\}) = L_1(\{A3F-23/2\}) =$$
$$L_1(\{A3F-35/2\}) = 0.7$$

The evidence that the missile is definitely a SAM is modeled using a second likelihood function:

$$L_1(\{SA-4\}) = L_1(\{SA-6\}) = L_1(\{SA-8\}) = L_1(\{SA-9\}) = 1$$
$$L_1(\{ZSU-23/4\}) = L_1(\{A3F-23/2\}) = L_1(\{A3F-35/2\}) = 0$$

The prior probability distribution, *P*, and the two likelihood functions, L_1 and L_2, can then be combined and normalized to create a posterior probability distribution, *P'*:

$$P'(\{SA-4\}) = 0.323, P'(\{SA-6\}) = P'(\{SA-8\}) = P'(\{SA-9\}) = 0.226$$

The ratio of the probabilities of states {SA-4} and {SA-6} are:

$$\frac{P'(\{SA-4\})}{P'(\{SA-6\})} = \frac{0.323}{0.226} = 1.43,$$

which is roughly the same conclusion as in the belief function model.　■

Computationally, D-S belief networks are more expensive to evaluate than Bayesian networks. The worst-case complexity of a Bayesian network solution is O(*n*), where *n* is the size of the state space of the largest clique in the join tree, whereas the worst-case complexity of a D-S belief network is O(2^n), with *n* defined equivalently. The size of the state space of the largest clique depends on the sizes of the state spaces of variables, the sizes of state spaces of valuations, and the structure of the graph.

Differences exist in the graphical and numerical representations of Bayesian and D-S belief networks; however, as the above examples show, these two frameworks can both be

utilized to model knowledge and evidence of varying types. Bayesian networks and D-S belief networks both allow representation of knowledge in uncertain domains. Bayesian networks are easier to construct in domains where knowledge is causal, whereas D-S belief networks better facilitate representation of non-causal knowledge. When multiple models constructed on the same domain are available—a combination of Bayesian models and D-S models—a framework for combining the knowledge that does not require experts to reassess one model or another using a different numerical or graphical representation is needed. Additionally, when knowledge is non-causal, building a D-S belief network and solving the network by translating it to a corresponding Bayesian network may be a computationally less expensive than solving the D-S belief network directly.

Consider the Bayesian network of Figure 2.3. For the variables Range (R) and Visibility (V), each possible state is considered equally likely in the original problem formulation. Additional information about V or R might be more easily modeled as a bpa, as opposed to a likelihood function. A device may be designed to provide evidence about range or visibility using a bpa; however, since other valuations have been established as conditional probabilities and since Bayesian networks are typically more efficient to solve than D-S belief networks, the bpa could be transformed to a likelihood function prior to solution of the Bayesian network.

The previous example which utilized D-S plausibilities to create an equivalent treatment of evidence in D-S and Bayesian models points to a possible solution for combining evidence from the two types of models and translating D-S belief networks to Bayesian networks for solution during implementation.

Proponents of the Dempster-Shafer's theory of belief functions claim that the D-S theory is more expressive than probability theory since it can distinguish between vacuous knowledge (represented by a vacuous belief function) and knowledge that all states of a variable are equally likely (represented by a Bayesian belief function in which all focal elements are singleton subsets with the same probability mass). When such belief functions are converted to equivalent probability functions, they both reduce to an equally likely probability distribution. From a

descriptive point of view, Ellsberg [1961] argues that human decision makers react quite differently to the two different situations. From a normative point of view (e.g. Savage [1950]), the two situations have no significant differences for making decisions. We note, however, that there is no *coherent* decision theory for the Dempster-Shafer's theory that can take advantage of this expressiveness. The decision theory proposed by Jaffray [1989] and Strat [1990] requires the choice of an ad-hoc parameter that in essence reduces an expectation with respect to a belief function to an expectation with respect to a probability function. The decision theory proposed by Kennes and Smets [1994] reduces a belief function to a probability function prior to making decisions using Bayesian decision theory. Neither decision theory is able to exploit the so-called expressiveness of belief functions in representing ignorance.

5.   Conclusions and Summary

The main goal of this paper is to compare two seemingly disparate calculi for uncertain reasoning. While there are many differences, there are also many commonalities. Our main conclusion is that the two calculi have roughly the same expressive power. The reason we need these calculi is that they have different semantics and if our knowledge of the domain fits a particular semantic, then we should use the appropriate calculus to build a model in that domain. This does not mean that other calculi cannot represent the knowledge in the domain. An analogy is having a toolbox with many tools. If one has a nail, use a hammer. If one has a screw, use a screwdriver. We can drive a screw with a hammer, but the results may not be so elegant. We are skeptical of claims such as one tool is sufficient for all jobs or that one tool is superior to another. Thus, e.g., we should not be restricted to using just one calculus. We should be more concerned with the models we build using these calculi. The theory of belief functions provides us with some semantics for building models. If these semantics are appropriate for the domain we are trying to model, we should model the domain using belief functions.

The thesis that the theories of belief functions and probability have roughly the same expressive power suggests that one can translate a belief function model to an equivalent probability model. This topic is explored in detail in Cobb and Shenoy [2003].

Bibliography

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

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.

Cobb, B. R., P. P. Shenoy. 2003. On transforming belief function models to probability models. Working Paper No. 293, University of Kansas School of Business, Lawrence, KS.

Dawid, A. P. 1979. Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society Series B*, **41**(1) 1–31.

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

Dubois, D., H. Prade. 1986. The principle of minimum specificity as a basis for evidential reasoning. In: B. Bouchon, R. R. Yager, eds. *Uncertainty in Knowledge-Based Systems*, Lecture Notes in Computer Science No. 286, 75–84.

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

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

Jensen, F. V., S. L. Lauritzen, K. G. Olesen. 1990. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, **4** 269–282.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, H. G. Leimer. 1990. Independence properties of directed Markov fields. *Networks*, **20**(5) 491–505.

Lauritzen, S. L., D. J. Spiegelhalter. 1988. Local computation with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society Series B*, **20**(5) 157–224.

Pearl, J. 1986. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, **29** 241–288.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

Savage, L. J. 1950. *The Foundation of Statistics*. New York, NY, John Wiley & Sons.

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

Shafer, G. 1982. Belief functions and parametric models (with discussion). *Journal of the Royal Statistical Society Series B*, **44**(3) 322–352.

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

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. 1994. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, **10**(3) 203–234.

Shenoy, P. P. 1994b. Representing Conditional Independence Relations by Valuation Networks. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **2**(2), 143–165.

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. 1978. Un modéle mathématico-statistique stimulant le processus du diagnostic médical. Doctoral dissertation, Université Libre de Bruxelles.

Smets, P. 1993a. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, **9** 1–35.

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

Srivastava, R. P. 1997. Audit decisions using belief functions: A review. *Control and Cybernetics*, **26**(2) 135–160.

Strat, T. M. 1990. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, **4** 391–418.

Zarley, D.K., Y.T. Hsia, G. Shafer. 1988. Evidential reasoning using DELIEF. *Proceedings of the Seventh National Conference on Artificial Intelligence*, **1**, 205–209.