Characterizing Belief with Minimum Commitment*

Yen-Teh Hsia
IRIDIA, University Libre de Bruxelles
50 av. F. Roosevelt, CP 194/6
1050, Brussels, Belgium
r0 1509@ bbrbfu0 1 .bitnet

Abstract

We describe a new approach for reasoning with belief functions in this paper. This approach is fundamentally unrelated to probabilities and is consistent with Shafer and Tversky's canonical examples.

1 Introduction

Belief functions [Shafer, 1976; Smets, 1988] serve as a way to quantify human beliefs. It is a non-additive formalism. That is, $Bel(A) + Bel(A^c) \leq 1$, where A is a set formalizing some proposition and A^c is A's complement with respect to some underlying universe of discourse. This is different from probability theory where $P(A) + P(A^c) = 1$ is assumed. The reasoning mechanism of belief functions consists of two rules: Dempster's rule of conditioning allows us to update a given belief function *in* light of new information about the actual situation, whereas Dempster's rule of combination allows us to combine "distinct" or "independent" belief functions [Shafer, 1976; Smets, 1990].[1]

Researchers interested in belief functions often try to understand this formalism from the perspectives of probability theory (e.g., [Halpern and Fagin, 1990; Kyburg 1987; Nguyen, 1978; Pearl, 1988, Chapter 9]). This is understandable, as the origin of belief functions lies in the seminal paper of Dempster [1967], where he set out to study a particular subclass of upper and lower probabilities. However, Dempster's original view of belief functions is nowhere to be found in Shafer's 1976 monograph, where he offered a re-interpretation of Dempster's work and coined this formalism the term "belief functions."

Nevertheless, three questions about belief functions are of interests. One, why should human beliefs be quantified by belief functions? Two, why should Dempster's rule of conditioning be used for updating belief functions? Three, when is Dempster's rule of combination applicable in combining belief functions "representing" evidence?

Shafer and Tversky [1985] used canonical examples to answer the first and third questions, and their approach is not too far away from Dempster's original ideas.[2] In essence, Shafer and Tversky's approach is that we first compute the probability distribution on a space Θ_1, and then, by establishing a special relationship between Θ_1 and a second space Θ_2, we obtain our belief (a belief function) on Θ_2.

As we see it, there may be two drawbacks with Shafer and Tversky's canonical examples. The first drawback is that one may be tempted to question the legitimacy of using Dempster's rule of conditioning for updating the belief function on Θ_2. Because if we view the belief function on Θ_2 as the lower bound of a family of probabilities on Θ_2 (which is something we are not obliged to do, and both Shafer [19901 and Smets [1990] have explicitly rejected this idea), then we will have to use a different conditioning rule [Dempster, 1967; Fagin and Halpern, 1990; Jaffray, 1990]. The second drawback is more pragmatic in nature: as Shafer and Tversky only offered a pragmatic recommendation for reasoning with belief functions, it does not make belief functions adequately "protected" against potential misuse of this theory. In other words, one might still use belief functions (Dempster's rule of combination in particular) in a way that does not correspond to the canonical examples (see, for example, the analysis of [Pearl, 1990]).

In an attempt to remedy these drawbacks of the current belief-function framework (in the sense of Shafer and Tversky), we propose a radical "restructuring" of this framework as follows. First, we take away Dempster's rule of combination and make Dempster's rule of conditioning the one and only way for making inferences. Then, we suggest that we construct a belief function that is "minimum committed" in characterizing our intuitions.[3]

The benefits of our approach for reasoning with belief functions are four-fold. The first benefit is that we can now use belief functions directly. No references to probabilities are needed. The second benefit is that *as* belief functions arc no longer linked to probabilities, there is no more reason why we should reject the use of Dempster's rule of

* This work was supported in part by the DRUMS project funded by the Commission of the European Communities under the ESPRIT II-Program, Basic Research Project 3085.

formally, Dempster's rule of conditioning is a special case of Dempster's rule of combination. However, the underlying intuitions are completely different. One is concerned with belief updating, while the other is concerned with belief combination.

[2]But see [Halpern and Fagin, 1990; Ruspini, 1987; Smets, 1990] for some alternative answers.

[3]A belief function Bel_1 is *not as committed as* a belief function Bel2 if and only if for every proposition A, $Bel_1(A) \leq Bel_2(A)$.

conditioning for updating belief functions.[4] The third benefit is that we can now use Dempster's rule of conditioning to justify Dempster's rule of combination, thereby investigating the issue of "when is Dempster's rule of combination applicable?". The fourth and final benefit is that our restructured belief-function framework is more "robust" against potential misuse of belief functions. Because now whenever one wants to use Dempster's rule of combination to combine belief functions within our framework, he or she will have to explicitly justify the use of this rule. Without such a justification, the use of Dempster's rule of combination would only amount to what we call "inappropriate use of Dempster's rule of combination."

The reader might rightfully ask at this point: is this proposed reasoning approach consistent with Shafer and Tversky's canonical examples? The answer is "yes", and we will give more detail about this answer later in the paper.

The remainder of this paper is organized as follows. Sections Two and Three describe our restructured belief-function framework. Section Four shows that our approach is consistent with Shafer and Tvcrsky's canonical examples. Finally, Section Five concludes.

2 Basic concepts

The purpose of this section is to introduce the basic concepts of belief functions. We wish to emphasize the fact that only Dempsters rule of conditioning is introduced here, as Dempster's rule of combination is no longer considered an integrated part of our belief-function framework.

Let X = (X_1 X2, ..., XN) be a finite non-empty set of *variables* and let $\Theta_1, \Theta_2, \ldots, \Theta N$ be the respective *frames* of these variables (each Θi is a finite non-empty set of values Xi can take; these values arc mutually exclusive and exhaustive). Xi is boolean if Θ = [Yes, No). Let h be a non-empty subset of X.Θ_h is the Cartesian product of the frames of the elements of h. Θx, the set of *all possible situations,* is abbreviated as 0. By the "Xi-value" (1 < i < N) of an element <a_1, a_2, ..., aN> of Θ, we mean a_i.

We will need to work with subsets of Θ in specifying a belief function. However, it is often desirable that we only work with some of the variables in specifying a particular fragment of our belief. Therefore we allow the use of *logical formulas* in referring to subsets of 0, and we list in the appendix the formal correspondence between f, a formula, and [fj, f s corresponding subset of 0. This allows us to use a notation like [(Rain = Yes) ⊃ (Wet = Yes)] to refer to *Θ "minus" all those situations (elements of S) that have Rain-value Yes and Wet-value No.* It also allows us to use [-.(Temp = high)] in referring to *S "minus" all those situations (elements of Θ) that have Temp-value high.* This is a rather effective way to refer to subsets of 0. (In addition, we will use "Rain" as a shorthand for "Rain = Yes" in the case of boolean variables.) This way, we can unambiguously refer to subsets of 0 without committing ourselves to explicitly stating what variables are in x.

A *belief function* on 0 is a function Bel: 2Θ —» [0, 1] which is characterized by an *m-value function* m_{Bel} (written as "m" whenever confusions can be avoided; m is also called "the m-values of Bel"), where m: 2^Θ -* [0, 1] satisfies two conditions:

(1) $m(\varnothing) = 0$, and

(2) $\sum_{A:\ A \subseteq \Theta} m(A) = 1;$

and for every subset B of 0, Bel(B) is defined as $\sum_{A: A \subseteq B} m(A)$.[5] A subset A of 0 is called *a focal element* of Bel if m(A) > 0. When Bel is such that m(0) = 1, we call Bel *the vacuous belief function.*

Intuitively, $Bel(A) = c > 0$ means that "I believe that the actual situation is one of the situations in A, and c corresponds to how confident I am in entertaining this belief', Bel(A) = 0 means that "I do not entertain the belief that the actual situation is one of the situations in A",[6] and m(A) = d means that "in the course of establishing Bel, A is found to be the *most specific* subset of 0 that deserves this particular amount (d) of intuitive support." As such, Bel serves to characterize (part of) some distinguished state of mind, with m being the "internal structure" of Bel.

Once we accept this intuitive view of Bel and m, it is only natural that we extend this intuitive interpretation to Bel(. I B) and m(. I B), where, for example, Bel(A I B) = c> 0 means that "given that the actual situation is in B, I believe that the actual situation is in A, and c corresponds to how confident I am in entertaining this belief." This gives rise to the following conditioning rule known as Dempster's rule of conditioning [Shafer, 1976]. Let Bel be a belief function on 0 and m be its associated m-values. Let B be a non-empty subset of 0 such that $Bel(B^c)$ =/1.

$\forall C \subseteq \Theta$, if $C \subseteq B$

then $m(C \mid B)$ df= $\sum_{D:\ D \subseteq B^c} m(C \cup D) / K$

else $m(C \mid B)$ df= 0,

where $K = 1 - Bel(B^c)$ is the normalization constant.

(Note that for every subset S of Θ, $Bel(S \cap B \mid B) = Bel(S \mid B)$, but in general, $m(S \cap B \mid B) \neq m(S \mid B)$.) Intuitively, Dempster's rule of conditioning may be understood from two perspectives. First perspective ($C \subseteq B$): originally wc

[4]We acknowledge that this does not make Dempster's rule of conditioning any more feasible than any other updating rule one might think of. However, the point here is that Dempster's rule of conditioning now becomes *as competitive as* any other updating rule. What we need to do, then, is perhaps to find some axiomatic justification of this rule.

[5]This definition is consistent with [Shafer, 1976]. Smets [1988] has a slightly more general definition (called an "open world" definition) in which m(0) does not have to be 0 and Bcl(A) is defined as the sum of the m-values of those *non-empty* subsets of A.

[6]Consider the belief that the actual situation is in A. Here, according to our interpretation, an agent either *entertains* this belief or *does not entertain* this belief. And when the agent does entertain this belief, he/she/it is entitled to a degree of confidence (c) in doing so. In other words, we do *not* think of Bel(A) as *the extent to which an agent entertains this belief.*

committed $m(C \cup D) = s$ to $C \cup D$, as we thought $C \cup D$ as a whole deserves this much (s) intuitive support and we did not want to further "split" s among the elements of $C \cup D$; now we learn that the actual situation is in B; as a result, we decide that C should "inherit" s, as we *still* think C as a whole deserves this much intuitive support and we *still* do not want to further "split" s among the elements of C. Second perspective ($C \subsetneq B^c$): originally we considered C the most specific subset of B that deserves $m(C) = v$; now we learn that the actual situation is *not* in B^c; as our intuitions satisfy $Bel(B \mid B) = 1$ and $Bel(C \mid B) = 0$, rationality requires that we make $m(C \mid B)$ zero and redistribute v in some way; what we do then is that we redistribute v among the focal elements of $Bel(. \mid B)$ by proportions - a normalization process that is similar in spirit to what Bayes' rule of conditioning does.

3 A reasoning paradigm

Now we are in the position to present our approach for reasoning with belief functions. We start by posing the following question: *suppose* we are able to come up with fragmentary specifications of what our intuitions satisfy, where a fragmentary specification is either a *marginal* (e.g., Bel([WET]) = 0), a *conditional* (e.g., Bel([RAIN] I [WET]) = .3), or a mathematical relation among some of the marginals and conditionals (e.g., Bel([PARTY]) = Bel([PARTY] | [RAIN])), how should the system make inferences from these fragmentary specifications?

In answering this question, we simply regard all fragmentary specifications as constraints that a belief function must satisfy, and we ask the system to identify or construct a belief function that has the minimum commitment property (defined below) among all belief functions satisfying the specified constraints; if there is such a belief function, the system uses it to make inferences (by conditioning this belief function on *the current context* - information we currently have about the actual situation); if such a belief function does not exist, something else would have to be done, and we will briefly discuss about this problem in Section Five.

The principle of minimum commitment:[7] Let $c_1, \ldots, c_M$ be an enumeration of fragmentary specifications (i.e., constraints), and let $\mathbb{B}$ be the set of belief functions {Bel: Bel satisfies $c_1, \ldots,$ and c_M}. If there is an element σ of $\mathbb{B}$ such that $\forall \tau \in \mathbb{B}, \forall A \subseteq \Theta$, $\sigma(A) \leq \tau(A)$, then σ should be preferred to all other elements of $\mathbb{B}$ (i.e., σ should be used for making inferences), and we say σ is *a characterization* of (our intuitions) $c_1, \ldots,$ and c_M *with minimum commitment.*

Why this principle? Well, assuming that the user was serious in providing the fragmentary specifications (i.e., constraints that his or her intuition satisfies), we think that the principle of minimum commitment can serve as a useful "general agreement" between the user and the system. In essence, what this agreement says is that *"Bel(A) is, by default, as small as it is formally allowed."* In other words, the principle of minimum of commitment formalizes the following intuition: if I (the user) do not say whether I entertain a belief (or how confident I am in entertaining this belief), then it would be an error in inferring that the extent to which I *am* confident in entertaining this belief (if I entertain this belief at all) is larger than what it formally *has to be* the case. Thus, for example, if the user only specifies an empty set of constraints (and clearly any belief function satisfies these constraints), then according to the principle of minimum commitment, what the user actually means is that Bel(A) should be 0 for every subset A of Θ (and for obvious reasons, the user simply did not bother to specify it).

To recap, we summarize our reasoning approach as a two-step process below.

Step One - knowledge solicitation: the user specifies what his or her intuition satisfies. The resulting specifications are $c_1, \ldots, c_M$.

Step Two - reasoning: given the fragmentary specifications $c_1, \ldots, c_M$, the system tries to identify a characterization of the c_i's with minimum commitment; if there is such a belief function, the system uses it to make inferences (by conditioning the constructed belief function on the current context).

4 Relation with the canonical examples

One interesting aspect of our reasoning approach is that it is consistent with Shafer and Tversky's canonical examples. We now show that this is the case. We first introduce some terminology. Then, we give a general independence result. As it turns out, Shafer and Tversky's canonical examples are special applications of this general independence result.

4.1 Some definitions

We start by introducing Dempster's rule of combination ($\oplus_h$), which is regarded as a purely syntactical operation by itself in our framework.

Let h be a non-empty subset of χ. Let Bel_1 and Bel_2 be two belief functions on Θ_h (and let m_1 and m_2 be their respective m-values). $Bel_1 \oplus_h Bel_2$ is defined to be the following belief function Bel on Θ_h:

$\forall S \subseteq \Theta_h$, $m_{Bel}(S) = \sum_{A,B:\, A \cap B = S} m_1(A)m_2(B) / K$,

where $K = \sum_{A,B:\, A \cap B \neq \emptyset} m_1(A)m_2(B)$ is the renormalization constant.

(In what follows, $\oplus_\chi$ is abbreviated as $\oplus$.)

Next come the notions of belief projection and belief extension. Let h and g be two subsets of χ such that $h \subseteq g$ (g is χ by default).

The *projection* of an element x of Θ_g (e.g., g = {X_1, X_2, X_3, X_5, X_7} and x = <a1, a2, a3, a5, a7>) to Θ_h, denoted as $x^{\downarrow h}$, is simply this element with the extra coordinates dropped (e.g., h = {X_1, X_3, X_5} and $x^{\downarrow h}$ = <a1, a3, a5>).

The *projection* of a subset A of Θ_g to Θ_h, denoted as $A^{\downarrow h}$,

[7] Formally, the principle of minimum commitment is a variant of the principle of minimum specificity [Dubois and Prade, 1986a].

is $\{x^{\downarrow h} : x \in A\}$.

Let Bel be a belief function on Θ_g (and let m be its associated m-values), *the projection (or marginalization) of m to h*, denoted as $m^{\downarrow h}$, is defined as

$m^{\downarrow h}(A) = \sum_{B:\, B^{\downarrow h}=A} m(B)$.

$Bel^{\downarrow h}$ (*the projection of Bel to h*) is, as usual, characterized by $m^{\downarrow h}$.

The *extension* of S ($S \subseteq \Theta_h$) to Θ_g, denoted as $S^{\uparrow g}$, is the set $\{x: x \in \Theta_g$ and $x^{\downarrow h} \in S\}$.

Let Bel be a belief function on Θ_h (and let m be its associated m-values), *the extension of m to g*, denoted as $m^{\uparrow g}$, is defined as

$\forall A \subseteq \Theta_h$, $m^{\uparrow g}(A^{\uparrow g}) = m(A)$, and $m^{\uparrow g}(B) = 0$ for all other $B \subseteq \Theta_g$.

$Bel^{\uparrow g}$ (*the extension of Bel to g*) is, as usual, characterized by $m^{\uparrow g}$.

With all these notations in place, we now have the problem of specifying sequences of conditioning, projections, and extensions (i.e., which occurs before which). As a convention, we use $m^{\downarrow h}(A \mid B)$ to mean first conditioning and then projection (and then the value of A). The meaning of $m^{\uparrow h}(A \mid B)$ is similar. To signify a particular sequence, parentheses are also used. For example, $(m^{\downarrow h})(A \mid B)$ means projection first and conditioning second, whereas $((m^{\uparrow g})(A \mid B))^{\downarrow h}$ means extension first, conditioning second, and projection last. This clumsiness in the notation is very unfortunate, and we ask the reader to bear with us.

In what follows, we will be using a notation such as $\oplus_{XYZ}$ (or $Bel^{\downarrow XY}$) instead of its corresponding "legal" notation (e.g., $\oplus_{\{X,Y,Z\}}$ or $Bel^{\downarrow\{X,Y\}}$) whenever confusions can be avoided. It is also convenient to write $Bel^{\uparrow\Theta}$, $m^{\uparrow\Theta}$, or $A^{\uparrow\Theta}$ instead of $Bel^{\uparrow\chi}$, $m^{\uparrow\chi}$, or $A^{\uparrow\chi}$.

4.2 General independence

The notion of general independence is the belief-function counterpart of the notion of stochastic independence in probability theory. It formalizes what we mean by two "independent domains." In essence, the following definition of general independence says: whatever information we learn regarding the Y domain (i.e., Θ_Y), as long as we are not expecting the contrary with total confidence, this information will not have any effect on our (marginal) belief regarding the X domain (i.e., Θ_X); similarly, our belief regarding the Y domain is unaffected by information concerning the X domain.

Let X and Y be two variables (or, alternatively, two disjoint sets of variables g and h). Let Bel_X be a belief function on Θ_X and Bel_Y be a belief function on Θ_Y. A belief function Bel (on Θ) satisfies *general independence* with respect to X and Y, with Bel_X and Bel_Y being the two corresponding marginals, if and only if

(1) $Bel^{\downarrow X} = Bel_X$,

(2) $Bel^{\downarrow Y} = Bel_Y$,

(3) $\forall B \subseteq \Theta_Y$ such that $\exists B' \subseteq \Theta_Y$, $B \subseteq B'$ and $Bel_Y(B') \neq 0$, we have $Bel^{\downarrow X}(. \mid B^{\uparrow\Theta}) = Bel_X$, and

(4) $\forall A \subseteq \Theta_X$ such that $\exists A' \subseteq \Theta_X$, $A \subseteq A'$ and $Bel_X(A') \neq 0$, we have $Bel^{\downarrow Y}(. \mid A^{\uparrow\Theta}) = Bel_Y$.

What Theorem 1 below tells us, then, is that we can use Dempster's rule of combination to obtain *the* marginal belief on Θ_{XY} if our belief satisfies general independence with respect to X and Y, with Bel_X and Bel_Y being the two corresponding marginals.

Theorem 1 Let $\mathbb{B}$ be the set {Bel: Bel satisfies general independence with respect to X and Y, with Bel_X and Bel_Y being the two corresponding marginals}. The minimum committed belief function in $\mathbb{B}$ is $(Bel_X^{\uparrow XY} \oplus_{XY} Bel_Y^{\uparrow XY})^{\uparrow\Theta}$.

Proof of Theorem 1: given in the appendix.

4.3 Shafer and Tversky's canonical examples

We mentioned earlier that Shafer and Tversky's canonical examples may be viewed as special applications of our reasoning approach. In this section, we describe such an application. It is slightly more general than Shafer and Tversky's canonical examples, however - instead of working with probabilities P_X and P_Y (as Shafer and Tversky do), we work with belief functions Bel_X and Bel_Y.

The first part of the (slightly generalized) canonical examples is that we specify, through the use of fragmentary specifications, a marginal belief Bel_X on some space Θ_X. If this is all we specify, the system will construct $Bel_X^{\uparrow\Theta}$ and uses it to make inferences, and $(Bel_X^{\uparrow\Theta})^{\downarrow Z}$ ($= (Bel_X^{\uparrow XZ})^{\downarrow Z}$) is vacuous. Suppose we now receive some information regarding the actual situation and this information comes in the form of a compatibility relation $\mathfrak{C}$ (defined below) between Θ_X and Θ_Z, then the updated (and projected) belief on Θ_Z will be $((Bel_X^{\uparrow\Theta})(. \mid \mathfrak{C}^{\uparrow\Theta}))^{\downarrow Z}$ ($= ((Bel_X^{\uparrow XZ})(. \mid \mathfrak{C}))^{\downarrow Z}$). In general, this updated belief is not the same as the original $(Bel_X^{\uparrow XZ})^{\downarrow Z}$.

A *compatibility relation* $\mathfrak{C}$ between Θ_X and Θ_Z is a subset of $\Theta_X \times \Theta_Z$ such that $\forall a \in \Theta_X$, $\exists c \in \Theta_Z$, $(a, c) \in \mathfrak{C}$.[8]

The second part of the canonical examples involves a third frame Θ_Y. So let us change the story a little bit. We still specify, through the use of fragmentary specifications, a belief Bel_X on Θ_X. But now we also specify, again through the use of fragmentary specifications, another belief Bel_Y on Θ_Y. On top of that, we *tell*[9] the system that what we *really* want to specify is $Bel_X^{\uparrow XY} \oplus_{XY} Bel_Y^{\uparrow XY}$ on Θ_{XY}, as our intuitions satisfy general independence with respect to X and Y, with Bel_X and Bel_Y formalizing the two corresponding marginals. Upon receiving this instruction, the system replaces Bel_X and Bel_Y with $Bel_X^{\uparrow XY} \oplus_{XY} Bel_Y^{\uparrow XY}$.

Again, if this is all we specify, the system will construct

[8]Shafer [1987] also requires that $\forall c \in \Theta_Z$, $\exists a \in \Theta_X$, $(a, c) \in \mathfrak{C}$.

[9]That is, we assume that the system has some pre-processing ability and, as such, it can assist us in making fragmentary specifications

$(Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY})^{\uparrow\Theta}$ and uses it to make inferences, and $((Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY})^{\uparrow\Theta})^{\downarrow Z}$ $(= ((Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY})^{\uparrow XYZ})^{\downarrow Z} = (Bel_X{}^{\uparrow XYZ} \oplus_{XYZ} Bel_Y{}^{\uparrow XYZ})^{\downarrow Z})$ is vacuous. Suppose we now receive some information in the form of two compatibility relations $\mho_{XZ}$ and $\mho_{YZ}$, where $\mho_{XZ}$ is between Θ_X and Θ_Z and $\mho_{YZ}$ is between Θ_Y and Θ_Z. Then the updated (and projected) belief on Θ_Z is

$$(((Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY})^{\uparrow\Theta})(.\mid \mho_{XZ}{}^{\uparrow\Theta} \cap \mho_{YZ}{}^{\uparrow\Theta}))^{\downarrow Z}$$
$$= ((Bel_X{}^{\uparrow XYZ} \oplus_{XYZ} Bel_Y{}^{\uparrow XYZ})(.\mid \mho_{XZ}{}^{\uparrow XYZ} \cap \mho_{YZ}{}^{\uparrow XYZ}))^{\downarrow Z}$$
$$= ((Bel_X{}^{\uparrow XYZ})(.\mid \mho_{XZ}{}^{\uparrow XYZ}) \oplus_{XYZ} (Bel_Y{}^{\uparrow XYZ})(.\mid \mho_{YZ}{}^{\uparrow XYZ}))^{\downarrow Z}$$
$$= (((Bel_X{}^{\uparrow XZ})(.\mid \mho_{XZ}))^{\uparrow XYZ} \oplus_{XYZ} ((Bel_Y{}^{\uparrow YZ})(.\mid \mho_{YZ}))^{\uparrow XYZ})^{\downarrow Z}$$
$$= ((Bel_X{}^{\uparrow XZ})(.\mid \mho_{XZ}))^{\downarrow Z} \oplus_Z ((Bel_Y{}^{\uparrow YZ})(.\mid \mho_{YZ}))^{\downarrow Z}.$$

Of course, this updated (and projected) belief is, in general, not the same as the original $(Bel_X{}^{\uparrow XYZ} \oplus_{XYZ} Bel_Y{}^{\uparrow XYZ})^{\downarrow Z}$.

5 Conclusion

So what have we achieved? We have restructured the current belief-function framework in such a way that belief functions are no longer linked to probabilities. We have also provided some ingredients that we feel arc necessary in order to reason with belief functions. Our reasoning approach, as we have demonstrated in the last section, is in line (at least on the formal level) with Shafer and Tverskys recommendation for reasoning with belief functions. It is also not terribly limited, as formally both propositional logic and (Bayesian) probability may be viewed as special applications of this reasoning approach [Hsia, 1990].[10]

Nevertheless, we did not provide a specification methodology that, when followed, would allow the system to obtain a minimum committed belief function from the specified constraints. This, however, does not render our belief-function framework useless. In fact, what we have achieved is the *setting up* of a formal framework that would allow researchers to identify various specification methodologies (e.g., a generalized canonical example) that can be used under different circumstances.

As for the problem of finding a *genera!* specification methodology that we can use in tackling *any* problem, we do not consider it feasible in pursuing in this direction. The now famous Republican-Quaker-Pacifist problem is a good example. What should the system do if the user does not specify $Bel([P] \mid [R \wedge Q])$ and $Bel([\neg P] \mid [R \wedge Q])$? As it is entirely possible that the user himself/herself cannot make up his/her mind about these two values, there is no reason why the system should come up with an "answer". How, then, should we go about it in performing reasoning in such cases? One possible alternative may be to ask the system to infer properties that are shared by *all* minimally committed belief functions satisfying the same set of constraints. This appears to be an issue that is worthy of further explorations.

essence, propositional logic corresponds to the case in which we only specify constraints of the form $Bel(A) = 1$, whereas Bayesian probability corresponds to a complete specification of prior probabilities and conditional probabilities.

Acknowledgement

The author thanks Philippe Smets for being the mentor of this work, Robert Kennes for explaining Yager's ordering in terms of belief refinement, Alessandro Saffiotti for demanding a better presentation, and two anonymous referees for very helpful comments.

Appendix A: Logical formulas and subsets

Let $x \in \Theta$ and $a \in \Theta_i$ $(1 \le i \le N)$, we recursively define what a formula f is and whether x *satisfies* the formula f.

Case 1. f is "$X_i = a$": x *satisfies* f if and only if the X_i-value of x is a.

Case 2. f is "$\neg g$", where g is a formula: x *satisfies* f if and only if x does not satisfy g.

Case 3. f is "$g \vee h$", where g and h are formulas: x *satisfies* f if and only if x satisfies at least one of g and h.

Case 4. f is "$g \wedge h$", where g and h are formulas: x *satisfies* f if and only if x satisfies the formula "$\neg(\neg g \vee \neg h)$".

Case 5. f is "$g \supset h$", where g and h are formulas: x *satisfies* f if and only if x satisfies the formula "$\neg g \vee h$".

Let f be a formula. By *the subset (of Θ) the formula f refers to* (or, alternatively, *the subset (of Θ) the formula f corresponds to*), we mean the set $\{x: x \in \Theta$ and x satisfies f$\}$, which we denote as $[f]$.

Appendix B: Proof of Theorem 1

Basically what we need to show is that for every element Bel of $\mathbb{B}$, $Bel^{\downarrow XY} = Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY}$. Once this is proved, it follows that $(Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY})^{\uparrow\Theta}$ is a minimum element in $\mathbb{B}$ with respect to a partial ordering concept originally proposed by Yager [1985] (see also [Moral, 1985]) and later discussed in detail in [Dubois and Prade, 1986b]. Moreover, Dubois and Prade [1986b] showed that whenever $\mathbb{B}$ has a minimum element, that element (a belief function) must be the minimum committed belief function in $\mathbb{B}$. The interested reader is referred to [Hsia, 1990] for the details of Yager's ordering and Dubois and Prade's lemma.

Our task now is reduced to showing that for every element Bel of $\mathbb{B}$, $Bel^{\downarrow XY} = Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY}$.

Let Bel be an element of $\mathbb{B}$, m be the m-values associated with Bel, and $m^{\downarrow XY}$ be the m-values associated with $Bel^{\downarrow XY}$. Let $CORE_X$ be the union of all $A \subseteq \Theta_X$ such that $m_X(A) \neq 0$, and let $CORE_Y$ be the union of all $B \subseteq \Theta_Y$ such that $m_Y(B) \neq 0$. We know the following are satisfied.

$Bel^{\downarrow X} = Bel_X$, $Bel^{\downarrow Y} = Bel_Y$,

$\forall b \in CORE_Y$, $Bel^{\downarrow X}(.\mid \{b\}^{\uparrow\Theta}) = ((Bel^{\downarrow XY})(.\mid \{b\}^{\uparrow XY}))^{\downarrow X}$
$= Bel_X$, and

$\forall a \in CORE_X$, $Bel^{\downarrow Y}(.\mid \{a\}^{\uparrow\Theta}) = ((Bel^{\downarrow XY})(.\mid \{a\}^{\uparrow XY}))^{\downarrow Y}$
$= Bel_Y$.

Therefore $Bel^{\downarrow XY}$ must satisfy the following property:

$\forall S \subseteq \Theta_{XY}$, $m^{\downarrow XY}(S) \neq 0$ if and only if $\exists S_X \subseteq \Theta_X$, $\exists S_Y \subseteq \Theta_Y$, $m_X(S_X) \neq 0$, $m_Y(S_Y) \neq 0$, and $S = S_X \times S_Y$ (m_X and m_Y are the m-values associated with Bel_X and Bel_Y, respectively).

Suppose $A \subseteq \Theta_X$ and $B \subseteq \Theta_Y$ are such that $m_X(A) \neq 0$, $m_Y(B) \neq 0$, and either $A \neq CORE_X$ or $B \neq CORE_Y$. Without loss of generality, assume it is the case that $B \neq CORE_Y$. We have (due to the stated independence conditions) $Bel(A^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta} \mid CORE_Y{}^{\uparrow\Theta} \backslash B^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta} \mid \Theta \backslash B^{\uparrow\Theta}) = (Bel(A^{\uparrow\Theta} \cup B^{\uparrow\Theta}) - Bel(B^{\uparrow\Theta}))/(1 - Bel(B^{\uparrow\Theta}))$ (see [Shafer, 1976, p. 67] for this equivalent formulation of Dempster's rule of conditioning).

We expand $Bel(A^{\uparrow\Theta} \cup B^{\uparrow\Theta})$ by observing the following:

$\forall S_1 \subseteq A^{\uparrow\Theta} \cap B^{\uparrow\Theta}$ (S_1 may or may not be $\varnothing$), $\forall S_2 \subseteq A^{\uparrow\Theta} \backslash (A^{\uparrow\Theta} \cap B^{\uparrow\Theta})$ ($S_2 \neq \varnothing$), $\forall S_3 \subseteq B^{\uparrow\Theta} \backslash (A^{\uparrow\Theta} \cap B^{\uparrow\Theta})$ ($S_3 \neq \varnothing$), it is not the case that $\exists C \subseteq \Theta_X$, $\exists D \subseteq \Theta_Y$, such that $C \times D = S_1 \cup S_2 \cup S_3$.

In other words, $\forall S \subseteq A^{\uparrow\Theta} \cup B^{\uparrow\Theta}$ ($S \neq \varnothing$), if neither $A^{\uparrow\Theta}$ nor $B^{\uparrow\Theta}$ is a superset of S, then $S \neq C \times D$, $\forall C \subseteq \Theta_X$, $\forall D \subseteq \Theta_Y$ (and therefore S must have 0 as its m-value (i.e., $m(S) = 0$) - due to the above described property of $Bel^{\downarrow XY}$).

Thus, $Bel(A^{\uparrow\Theta} \cup B^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta}) + Bel(B^{\uparrow\Theta}) - Bel(A^{\uparrow\Theta} \cap B^{\uparrow\Theta})$.

Replacing this into the previous conditioning formula, we get $Bel(A^{\uparrow\Theta} \cap B^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta})Bel(B^{\uparrow\Theta})$, and this is true for any $A \subseteq \Theta_X$ and $B \subseteq \Theta_Y$ where $m_X(A) \neq 0$ and $m_Y(B) \neq 0$, and either $A \neq CORE_X$ or $B \neq CORE_Y$.

We observe further that $Bel(CORE_X{}^{\uparrow\Theta} \cap CORE_Y{}^{\uparrow\Theta}) = 1 = Bel(CORE_X{}^{\uparrow\Theta})Bel(CORE_Y{}^{\uparrow\Theta})$.

Thus, for every $A \subseteq \Theta_X$ and every $B \subseteq \Theta_Y$ where $m_X(A) \neq 0$ and $m_Y(B) \neq 0$, $Bel(A^{\uparrow\Theta} \cap B^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta})Bel(B^{\uparrow\Theta})$.

Since the only subsets of Θ_{XY} that are "eligible" (in $Bel^{\downarrow XY}$) for non-negative m-values are those that are the Cartesian product of some A in Θ_X and some B in Θ_Y where $m_X(A) \neq 0$ and $m_Y(B) \neq 0$ *and*,

for any such subset $A \times B$, $Bel^{\downarrow XY}(A \times B) = Bel((A \times B)^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta} \cap B^{\uparrow\Theta}) = Bel(A^{\uparrow\Theta})Bel(B^{\uparrow\Theta}) = Bel^{\downarrow X}(A)Bel^{\downarrow Y}(B) = Bel_X(A)Bel_Y(B)$,

it follows that $m^{\downarrow XY}(S)$ (where $S \subseteq \Theta_{XY}$) must be such that

if $\exists S_X \subseteq \Theta_X$, $\exists S_Y \subseteq \Theta_Y$ such that $S = S_X \times S_Y$, **then** $m^{\downarrow XY}(S) = m_X(S_X)m_Y(S_Y)$ **else** $m^{\downarrow XY}(S) = 0$.

This means $Bel^{\downarrow XY} = Bel_X{}^{\uparrow XY} \oplus_{XY} Bel_Y{}^{\uparrow XY}$. **Q.E.D.**

References

Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics,* 38, 325-339.

Dubois, D. and Prade, H. (1986a). The principle of minimum specificity as a basis for evidential reasoning. In *Uncertainty in Knowledge-Based Systems* (Bouchon and Yager eds.), Springer-Verlag, Berlin, 75-84.

Dubois, D. and Prade, H. (1986b). A set-theoretic view of belief functions - Logical operations and approximations by fuzzy sets. *International Journal of General Systems,* 12, 193-226.

Fagin, R. and Halpern, J.Y. (1990). A new approach to updating beliefs. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence,* Cambridge, Massachusetts, 317-325.

Halpern, J.Y. and Fagin, R. (1990). Two views of belief: Belief as generalized probability and belief as evidence. Research Report RJ 7221, IBM. (A shortened version appeared in *Proceedings of the Eighth National Conference on Artificial Intelligence,* American Association for Artificial Intelligence, Boston, Massachusetts, 112-119.)

Hsia, Y.-T. (1990). Characterizing belief with minimum commitment. Technical Report TR/IRIDIA/90-19, IRIDIA, University Libre de Bruxelles.

Jaffray, J.-Y. (1990). Bayesian updating and belief functions. In *Proceedings of the Third International Conference on Information Processing and Management of Uncertainty,* 449-451.

Kyburg, Jr., H.E. (1987). Bayesian and non-Bayesian evidential updating. *Artificial Intelligence,* 31, 271-293.

Moral, S. (1985). Informacion difusa. Relaciones entre probabilidad y posibilidad. Tesis Doctoral, Universidad de Granada

Nguyen, H.T. (1978). On random sets and belief functions. *Journal of Mathematical Analysis and Applications,* 65, 531-542.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Morgan Kaufmann Publishers, Inc., San Mateo, California.

Pearl, J. (1990). Reasoning with belief functions: an analysis of compatibility. *International Journal of Approximate Reasoning,* 4, 363-389.

Ruspini, E. H. (1987). Epistemic logics, probability, and the calculus of evidence. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence.*

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press.

Shafer, G. (1987). Belief functions and possibility measures. In *The Analysis of Fuzzy Information,* James C. Bezdek, ed., CRC Press.

Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning,* 4, 323-362.

Shafer, G. and Tversky, A. (1985). Languages and designs for probability judgment. *Cognitive Science,* 9, 309-339.

Smets, P. (1988). Belief functions. In *Non-Standard Logics for Automated Reasoning* (P. Smets, E. H. Mamdani, D. Dubois and H. Prade eds.). Academic Press, London.

Smets, P. (1990). The transferable belief model and other interpretations of Dempster-Shafer's model. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence,* Cambridge, Massachusetts, 326-333.

Yager, R. (1985). The entailment principle for Dempster-Shafer granules. Technical Report Mil 512, Iona College, New Rochelle, New York (also appeared in *International Journal of Intelligent Systems, 1,* 247-262).