THE COMBINATION OF EVIDENCE IN THE TRANSFERABLE BELIEF MODEL.

Philippe SMETS.
I.R.I.D.I.A.
Université Libre de Bruxelles
50 Av. F. Roosevelt, 1050 Brussels, Belgium.

SUMMARY:

Description of the transferable belief model to quantify degrees of belief, based on belief functions. The impact of open- and closed-world assumption on conditioning. Presentation of a set of axioms justifying Dempster's rule for the combination of belief functions induced by 2 distinct evidences.

KEYWORDS:
Belief function, plausibility function, degree of belief, combination of evidences, Dempster's rule of combination, open-world assumption, transferable belief model.

1. INTRODUCTION.

Ignorance can take 2 forms: imprecision and uncertainty. A statement is imprecise if it contains predicates that do not specify uniquely the value of a variable within its referential (e.g. 'John is less than 30 years old', 'John is young'). A statement is uncertain if one cannot evaluate its truth or its falsity given the available data (Dubois et al. [7])

Imprecision is represented by intervals, fuzzy sets (Dubois and Prade [8], Mamdani and Gaines [20]) or possibility functions (Zadeh [46], Dubois and Prade [9], Magrez [19]).

The most classical models to quantify **uncertainty** are based on probability functions and the Bayesian approach is often considered as the most appropriate (Fine [10]).

Nevertheless it is unrealistic to claim that all forms of uncertainties can and should be fitted by probability functions as argue in Lindley [17].

In this presentation, we consider that form of uncertainty that leads to a subjective, personalistic degree of belief. This is indeed the topic studied by the Bayesian probabilists. But another model based on **belief functions** has been proposed by Shafer [25, 27, 28] that seems more appropriate to represent someone's degree of belief.

The need for a mathematical model to simulate the belief process can be justified by cognitive, normative or pragmatic arguments:

1) *cognitive:* it helps in obtaining a better understanding of the underlying psychological process.
2) *normative:* it establishes rules of behavior that should be applied by everyone.
3) *pragmatic:* it provides a mathematical construct that can be implemented on computers, as in Expert Systems, in order to simulate a cognitive process where the concept of belief is relevant.

Whereas the probability model is surely the most popular, it is not exempt of real limitations. It hardly fits the human behavior it sets out to represent (Kahneman [15]). Normative Bayesians claim that this lack of fit is irrelevant. It only proves that the human being is a poor estimator of probabilities. The fact that humans cannot correctly guess the square root of 2 is not a criticism of arithmetic, it merely implies that one should use arithmetic in order to compute it. This normative justification would be definite if the bayesian model were absolutely convincing. Unfortunately all axiomatic models developed to justify the use of probability functions include some conditions open to criticism (Fine [10], Rivett [21]) and furthermore they suffer from real weaknesses when it comes to represent total ignorance.

From a pragmatical point of view, the developers of Expert Systems had to choose a mathematical model to quantify uncertainties. Some choose the probability or fuzzy sets theory. MYCIN (Shortliffe [32]) uses a model based on measures of belief and of disbelief and a certainty factor that does not obey probability axioms [3, 31]. These authors convincingly criticize the probability model but only provide an ad hoc model that also presents some weaknesses. More recently the Artificial Intelligence community has started to show interest for what they call the Dempster-Shafer's model, i.e. the use of belief functions [1, 11, 13, 14, 16, 18, 43]. Unfortunately confusion emerged between a model based on upper and lower probabilities and one based on transferable beliefs, which we are here presenting in detail. Initial work by Shafer [25] looked closer to the transferable belief model, but his more recent work is essentially based on the upper and lower probability model. Most work in Artificial Intelligence is based on the upper and lower probability interpretation.

The **upper and lower probability model** is based on works by Good [12], Smith [42] and Dempster [5, 6]. This model is based on the idea that there is a well-defined probability assigned to each proposition, but that these numerical values are only known by us to be within intervals whose boundaries are the so called upper and lower probabilities. This model generalizes the probability model.

Using this mathematical tool, Shafer [25] suggested that degrees of belief are quantified by **belief functions** and that belief functions induced by distinct pieces of evidence are combined through Dempster's rule of combination. The fact that this model is completely unrelated to the probability model was not clear, and this distinction was hardly enhanced by recent works. So we feel the necessity to present explicitly the **transferable belief model**, a normative model we hope might have some cognitive and pragmatic interest.

We insist on the fact that:
1) the transferable belief model is built without ever introducing explicitly or implicitly any concept of probability.
2) Dempster's rule of conditioning is one of the natural ingredients of the transferable belief model. It is not ad hoc [28]. It is at the centre of the model.
3) Dempster's rule of combination will be derived from the transferable belief model, the strongest postulate being the autofunctionality axiom A6, and not the conditioning axiom A4, as A4 is just a formalization of one of the constituents of the transferable belief model.

For the clarity of this paper, it is capital that the points above be borne in mind. The reader should also temporarily forget his/her previous indoctrination in probability theory and the fact that Dempster's rule of conditioning is often presented as a special case of Dempster's rule of combination (which it fortunately is, but which may not be accepted a priori, otherwise the whole derivation of Dempster's rule of combination would be a cyclical reasoning).

Parallelism will often be established with the probability model as it is of course the best known and major contender of the transferable belief model.

Part 2 discusses the nature of the frame of discernment on which a degree of belief will be established, and presents the distinction between the open and closed-world assumptions that are too often neglected. This distinction is essential in order to understand the normalization problem and to avoid Zadeh's paradox.

Part 3 presents the transferable belief model. Shafer [25] introduces his model by postulating that degrees of belief are quantified by Choquet's capacities [4], but the numerous inequalities that underlie these capacities are hardly satisfactory to convince people of the appropriateness of the model. Hence, the need to redevelop the model.

Part 4 presents Dempster's rule of conditioning.

Part 5 presents an axiomatic justification of Dempster's rule of combination. This rule - known in the 18th century [26] - may seem natural. Nevertheless an axiomatic derivation is useful in order to show its meaning and its relevance.

Part 6 discusses the impact of the closed-world assumption and hopefully resolves a criticism put forward by Zadeh [47] about the appropriateness of the renormalization encountered in Dempster's rules as defended by Shafer [29].

Part 7 concludes and summarizes the results.

Computational feasibility of belief functions is an open question. It is usually claimed that belief functions are computationaly intractable as works is done on power sets. Though theoretically true, the claim is not correct in practice. We even feel that belief functions might require less computational efforts than the use of probability functions. This paper being a theoretical presentation of belief functions, the problem is not investigate further.

2. THE FRAME OF DISCERNMENT.

Models on reasoning postulate a finite Boolean algebra of propositions Ω on which beliefs will be constructed, some propositions of Ω being believed more than other. (Ω is also called the frame of discernment, the universe of discourse, the domain of reference.) Usually only those propositions are considered, any proposition not included in Ω is claimed as impossible.

In reality, the cognitive process is hardly as simple. One first constructs a set **KP** of those propositions **Known as Possible**. But one must also consider 1) the set **UP** of **Unknown Propositions** that are not considered and 2) the set **KI** of those propositions **Known as Impossible**. In the classical approach like the bayesian, UP is postulated empty, one accepts a highly idealized closed-world assumption, i.e. that the truth is necessarily in KP.

The content of the three sets depends not only on the problem under scrutiny, but also on the available pieces of evidence. As evidence becomes available, propositions are redistributed among the three sets:

1) a proposition A is transferred from KP to KI if the evidence is sufficient to claim that A is impossible. It corresponds to conditioning.
2) a proposition A is transferred from UP to KP if the evidence induces us to consider as possible some forgotten propositions.
3) a proposition A is transferred from UP to KI if the evidence induces us to consider that some forgotten propositions are in fact impossible. This has no practical effect as the degrees of belief are constructed only on KP.
4) transfer from KI to KP or UP and from KP to UP would be inconsistent with the definition of the three sets, if one accepts, as here, that the allocation of any proposition to one of the three sets is always correct. A true proposition may be correctly allocated to KP and UP, and a false proposition may be correctly allocated to KP, KI or UP.

A true proposition may not be allocated to KI, and any proposition allocated to KI will stay in KI, inducing monotony for the impossible (false) propositions.

The **closed-world assumption** corresponds to an a priori empty UP set. The **open-world assumption** admits the existence of a non-empty UP set, and the fact that the truth might be in UP.

As mentioned before, it is admitted here that the truth may never be in KI. Generalization could be considered by accepting that a true proposition might be in KI and constructing some meta-belief function on the set of all propositions, that expresses the degree of belief that each proposition can belong to any of the three sets. It is not studied in this paper but can be resolved by methods developed in [33] in the framework of the generalized bayesian theorem.

3. QUANTIFICATION OF THE DEGREE OF BELIEF.

Let KP be a finite Boolean algebra of propositions Ω and let Δ be the set of elementary propositions of Ω. Let $\mathbf{1}_\Omega$ be the tautology relative to Ω i.e. $\mathbf{1}_\Omega$ is the disjunction of all elementary propositions of Δ . Let $\mathbf{0}_\Omega$ be the contradiction relative to Ω i.e. none of the propositions of Δ implies $\mathbf{0}_\Omega$. The conjunction of any two distinct propositions of Δ is

$\mathbf{0}_\Omega$. Any proposition of Ω different from $\mathbf{0}_\Omega$ is called a non vacuous proposition.

3.1. Notations.

One writes $\neg$, $\vee$,& for the negation, the disjunction and the conjunction connectives. Negation of any proposition A in Ω is taken relative to Δ. So $\neg$A is the disjunction of the elementary propositions of Δ not implying A.
The set UP is denoted Θ. No details about its structure and about KI are needed.

The symbols $\in$, $\rightarrow$, $\not\rightarrow$ are used with the following meanings:
$A\in\Delta$: A is an element of Δ
$A\in\Omega$: A is a proposition of Ω
$A\rightarrow B$: "it is true that A implies B" where $A, B\in\Omega$
$A\not\rightarrow B$: "it is true that A does not imply B" where $A, B\in\Omega$

We write $\sum_{A\rightarrow B} m(A)$ and $\sum_{A\not\rightarrow B} m(A)$ to mean that the sums must be taken respectively on all $A\in\Omega$ that imply B and on all $A\in\Omega$ that do not imply B. Note that $\mathbf{0}_\Omega$ implies both A and $\neg$A for all $A\in\Omega$.

3.2. The transferable belief model.

Suppose an evidence that induces some belief in us concerning the truth of the propositions A of Ω. It is postulated that there exists some finite amount of belief that can be spread among the various propositions A of Ω, and that given the available pieces of evidence, one allocates parts of that total amount of belief to each proposition.

For instance suppose Mr. White has been murdered and we have three suspects: Henry, Tom and Sarah. Thus Δ = {Henry, Tom, Sarah}. Given the available pieces of evidence, parts of the amount of belief are allocated to each of the three potential murderers. But some evidence might also point to more than just one of the three persons. Such is the case of the evidence "the murderer is a male". This evidence points to A = "Henry $\vee$ Tom" and we must allocate some part m of our total mass of belief to A without being able to split it between the two components of A. In such a situation, probabilists usually invoke the Principle of Insufficient Reason or an argument of symmetry to decide that the mass m must be split into two equal parts, one for Henry and one for Tom. The originality (and the power) of the transferable belief model is that it does not ask for such principles and leaves the mass m allocated to the proposition A.

The model thus postulates a finite amount of belief arbitrary but conveniently scaled to 1, a frame of discernment Ω and a distribution of this total unitary mass among the propositions A of Ω. The non negative mass m(A) allocated to the proposition $A \in \Omega$ that cannot be allocated to any proposition A' such that $A' \rightarrow A$, $A' \neq A$ is called the basic belief mass (the basic probability number of Shafer [25]).

A **basic belief assignment** is a function $m: \Omega \rightarrow [0,1]$ such that:

$$\sum_{A \rightarrow \mathbf{1}_\Omega} m(A) = 1$$

(remember that Ω is the power set build from the elementary propositions of Δ).

Shafer's model includes the further requirement $m(\mathbf{0}_\Omega)=0$. We feel it unnecessary and it leads to unsatisfactory results. $m(\mathbf{0}_\Omega)$ represents the amount of belief that cannot be allocated to any non vacuous propositions of Ω. To understand the meaning of some mass $m(\mathbf{0}_\Omega)$ given to $\mathbf{0}_\Omega$, one must accept the open-world assumption and consider that any amount of belief allocated to a proposition $A \rightarrow \mathbf{1}_\Omega$ is in fact allocated to $A \vee \Theta$ where Θ is the set UP considered in part 2. Then $m(\mathbf{0}_\Omega)$ represents the mass allocated to Θ.

In the Mr. White's case, $m(\mathbf{0}_\Omega)$ corresponds to that amount of belief allocated to none of the three suspects. We must always keep in mind that the murderer might be someone else, e.g. evidence pointing to Sarah and not to Henry and Tom, point in fact to "Sarah or someone not in Δ". In particular $m(\mathbf{0}_\Omega)$ is the amount of belief allocated to the proposition that none of the three suspects is the murderer. Had we received the evidence that the murderer must be one of the three suspects i.e. that the closed-world assumption is true, then this new evidence would have induced some conditioning that would have implied $m(\mathbf{0}_\Omega)=0$ (see part 6). The fact that $m(\mathbf{0}_\Omega)$ might be non null implies that evidence impact is by nature essentially negative in that it allows some propositions to be discarded. Indeed evidence pointing to Sarah essentially does not support 'Henry or Tom'. The method of reasoning simulated by this approach is closer to an elimination process than to a constructive process. A support to a proposition is a non support to its negation taken relative to a closed world.

Shafer's approach postulates beforehand the closed-world assumption. If one defines Ω to include Θ, this would lead to the same results as with the open-world assumption if one is careful never to allocate some masses to propositions of Ω that do not include Θ. We feel it easier to use the restricted Ω and to allow positive masses to $\mathbf{0}_\Omega$, bearing in mind that all masses given to propositions $A \in \Omega$, are always allocated to $A \vee \Theta$, except if the closed-world assumption is explicitly expressed. For simplicity's sake, we drop the Θ, and use the notation A to denote $A \vee \Theta$. This implies that $A \& \neg A = \mathbf{0}_\Omega$ means

$(A \vee \Theta) \& (\neg A \vee \Theta) = \Theta$.

The quantity m(A) measures the amount of belief that is exactly committed to A, not the total belief that is committed to A. Each mass m(A) also supports any proposition B that is implied by A. Therefore the belief that a proposition A is true is obtained by adding all the masses m(B) allocated to propositions B that imply A without implying ¬A (which means that $\mathbf{0}_\Omega$ must be discarded from the sum). The **degree of belief** given to A is quantified by the belief function $bel: \Omega \rightarrow [0,1]$ with:

$$bel(A) = \sum_{\substack{B \rightarrow A \\ B \neq \mathbf{0}_\Omega}} m(B)$$

$$bel(\mathbf{0}_\Omega) = 0$$

Given this definition, it can be shown that any belief function is a capacity of order infinite [4], i.e. satisfies the following inequalities:

1) $bel(\mathbf{1}_\Omega) = 1 - m(\mathbf{0}_\Omega) \leq 1$

2) for every n>0 and every collection $A_1, A_2 \ldots A_n \in \Omega$,

$$bel(\bigvee_i A_i) \geq \sum_i bel(A_i) - \sum_{i>j} bel(A_i \& A_j) \ldots -(-1)^n bel(A_1 \& A_2 \ldots \& A_n) \qquad (3.1)$$

Shafer starts his presentation by requiring that degrees of belief satisfy inequalities (3.1) arguing that the belief in the disjunction of two propositions should at least contain the sum of the belief allocated to each reduced by the belief allocated to their conjunction, equality being unjustified. Unfortunately, this requirement is not sufficient to define belief functions and one must postulate inequalities (3.1) for all n. Critics of Shafer's approach [28] argue against having to postulate all these inequalities, an excessive and not very natural requirement. These criticisms justify why our presentation starts with basic belief masses, not with belief functions.

Given a belief function bel or a basic belief assignment m, the **plausibility** of a proposition A is the sum of the parts of belief that are allocated to propositions B that do not imply ¬A:

$$pl(A) = \sum_{B \not\rightarrow \neg A} m(B)$$

It is related to bel trough

$$pl(A) = bel(\mathbf{1}_\Omega) - bel(\neg A)$$

The meaning of 'belief' and 'plausibility' is still controversial. One might prefer to call bel(A) the degree of minimal (or necessary) entailment (or support) for A, and pl(A) the degree of maximal (or potential) entailment (or support) for A. We will hereafter use the words belief and plausibility as they are those most often used in the present context.

The **commonality function** $q:\Omega \rightarrow [0,1]$ with

$$q(A) = \sum_{B \rightarrow \neg A} m(A \vee B)$$

has no immediate intuitive interpretation (except in the case of disjunctive evidence, not covered here). Its usefulness will appear when one considers Dempster's rule of combination. It represents in fact the maximum value that the mass given to A might reach after combining the belief function that corresponds to q with the belief function induced by any other evidence.

The four functions m, bel, pl and q define each other uniquely. Among them, one has:

$$m(A) = \sum_{B \rightarrow \neg A} (-1)^b q(A \vee B) \geq 0 \qquad \text{with } b = |B| \qquad (3.2)$$

Shafer defines these functions differently: he postulates a null basic belief mass for the contradiction $\mathbf{0}_\Omega$. This difference results from the closed-world assumption accepted a priori by Shafer.

Total ignorance is described by the so called vacuous belief function where:

$m(\mathbf{1}_\Omega) = 1$

$bel(A) = 0$ for all $A \neq \mathbf{1}_\Omega$

$bel(\mathbf{1}_\Omega) = 1$

Total ignorance has always bothered the Bayesians, leading to strong controversies. It is either simply rejected as non existing, a procrustean solution not followed here, or solved by the application of the Principle of Insufficient Reason: if one has k elementary propositions and there is no reason why any should be more supported (credible) than any other, then split the probability mass equally among them. But this does not represent

Total Ignorance. There is no reason why some disjunction of elementary propositions should be more supported than any other. So one must have bel(A) equals some constant $c \geq 0$ for all A in Ω, and not only for the elementary propositions of Δ. Of course such a requirement is impossible with probability functions. With belief functions, this means that with A and B such that $A \& B = \mathbf{0}_\Omega$, one has the inequality $bel(A \vee B) \geq bel(A) + bel(B)$ thus $c \geq 2c$, therefore $c=0$ is the only solution, and it does indeed satisfy all the inequalities (3.1). It corresponds to the highly logical basic belief assignment by which $m(\mathbf{1}_\Omega)=1$ and all other masses are null, $\mathbf{1}_\Omega$ being indeed the only supported proposition.

4. CONDITIONING.

Suppose a basic belief assignment m on Ω obtained after considering some initial evidence. Then, suppose we learn from a new evidence that the truth is necessarily in $B \in \Omega$, thus that all non vacuous propositions implying $\neg B$ should be transferred into KI, the set of propositions known as impossible. How does this evidence modify our basic belief assignment.

Let m' be the basic belief assignment obtained after taking the new evidence into account. To construct m'(A), three situations must be considered depending on the relation between $A \in \Omega$ and the conditioning proposition B.

1) $A \rightarrow B$. The evidence that the truth is in B does not modify the part of our total belief mass supporting A.

2) $A \& B = A_1 \neq \mathbf{0}_\Omega$ and $A \& \neg B = A_2 \neq \mathbf{0}_\Omega$. The mass A was allocated by m to $A_1 \vee A_2$ with $A_1 \rightarrow B$ and $A_2 \rightarrow \neg B$. We learn that the truth is in B, therefore the mass that was allocated to $A_1 \vee A_2$ is transferred to A_1, the only part of A that is compatible with the new evidence that asserts 'the truth is in B'.

3) $A \rightarrow \neg B$. The evidence that the truth is in B tells us that all elementary propositions in A are impossible. Thus the mass m(A) is transferred to $\mathbf{0}_\Omega$ (provided that - being in an open-world - $\mathbf{0}_\Omega$ represents in fact Θ).

Therefore

$$m'(A) = \sum_{C \rightarrow \neg B} m(A \vee C) \qquad \text{for all } A \rightarrow B$$

$$= 0 \qquad \text{otherwise}$$

The major ingredient of the **transferable belief model** is seen in case 2, cases 1 and 3 being particular cases of 2.

When a mass m is allocated to some proposition A, it is acknowledged that this mass could eventually be allocated to any subproposition of A if further evidence became available, but that given the available evidence it cannot be allocated more specifically. That mass m was given to A as there was no reason to allocate it to a more specific proposition. Once the evidence "B is true" becomes available (with B compatible with A, i.e. $A\&B\neq\mathbf{0}_\Omega$), the mass m should be allocated henceforth to A&B, m is transferred to A&B, therefore the name of the model.

Case 3 was not considered by Shafer who renormalizes the basic belief masses, multiplying each m' such that their sum remains 1 and $m'(\mathbf{0}_\Omega)=0$. Up to that renormalization, the proposed model for conditioning corresponds to **Dempster's rule of conditioning**.

It implies:	$m'(\mathbf{0}_\Omega)$	$= bel(\neg B) + m(\mathbf{0}_\Omega)$	
	$bel'(A)$	$= bel(A\vee\neg B) - bel(\neg B)$	for all $A\in\Omega$, $A\neq\mathbf{0}_\Omega$
	$pl'(A)$	$= pl(A\&B)$	for all $A\in\Omega$, $A\neq\mathbf{0}_\Omega$
	$q'(A)$	$= q(A)$	for all $A\rightarrow B$
		$= 0$	otherwise

Returning to Mr. White's case, suppose the evidence "Henry is not the murderer", then B= {Tom, Sarah}. The portion of belief that was allocated to Tom and/or Sarah remains theirs. The portion that was given to 'Henry or Tom' now supports Tom alone, and the portion that was given to Henry is transferred to $\mathbf{0}_\Omega$.

This description of the nature of the transferable belief masses justifies the resulting rule of conditioning. It is part of the whole model and not ad hoc as is often felt.

There are two ways of presenting the use of belief functions. Shafer [25] starts with the concept of a degree of belief represented by a belief function, postulates the inequalities (3.2), derives the non negative masses, postulates Dempster's rule of combination and derives Dempster's rule of conditioning. The transferable belief model first presents the masses, derives the belief functions and the inequalities, and from the intrinsic nature of the masses derives Dempster's rule of conditioning. Dempster's rule of combination is then also derived (see part 5). Shafer's approach naturally leads to the "feeling" that Dempster's rules are ad hoc. Our approach tries to eradicate that erroneous opinion.

5. COMBINATION OF TWO BELIEF FUNCTIONS.

Suppose two belief functions bel_1 and bel_2 induced by two distinct evidences. The question is to define a belief function $bel_{12}=bel_1 \oplus bel_2$ resulting from the combination of the two belief functions, where the $\oplus$ symbolizes the combination operator. Shafer's proposal was to derive bel_{12} from the so called Dempster's rule of combination: the product $m_1(X) \cdot m_2(Y)$ is supporting X&Y

$$m_{12}(A) = \sum_{X\&Y=A} m_1(X) \cdot m_2(Y)$$

This implies the most useful relation that explains the usefulness of the commonality function:

$$q_{12}(A) = q_1(A) \cdot q_2(A)$$

Suppose further a belief function bel_1 and the evidence "B is true". Let bel_{1B} be the result of the conditioning of bel_1 as derived in part 4. Suppose one defines m_B such that $m_B(B)=1$. Then $m_{1B} = m_1 \oplus m_B$, therefore Dempster's rule of conditioning happens to be a particular case of Dempster's rule of combination.

Even though Dempster's rule of combination is natural, some justification is required. A set of axioms is given that indeed leads to Dempster's rule of combination. The importance of these axioms rests in the fact that if Dempster's rule of combination is refuted, some of the axioms must explicitly be rejected. A discussion on the adequacy of the axioms is easier than a discussion based directly on the rule itself.

A1: compositionality axiom.

$bel_{12}(A)$ is a function of A, bel_1 and bel_2 only.

A2: symmetry:

$bel_1 \oplus bel_2 = bel_2 \oplus bel_1$

A3: associativity:

$(bel_1 \oplus bel_2) \oplus bel_3 = bel_1 \oplus (bel_2 \oplus bel_3)$

A4: conditioning:

if bel_2 is such that $m_2(B)=1$, then

$$m_{12}(A) = \sum_{C \to \neg B} m_1(A \vee C) \qquad \text{for all } A \to B$$

$$= 0 \qquad \text{otherwise}$$

The axiom of compositionality A1 claims that the combination is a functional of both belief functions and may be A, but nothing else. This is essentially what was meant by distinct evidences. A concept of distinctness is defined in Smets [37]. Two pieces of evidence are distinct if the knowledge of one of them does not induce a non vacuous belief in the truth of the other. Once Dempster's rule of combination is accepted, distinctness implies compositionality. This concept of distinctness could have been postulated in place of the compositionality, but at the cost of higher complexity.

The axiom of symmetry A2 and the axiom of associativity A3 tell us that the result of the combination of pieces of evidence is independent of the order in which they are considered and/or they are associated.

The axiom of conditioning A4 has been justified in part 4. It implies that if bel_2 is vacuous, $bel_{12}=bel_1$.

In order to prove the unicity of Dempster's rule of combination, it is much easier to work with commonalty functions as Dempster's rule of combination is $q_{12}(A)=q_1(A)q_2(A)$. Many theorems are conveniently described with such commonalty functions. As belief functions are uniquely related to commonalty functions, both approaches are equivalent. Proofs are given in appendix 3.

Theorem 1. Given axioms A1 to A4, there is a function f such that:

$$q_{12}(A) = f(A , \{q_1(B):B\rightarrow A , q_2(B):B\rightarrow A\})$$

Axiom A5 expresses the idea that the result of the combination will not be modified by a permutation among the elementary propositions of Δ.

A5: internal symmetry.

Let $\Delta = (A_1, A_2...A_n)$. Let the propositions $B_1,B_2...B_n$ be a permutation of the propositions $A_1, A_2...A_n$. Let m_i and m_i' be two sequences of basic belief masses with:

$m_i = (m_i(A_1), m_i(A_2), m_i(A_1\vee A_2), m_i(A_3)...m_i(A_1\vee A_2...A_n))$

$m_i'= (m_i(B_1), m_i(B_2), m_i(B_1\vee B_2), m_i(B_3)...m_i(B_1\vee B_2...B_n))$

Let $bel_{12}(A)=g(A,m_1,m_2)$.

Then $\quad g(A,m_1,m_2)=g(A,m_1',m_2')$

Axiom A6 considers that the mass given by m_{12} to $A\in\Omega$ is independent of the masses given by m_1 (and m_2) to propositions $B\rightarrow\neg A$.

A6: autofunctionality:

$\forall A \in \Omega$, $A \neq \mathbf{1}_\Omega$, $m_{12}(A)$ does not depend on $m_1(X)$ for all $X \rightarrow \neg A$.

Theorem 2: Given axioms A1 to A6, there is a function f such that

$$q_{12}(A) = f(\ A\ ,\ q_1(A)\ ,\ q_2(A)\ )$$

Two further technical axioms are necessary to prove the final theorem 3.

A7: three-element:

There are at least three elementary propositions in Δ.

A8: continuity:

Let $m_2(A) = 1-\varepsilon$, $m_2(\mathbf{1}_\Omega) = \varepsilon$. Let $m_A(A) = 1$. For any bel_1 defined on Ω, let $bel_{1A} = bel_1 \oplus bel_A$. Then for all $X \in \Omega$,

$$\lim_{\varepsilon \to 0} m_{12}(X) = m_{1A}(X).$$

The three-element axiom can hardly be criticized.The continuity axiom is in fact needed only to eliminate an uninteresting degenerate solution of theorem 3:

$$\begin{aligned} q_{12}(A) &= q_1(A) && \text{if } q_2(A)=1 \\ &= q_2(A) && \text{if } q_1(A)=1 \\ &= 0 && \text{otherwise} \end{aligned}$$

Axiom A8 could be replaced by the requirement that $q_{12}(A)=f(A,a,b)$, $a,b \in [0,1]$, should be non null somewhere in the open interval (0,1)x(0,1), or the degenerate solution could be explicitly rejected.

Theorem 3: Given axioms A1 to A8, for all $A \in \Omega$

$$q_{12}(A) = q_1(A) \cdot q_2(A)$$

Theorem 3 proves the unicity of Dempster's rule of combination under axioms A1 to A8. The proof is based on the properties of triangular norms and absolute monotone functions presented in appendix 1 and 2.

6. NORMALIZATION.

When Shafer introduced his model, he postulated $m(\mathbf{0}_\Omega)=0$ and $bel(\mathbf{1}_\Omega)=1$. So after combining two belief functions, he had to normalize the results in order to get

$bel_{12}(\mathbf{1}_\Omega)=1$. This is obtained by computing $m_{12}(A)$ as done here and then proportionally rescaling it by a factor $1/(1-m_{12}(\mathbf{0}_\Omega))$. This normalization seems natural but has been seriously criticized by Zadeh [47] with the next counter example.

Suppose a murder case with three suspects: Δ = {Henry, Tom, Sarah}, and two witnesses. Table 1 presents the degrees of belief of each witness about who might be the murderer k. For witness 1, k is not Sarah, k is most probably Henry, but k might also be Tom. Witness 2 holds similar beliefs except for the permutation between Henry and Sarah.

	Witness 1	Witness 2	normalized m_{12}	unnormalized m_{12}
Henry	.99	.00	.00	.00
Tom	.01	.01	1.00	.0001
Sarah	.00	.99	.00	.00

How can these two quite contradictory pieces of evidence be combined? Shafer's normalized solution leads to the conclusion that Tom is certainly the murderer.

Zadeh does not accept this solution as it gives full certainty to a solution (Tom) that is hardly supported at all. In fact, in the totally different situation in which both witnesses might have been sure that Tom was the murderer, the result of the combination would have been the same. The unnormalized solution presented within our theory seems much more realistic as it shows Tom to be slightly supported but $\mathbf{0}_\Omega$ to be highly supported. Bearing in mind the semantic of $\mathbf{0}_\Omega$ given in part 2, the most obvious conclusion in the present situation is that the real murderer must be a fourth person, i.e. the solution is in the set UP and not in the set KP= Δ ={Henry, Tom, Sarah}.

There is of course another way of dealing with the present incoherence. The pieces of evidence are combined by a judge who obtains them from two witnesses, each of whom expresses his own belief. The judge must consider his own belief about the reliability of the witnesses. So one could introduce a meta-belief function representing the degree of belief held by the judge about assertions of each witness. Discounting [25, 33] is one way of taking into account this meta-belief. In the present paper we shall restrict ourselves to the case where the two witnesses are wholly reliable.

What further is it that represents the normalization in the present theory? Suppose we are presented with the evidence: 'The murder is necessarily Henry, Tom or Sarah'. How can we accommodate this 'closed-world conditioning' (UP is empty), i.e. how can we transform m_{12} into m'_{12} so that $m'_{12}(\mathbf{0}_\Omega)=0$. One must somehow reallocate $m_{12}(\mathbf{0}_\Omega)$ to non vacuous propositions of Ω in order to keep the sum of all the masses m'_{12} equal to 1.

The general solution is given by

$$m'_{12}(A) = m_{12}(A) + c(A,m_1,m_2)\, m_{12}(\mathbf{0}_\Omega) \qquad \forall A\in\Omega,\ A\neq\mathbf{0}_\Omega$$
$$m'_{12}(\mathbf{0}_\Omega) = 0$$

If c does not depend on m_1 (nor on m_2 by symmetry as m'_{12} must obey axiom A2), there is a problem if $m_2=m_A$ with $m_A(A)=1$, $A\in\Omega$. One must obtain $pl'_{12}(\neg A)=0$ as m_2 tells us that the truth may not be in $\neg A$, so that one must have $m'_{12}(X) = 0$ for all $X\rightarrow\neg A$. This can only be obtained if c depends on m_A (i.e. m_1 and m_2). Thus Yager's proposal [45] to take $c=0\ \forall A\in\Omega$, $A\neq\mathbf{1}_\Omega$, and $c(\mathbf{1}_\Omega)=1$, is not acceptable as his proposal leads to $pl'_{12}(\neg A) > 0$.

Shafer's solution corresponds to $c(A,m_1,m_2)=m_{12}(A)/(1-m_{12}(\mathbf{0}_\Omega))$. It can be obtained if one requires that relative degrees of belief (or plausibility) should stay constant after considering the closed-world conditioning.

Definition: the **closed-world conditioning** corresponds to the impact of the strictly certain proposition 'UP is empty'.

Axiom A9: Let bel' be the belief function obtained from $bel:\Omega\rightarrow[0,1]$ after closed-world conditioning. Then $\forall A,B\in\Omega$, $A,B\neq\mathbf{0}_\Omega$

$$bel'(A)/bel'(B)=bel(A)/bel(B)$$
$$pl'(A)/pl'(B)=pl(A)/pl(B)$$

and $m'(\mathbf{0}_\Omega) = 0$.

Axiom A9 implies that $bel'(A)=c\cdot bel(A)$ with c independent of A. As $bel'(\mathbf{1}_\Omega)=1-m'(\mathbf{0}_\Omega)=1$, then $c=1/(1-m(\mathbf{0}_\Omega))$, as in Shafer's solution.

The impact of the closed-world conditioning could in fact be represented by a meta-belief function defined on the sets UP, KP and KI. The solution is studied in Smets [33].

In this paper, the combination operator $\oplus$ has been considered with the open-world assumption and it is shown that Shafer's normalization can be assimilated to the impact of the closed-world conditioning. It takes in account Zadeh's criticisms because if the closed-world assumption is true, than the only murderer is Tom as Henry and Sarah have been eliminated by witnesses 1 and 2 respectively.

The real counter intuitive result observed within Zadeh's counter example results not so much from the normalization than from the acceptation of the closed-world assumption. In real world situation, it is obvious that if one can really believe both witnesses, then one should seriously question the closed-world assumption. Solution m_{12} has the advantage of showing the practical impact of the closed-world conditioning that was not visible with Shafer's solution.

7. CONCLUSIONS.

The present paper has presented the **transferable belief model** used to quantify someone's degree of belief about the truth of a set of propositions. A finite amount of belief is distributed among the propositions of a frame of discernment Ω. The non negative mass m(A) quantifies the amount of belief specifically allocated to proposition A, that cannot be allocated to any proposition $B \neq A$ that implies A but that might be allocated to such a proposition B if further evidence permits such a transfer. The degree of belief in a proposition A is the sum of the masses allocated to propositions B that imply A without implying $\neg A$. The degree of plausibility in a proposition A is the sum of the masses allocated to propositions B that are compatible with A.

This model must not be confused with the **upper and lower probabilities** model or some **interval valued probabilities** model, that corresponds usually to the interpretation given by those who use Dempster-Shafer's theory [16]. In these models, one postulates the existence of some probabilities that quantify our degrees of belief, but the exact value of each probability is only known to be between two boundaries. In the transferable belief model, no probability whatsoever is introduced. Probabilities are irrelevant.

The transferable belief model is also different conceptually and mathematically from the **bayesian model**. Conceptually the bayesian assumption that belief is quantified by probabilities rests on betting and decision arguments. The transferable belief model applies at the cognitive level, a level where the concepts of bets and decisions are not required. It is true that once a decision is involved, one must construct a probability function based on the belief function that describes the cognitive state. But nowhere in the bayesian approach is there an argument that requires that, at the cognitive level, beliefs be described by probability functions. The problem of decision making when beliefs are quantified by belief functions is solved in Smets (1989).

Mathematically, a **probability function** is a particular case of a belief function where positive masses are allocated only to elementary propositions of Ω. Dempster's rule of conditioning reduces itself into classical probabilistic conditioning. But Dempster's rule of combination does not have its real counterpart in probability theory. Dempster's rule of combination with closed-world assumption can be seen as identical to the combination of two a posteriori probability functions defined on Ω if the a priori probability function on Ω gives the same probability to every elementary proposition of Ω and if the two pieces of evidence are conditionally independent given each elementary proposition of Ω. Generalization to the case where the a priori probabilities are not constant is immediate if one is careful not to introduce it twice in the ⊕ combination [33]. These mathematical similitudes should not be interpreted as implying that the transferable belief model is a generalization of the bayesian model. We rather see them as two complementary models, the transferable belief model normatively describes cognitive state, the bayesian model normatively describes decision bevahiors.

Any measure of cognitive process gets its meaning only if one can provide an objective tool through which this measure can be assessed. Exchangeable bets is the one used by bayesians. The translator example is the one used for belief functions [30, 40].

One particularity of the transferable belief model is that it allows a **positive mass allocated to the contradiction**. The meaning of such allocation can be understood if one gives due consideration to the difference between the open and the closed-world assumption. The frame of discernment Ω is an a priori construct on which belief is distributed. But one should not ignore that this frame is usually nothing but an intellectual construct and that it may be that none of the propositions of Ω is true. The impact of the closed-world assumption is studied and a normalization coefficient is derived, the result is Shafer's model. The advantage of distinguishing between the open and the closed-world is that it allows evaluation of the degree of conflict among the pieces of evidence as far as Ω is concerned, therefore to judge the appropriateness of the frame of discernment Ω and of the closed-world assumption.

The second part of the paper provides axioms that imply the **Dempster's rule of combination** used to combine two belief functions derived from two distinct pieces of evidence. Distinctness is defined in relation to the compositionality axiom A1. The major axiom is the conditioning axiom A4 that results directly from the structure of the transferable belief model, i.e. a mass initially allocated to some proposition A that could be allocated (transferred) to any subproposition of A by further evidence.

It would have been interesting to construct counter examples for which the autofunctionality axiom A6 would not be fulfilled. This is easy if one restricts the belief functions to the so called **consonant belief functions** by Shafer [25] and necessity functions in Dubois and Prade [9]. These are belief functions for which the masses are allocated on propositions $A_1, A_2 \ldots A_n$ such that $A_i \rightarrow A_{i+1}$: i=1...n-1.. In that case,

$$\mathrm{bel}(A \& B) = \min\{\mathrm{bel}(A) , \mathrm{bel}(B)\}$$
$$\mathrm{pl}(A \vee B) = \max\{\mathrm{pl}(A) , \mathrm{pl}(B)\}.$$

Acceptable solutions are

$$q_{12}(A) = \min \{ T(q_1(t), q_2(t)) : t \in A \}$$

where T is any T-norm. They obey A6 iff $T=T_0$ (the min-based T norm). But consonant belief functions are only particular cases of belief functions and are too restrictive to describe someone's degree of belief.

In conclusion, the transferable belief model presents two characteristics: the masses allocation that leads to superadditive belief functions to describe someone's degree of belief and a rule to combine two distinct evidences. The interest of the first aspect is usually recognized. But the combination rule was felt to be ad hoc by critics [28], especially when they interpret the transferable belief model as an upper and lower probabilities model. This paper provides axioms that explain the meaning of Dempster's rule of combination within the transferable belief model.

Belief functions provide a model that should be most useful in developing Expert Systems that need to handle uncertainty. Its theoretical use for medical diagnosis was considered in Smets [33, 34, 35], the generalization of Bayes's theorem necessary for inferences is developed in Smets [33, 38, 39] and the concept of degree of belief in a fuzzy proposition in Smets [36].

APPENDIX 1. Triangular norms.

The concept of triangular norms (T-norms) are fully developed in Scheizer and Sklar [22, 23, 24], see also Weber [44].

Definition. A T-norm is a function T from [0,1]x[0,1] to [0,1] such that for all a, b, c, d $\in$ [0,1], one has:

1. $T(a,b) = T(b,a)$ symmetry
2. $T(a,T(b,c)) = T(T(a,b),c)$ associativity

3. $T(a,b) \geq T(c,d)$ if $a \geq c$ and $b \geq d$ monotony

4. $T(a,1) = a$ boundary conditions

For any T-norm T, there is:

$$T_W(a,b) \leq T(a,b) \leq T_0(a,b)$$

where T_W, and T_0 are T-norms such that:

$$T_W(a,b) = \begin{cases} a & \text{if } b = 1 \\ b & \text{if } a = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$T_0(a,b) = a \wedge b$$

where $\wedge$ denotes the minimum operator.

A particular T-norm is the product function T_1 with $T_1(a,b) = a \cdot b$.

APPENDIX 2. Monotone functions.

Let a function f(x) defined on a segment 0R. Let the successive differences $\Delta_i f$ be positive on the segment 0R for i=1, 2...n and any positive h

$\Delta_1 f(x) = f(x+h) - f(x) \geq 0$

$\Delta_2 f(x) = f(x+2h) - 2f(x+h) + f(x) \geq 0$

....

$\Delta_n f(x) = f(x+nh)-nf(x+\{n-1\}h)+...\pm f(x) \geq 0$

given all values in the inequalities belong to 0R.

If f satisfies those properties for $n \geq 2$, then f is continuous and admits continuous derivates up to the order n-2 inclusively and at each point a right and a left derivate of order n-1. (Proof is given in Bernstein [2], pg 190 et seq.).

Given n=3, f admits non negative first and second derivates and ([2] pg 193)

$$f(x) = f(0) + xf'(0) + (x^2/2) \int_0^1 (1-u)\, f''(ux)\, du$$

APPENDIX 3. Proofs of theorems 1 to 3.

Given the proofs are much easier when arrived at with commonalty functions, we translate axioms A1 to A4 into axioms Q1 to Q4, their counterpart based on commonality functions. We write q_{ij} for $q_i \oplus q_j$. $\forall A \not\to C$ means $\forall A$ such that $A \& \neg C \neq \mathbf{0}_\Omega$.

Q1: compositionality axioms:

$q_{12}(A)$ is a function of A, q_1 and q_2 only.

Q2: symmetry:

$q_1 \oplus q_2 = q_2 \oplus q_1$

Q3: associativity:

$(q_1 \oplus q_2) \oplus q_3 = q_1 \oplus (q_2 \oplus q_3)$

Q4: conditioning:

if q_2 is such that $m_2(B)=1$, then

$$q_{12}(A) = q_1(A) \quad \text{for all } A \to B$$
$$= 0 \quad \text{otherwise}$$

Theorem 1. Given axioms Q1 to Q4, there is a function f such that:

$$q_{12}(A) = f(\ A\ , \{q_1(B):B \to A\ \}, \{q_2(B):B \to A\ \})$$

Proof: Let q_1, q_2 and q_C be three commonality functions defined on Ω. Let $C \in \Omega$. q_C corresponds to a conditoning on C, i.e. $q_C(A)=1\ \forall A \to C$, and 0 otherwise.

By Q1, $q_{12}(A) = f(A,\{q_1(X):X \to \mathbf{1}_\Omega\},\{q_2(X):X \to \mathbf{1}_\Omega\})$

By Q3, $q_{(12)C} = q_{1(2C)}$.

By Q4, $q_{(12)C}(A)=q_{12}(A)\ \forall A \to C$, and 0 otherwise.

$\forall A \to C$, $q_{1(2C)}(A) = f(A,\{q_1(X):X \to \mathbf{1}_\Omega\},\{q_2(X):X \to C, 0:X \not\to C\})$

and $q_{1(2C)}(A) = q_{12C}(A)$

The comparison of the two f functions shows that $q_{12}(A)$ does not depend on $q_2(X)$ $\forall X \not\to C$. This is true for all C such that $A \to C$ and among others for C = A. Thus $q_{12}(A)$ does not depend on $q_2(X)\ \forall X \not\to A$. It depends only on $q_2(X)$ for $X \to A$.

By symmetry (axiom Q2) one concludes that $q_{12}(A)$ depends only on the $q_1(X)$ and $q_2(X)$ for $X \to A$.

QED.

Lemma 1: Let q_B be the commonality function corresponding to a conditioning on B. Then $q_{BB}=q_B \oplus q_B = q_B A$

Proof: By definition, $m_B(B)=1$, thus by axiom Q4 $m_{BB}(A)=1$ if A=B, and 0 otherwise. Therefore $m_B=m_{BB}$.

QED.

Lemma 2: If for some $B \rightarrow A$, $B \neq A \in \Omega$, $q_2(X)=0$ $\forall X \rightarrow A, \forall X \not\rightarrow B$, then $q_{12}(A)=0$

Proof: Let $B \rightarrow A$, $B \neq A \in \Omega$. Let q_B be the commonality function corresponding to a conditioning on B. By Q3, $q_{(12)B}=q_{1(2B)}$. By Q4 as $B \rightarrow A$, $B \neq A$, $q_{(12)B}(A)=0$ and $q_{2B}(X)=0$ $\forall X \rightarrow A$, $\forall X \not\rightarrow B$. Thus $0 = q_{(12)B}(A) =$

$f(A,\{q_1(X):X \rightarrow A\},\{q_{2B}(X):X \rightarrow A\}) = f(A,\{q_1(X):X \rightarrow A\},\{q_2(X):X \rightarrow B, 0:\forall X \not\rightarrow B\})$

This last term corresponds to $q_{12}(A)=0$ when q_2 satisfies the conditions of the hypothesis.

QED.

Lemma 3: Axioms A6 implies that $\forall A \in \Omega$, $A \neq \mathbf{1}_\Omega$, $q_{12}(A)$ does not depend on $m_1(X)$ for all $X \rightarrow \neg A$.

Proof: By theorem 1, for $A \in \Omega$, $q_{12}(A)$ may depend only on $q_1(B)$ for $B \rightarrow A$ and $q_1(B)$ depends only on those $m_1(C)$ such that $B \rightarrow C$. Thus $q_{12}(A)$ may depend on $m_1(C)$ only if $A \& C \neq \mathbf{0}_\Omega$.

QED.

Theorem 2: Given axioms A1 to A6, there is a function f such that:

$$q_{12}(A) = f(\ A\ ,\ q_1(A)\ ,\ q_2(A)\ )$$

Proof: For simplicity's sake, the domain of the m_1, q_1, m_{12} and q_{12} functions are not repeated. They are all defined on Ω.

For $X \in \Omega$, $|X|$ is the number of elementary propositions of Δ that imply X.

The set of independence pairs is defined as the set of ordered pairs (a,b) where $a=|A|$, $b=|B|$, $a>b$, and (a,b) means that $q_{12}(A)$ is independent of $q_1(B)$, irrespective of the interrelations between A and B.

1°) Consider $A \in \Omega$, $|\mathbf{1}_\Omega|=n$, $|A|=n-1$ and $\neg A=B$. By (3.2), $m_{12}(A) = q_{12}(A) - q_{12}(\mathbf{1}_\Omega)$. By axiom A6, $m_{12}(A)$ does not depend on $m_1(B)$. By lemma 3, $q_{12}(A)$ does not depend on $m_1(B)$. The only component of $q_{12}(\mathbf{1}_\Omega)$ that depends on $m_1(B)$ is $q_1(B)$. Thus $q_{12}(\mathbf{1}_\Omega)$ is independent of $q_1(B)$. By symmetry (axiom Q5), $q_{12}(\mathbf{1}_\Omega)$ is independent of $q_1(B)$ whenever $|B|=1$. So (n,1) belongs to the set of independence pairs.

2°) Consider the set of independence pairs with $i<k<n$,

(n,1)

(n-1,1) (n,2)

(n-2,1) (n-1,2) (n,3)

…

(n-k+2,1)…………..(n,k-1)

(n-k+1,1)...(n-k+i,i)

These independence pairs (r,s) are such that either r-s>n-k or r-s=n-k in which case s=1,2...i. Suppose these independence pairs hold, then the independence pair (n-k+i+1,i+1) holds also.

Consider $A\in\Omega$ such that |A|=n-k and $Y\rightarrow\neg A$ with |Y|=i+1≤k. Relation (3.2) can be written:

$$m_{12}(A) = \sum_{r=0}^{k} (-1)^r \sum_{s=0}^{r\wedge(i+1)} \sum_{\substack{B\rightarrow Y\\|B|=s}} \sum_{\substack{C\rightarrow\neg(A\vee Y)\\|C|=r-s}} q_{12}(A\vee B\vee C)$$

By A6, $m_{12}(A)$ is independent of $m_1(Y)$. Given A,B,C, we test if $q_{12}(A\vee B\vee C)$ may depend on $m_1(Y)$. Only $q_1(D)$ may depend on $m_1(Y)$ only if $D\rightarrow Y$, and by theorem 1, $q_{12}(A\vee B\vee C)$ may depend on q1(D) only if $D\rightarrow A\vee B\vee C$, so we must only test if $q_{12}(A\vee B\vee C)$ depends on $q_1(D)$ for $D\rightarrow B$.

To check if $q_{12}(A\vee B\vee C)$ is independent of $q_1(D)$, one must check if the pair (n-k+r,d) with d=|D| belongs to the set of accepted independence pairs. One has d≤s and $s\leq r\wedge(i+1)$. Therefore n-k+r-d≥n-k. Whenever n-k+r-d>n-k, the pair belongs to the set of accepted independence pairs. When n-k+r-d=n-k, thus r=d, the pair also belongs to the set of accepted independence pairs if d≤i.

The only case not considered is r=d=i+1 in which case r=s, D=B=Y, which corresponds to the only term $q_{12}(A\vee Y)$ that might be dependent only on $q_1(Y)$. As this is the only term that might depend on $m_1(Y)$ and as $m_{12}(A)$ does not depend on $m_1(Y)$, we have proved that $q_{12}(A\vee Y)$ does not depend on $q_1(Y)$. By axiom A5, it implies that the pair (n-k+i+1,i+1) is an independence pair.

3°) Consider the set of independence pairs as in 2°) but with its last term being (n,k-1), i.e. all the pairs (r,s) with r-s>n-k.

We must prove that (n-k+1,1) is also an independence pair. Consider |A|=n-k and |Y|=1, $Y\rightarrow\neg A$. $m_{12}(A)$ depends on $q_{12}(A)$, $q_{12}(A\vee Y)$, $q_{12}(A\vee C)$ and $q_{12}(A\vee Y\vee C)$ |C|=c≥1, $C\rightarrow\neg(A\vee Y)$. $q_{12}(A)$ and $q_{12}(A\vee C)$ do not depend on $q_1(Y)$ as $Y\rightarrow\neg A$. $q_{12}(A\vee Y\vee C)$ does not depend on $q_1(Y)$ as the pair (n-k+1+c,1) belongs to the set of accepted independence pairs as n-k+1+c-1>n-k. The only term that might depend on $q_1(Y)$ is $q_{12}(A\vee Y)$. As $m_{12}(A)$ does not depend on $q_1(Y)$ and as there is only one term in $m_{12}(A)$ that might depend on $q_1(Y)$, it is independent of $q_1(Y)$. By axiom A5, the pair (n-k+1,1) is an independence pair.

4°) By 1° (n,1) is an independence pair. By 3° (n-1,1) is IP. By 2°, (n,2) is an independence pair. By 3°, (n-2,1) is an independence pair etc...Thus $q_{12}(A)$ is independent of all $q_1(Y)$ whenever |Y|<|A|. Therefore $q_{12}(A)=f(q_1(A),q_2(A))$.

QED.

Lemma 4: The f function of theorem 2 is such that $f(A,a,0) = 0$.

Proof: The last relation in the proof of lemma 2 becomes: $0 = q_{(12)B}(A) = f(A,q_1(A),0)$.

QED.

Lemma 5: The f function of theorem 2 is such that $f(A,a,1)=a$.

Proof: Let q_A be the commonality function corresponding to the conditioning on A. Let $q_{1A}=q_1 \oplus q_A$. Then by axiom Q4, $q_{1A}(A)=q_1(A)$, thus $f(A,q_1(A),1)=q_1(A)$.

QED.

Lemma 6: Given axioms A1 to A7, there is a function T such that:

$$q_{12}(A) = T(\, q_1(A) \,,\, q_2(A) \,)$$

Proof: 1) Let $a_i=q_i(A)=q_i(A \vee X)$, i=1,2, with X an elementary proposition of Δ implying $\neg A$. By theorem 2, one has $q_{12}(A) = T_A(a_1,a_2)$. As all q functions are such that $q(A) \geq q(A \vee B)$, $T_A(a_1,a_2) \geq T_{A \vee X}(a_1,a_2)$. It holds for all $A \neq \mathbf{1}_\Omega$.

2) Let X and Y be 2 distinct elementary propositions implying $\neg A$.

Let $c_i = m_i(A \vee X \vee Y)$ and $1-c_i = m_i(A \vee X)$. One has:

$$m_{12}(A) = \sum_{B \to \neg A} (-1)^b \, q_{12}(A \vee B) \qquad \text{with } b=|B|.$$

Let $C=A \vee X \vee Y$, then:

$$m_{12}(A)=T_A(1,1) - T_{A \vee X}(1,1) - T_{A \vee Y}(c_1,c_2) - \sum_{Z \in \neg C} T_{A \vee Z}(0,0) + T_{A \vee X \vee Y}(c_1,c_2) + R$$

where R contains terms $T_B(0,0)$ with $B \not\to C$. By lemma 4, $T_B(0,0)=0$ for all B. By lemma 5, $T_A(1,1)=T_{A \vee X}(1,1)=1$. Thus

$m_{12}(A)=-T_{A \vee Y}(c_1,c_2)+T_{A \vee X \vee Y}(c_1,c_2) \geq 0$,

so $T_{A \vee X \vee Y}(c_1,c_2) \geq T_{A \vee Y}(c_1, c_2)$. It holds for all A such that there are 2 elementary propositions X and Y implying $\neg A$, which is possible from axiom A7.

3) Combining inequalities in 1) and 2), one has $T_{A \vee X}(c_1,c_2)=T_A(c_1,c_2)$ for all A. Thus T does not depend on A and it is true for all $A \in \Omega$ as far as there are at least 2 elementary propositions in Δ.

QED.

Lemma 7: Under lemma 6 conditions, T is non decreasing in its arguments.

Proof: By definition, $q(A) \geq q(A \vee B)$ for all $A,B \in \Omega$. Thus one has $T(q_1(A),q_2(A)) \geq T(q_1(A \vee B),q_2(A \vee B))$. It is equivalent to $T(x+\varepsilon, y+\delta) \geq T(x,y)$ for all ε,

$\delta \geq 0$.

QED.

Lemma 8: Under lemma 6 conditions, the T function is a T-norm.

Proof: T is a function from [0,1]x[0,1] to [0,1]. To be a T-norm, T must satisfy:

1: T(a,b)=T(b,a) which is true by axiom Q2.

2: T must be associative, which is true by axiom Q3.

3: T must be non decreasing in its arguments, which is true by lemma 7.

4: T(1,a)=a, which is true by lemma 5.

QED.

Definition: T'(x,y) and T''(x,y) are the first and second derivates of T(x,y) taken for x.

Lemma 9: Given axioms A1 to A8, the T-norm of lemma 8 is continuous on [0,1], admits non negative derivates of order 2 and a continuous first derivative.

Proof: For all $A \in \Omega$, the commonality functions q satisfy the following inequalities

$$m(A) = \sum_{B \rightarrow \neg A} (-1)^b \, q(A \vee B) \geq 0$$

Let $A \in \Omega$, $B,C,D \in \Delta$, B,C,D be pairwise distinct and $A \vee B \vee C \vee D = \mathbf{1}_\Omega$.

Construct a commonality function q_1 such that $q_1(X)$ depends only on $n=|X|$. Take $m_1(\mathbf{1}_\Omega)=x$ and $m_1(A \vee B \vee C)=m_1(A \vee B \vee D)=m_1(A \vee C \vee D)=\varepsilon$, $x+3\varepsilon \leq 1$, $m_1(X)=0$ for $X=A \vee B, A \vee C, A \vee D, B \vee C, B \vee D, C \vee D, A, B, C, D$. Then $q_1(A)=x+3\varepsilon$, $q_1(A \vee X)=x+2\varepsilon$, $X \rightarrow B \vee C \vee D$, $|X|=1$, $q_1(A \vee X)=x+\varepsilon$, $X \rightarrow B \vee C \vee D$, $|X|=2$, and $q1(A \vee B \vee C \vee D)=x$. Take a commonality function q_2 such that $q_2(A \vee X)=y$, $X \rightarrow B \vee C \vee D$. Let $q_{12}=q_1 \oplus q_2$. The inequalities for m(A) become,

$m_{12}(A \vee B \vee C) = T(x+\varepsilon,y) - T(x,y) \geq 0$

$m_{12}(A \vee B) = T(x+2\varepsilon,y) - 2\,T(x+\varepsilon,y) + T(x,y) \geq 0$

$m_{12}(A) = T(x+3\varepsilon,y) - 3\,T(x+2\varepsilon,y) + 3\,T(x+\varepsilon,y) + T(x,y) \geq 0$

for all $x,y,\varepsilon \geq 0$, given all terms are in [0,1]. With f(x)=T(x,y), f thus admits non negative differences of order 1 to 3 on [0,1]. Given A8 f is thus continuous on [0,1]. It admits non negative derivates f' and f'', and f' is continuous(see appendix 2).

QED.

Lemma 10: Under lemma 9 conditions and given A7, $T''(x,y+\delta) \geq T''(x,y)$.

Proof: Let $A \in \Omega$ and X,Y,Z be 3 distinct elementary propositions of Δ such that $X \vee Y \vee Z = \neg A$. Construct 2 belief functions as follows with $x+2\varepsilon \leq 1$, $y+\delta \leq 1$:

	A	A∨X	A∨Y	A∨Z	A∨X∨Y	A∨X∨Z	A∨Y∨Z	A∨X∨Y∨Z
m_1	0	0	0	0	ε	ε	0	x
q_1	x+2ε	x+2ε	x+ε	x+ε	x+ε	x+ε	x	x
m_2	0	0	0	0	0	0	δ	y
q_2	y+δ	y	y+δ	y+δ	y	y	y+δ	y

The inequality (3.2) for $m_{12}(A)$ becomes:

$T(x+2\varepsilon,y+\delta)-T(x+2\varepsilon,y)-2T(x+\varepsilon,y+\delta)+2T(x+\varepsilon,y)+T(x,y+\delta)-T(x,y) \geq 0$

$T(x+2\varepsilon,y+\delta)-2T(x+\varepsilon,y+\delta)+T(x,y+\delta) \geq T(x+2\varepsilon,y)-2T(x+\varepsilon,y)+T(x,y)$

As T admits a second derivate, one divides both terms by ε, and takes the limit for $\varepsilon \to 0$. One obtains $T''(x,y+\delta) \geq T''(x,y)$.

QED.

Theorem 3: Given axioms Q1 to Q8, q_{12} is such that for all $A \in \Omega$

$$q_{12}(A) = q_1(A) \,.\, q_2(A)$$

Proof: Given lemma 9, one knows that $T'(x,y)$ and $T''(x,y)$ exist and are non negative. Therefore one has the representation (see appendix 2):

$$T(x,y) = T(0,y) + x\, T'(0,y) + (x^2/2)\int_0^1 (1-u)\, T''(ux,y)\, du$$

where T is a T-norm. Thus $T(0,y) = 0$. As $T(x,1) = x$, $T''(x,1)=0$. By lemma 10, $T''(x,y+\delta) \geq T''(x,y)$. Thus $0=T''(x,1) \geq T''(x,y) \geq 0$, and $T''(x,y)=0$. Then $T(x,y) = xT'(0,y)$ for $x \in [0,1)$. T is symmetrical, $T(1,1)=1$ and $T(x,y)$ is continuous as $x \to 1$ by A8, so $T(x,y)=xy$.

QED.

REFERENCES:

1. BARNETT J.A. Computational methods for a mathematical theory of evidence, in: Proceedings Seventh Intern. Joint Conf. on Artificial Intelligence, Vancouver, BC (1981) 868-875.

2. BERNSTEIN S. Leçon sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle. Gauthier-Villars, Paris (1926)

3. BUCHANAN B.G. and SHORTLIFFE E.H. (Eds), Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-

Wesley, Reading, MA (1984)

4. CHOQUET G. Theory of capacities. Annales de l'Institut Fourier, Université de Grenoble, 5 (1953) 131-296.

5. DEMPSTER A.P. Upper and lower probabilities induced by a multplevalued mapping. Ann. Math. Statistics 38 (1967) 325-339.

6. DEMPSTER A.P. A generalization of Bayesian inference. J. Roy. Statist. Soc. B.30 (1968) 205-247.

7. DUBOIS D., FARRENY H. and PRADE H. Sur divers problèmes inhérents à l'automatisation des raisonnements de sens common. AFCET 1985, 5ème Congrès, Grenoble, Tome 1 (1985) 321-328.

8. DUBOIS D. and PRADE H. *Fuzzy sets and systems, theory and applications*. Academic Press (1980).

9. DUBOIS D. and PRADE H. *Theorie des possibilités.* Masson, Paris (1985).

10. FINE T. Theories of probability. Academoic Press, New York (1973).

11. GARVEY T.D., LOWRANCE J.D. and FISCHLER M.A. An inference technique for integrating knowledge from disparate sources, in: Proceedings Seventh Intern. Joint Conf. on Artificial Intelligence, Vancouver, BC (1981) 319-325.

12. GOOD I.J. Probability and the weighting of evidence. Hafner (1950)

13. GORDON J. and SHORTLIFFE E.H. The Dempster-Shafer theory of evidence, in BUCHANAN B.G. and SHORTLIFFE E.H. (Eds), Rule-Based Expert Systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA (1984) 272-292.

14. GORDON J. and SHORTLIFFE E.H. A method for managing evidential reasoning in a hierarchical hypothesis space. Artificial Intelligence 26 (1985) 323-357.

15. KAHNEMAN D., SLOVIC P. and TVERSKY A. Judgement under uncertainty: heuristics and biases. Cambridge Univ.Press, Cambridge (1982).

16. KANAL L.N. and LEMMER J. *Uncertainty in artificial intelligence*. North Holland, Amsterdam, (1966)

17. LINDLEY D.V. Scoring rules and the inevitabilty of probability. Intern. Statist. Rev. 50 (1982) 1-26.

18. LOWRANCE J.D. Dependency-graph models of evidential support. COINS Technical Report 82-26. (1982)

19. MAGREZ P. Modèles de raisonnement approchés. Doctoral dissertation, Université Libre de Bruxelles, Bruxelles (1985) .

20. MAMDANI E.H. and GAINES B.R. Fuzzy reasoning and its applications. Academic Press (1981) .

21. RIVETT B.H.P. Behavioural problems of utility theory. in WHITE D.J. and BOWEN K.C. (Eds), The role and effectiveness of theories of decision in practice. Hodder and Stoughton, London (1975) 21-27.

22. SCHWEIZER B. and SKLAR A. Associative functions and statistical triangle inequalities. Publ. Math. Debrecen, 8 (1961) 169-186.

23. SCHWEIZER B. and SKLAR A. Associative functions and abstract semi-groups. Publ. Math. Debrecen, 10 (1963) 69-81.

24. SCHWEIZER B. and SKLAR A. *Probabilistic metric spaces.* North Holland, New York (1983).

25. SHAFER G. *A mathematical theory of evidence*. Princeton Univ. Press, Princeton, NJ (1976).

26. SHAFER G. Nonadditive probabilities in the work of Bernoulli and Lambert. Arch. History Exact Sci. 19 (1978) 309-370.

27. SHAFER G. Belief functions and parametric models. J. Roy. Statist. Soc. B44 (1982) 322-352.

28. SHAFER G. Lindley's paradox. J. Amer. Statist. Ass. 77 (1982) 325-351.

29. SHAFER G. The combination of evidence. Working paper 162, School of Business, University of Kansas (1984)

30. SHAFER G. and TVERSKY A. Languages and designs for probability judgment. Cognitive Sc. 9 (1985) 309-339.

31. SHORTLIFFE E.H. and BUCHANAN B.G. A model of inexact reasoning in medicine. Math. Biosci. 23 (1975) 351-379.

32. SHORTLIFFE E.H. Computer based medical consultations: MYCIN. American Elsevier, New York (1976)

33. SMETS Ph. Un modèle mathématico-statistique simulant le processus du diagnostic médical. Doctoral dissertation, Université Libre de Bruxelles, Bruxelles, (1978) (Available through University Microfilm International, 30-32 Mortimer Street, London W1N 7RA, thesis 80-70,003)

34. SMETS Ph. Modèle quantitatif du diagnostic médical. Bulletin de l'Académie Royale de Médecine de Belgique 134 (1979) 330-343.

35. SMETS Ph. Medical diagnosis: fuzzy sets and degree of belief. Fuzzy Sets and Systems 5 (1981) 259-266.

36. SMETS Ph. The degree of belief in a fuzzy event. Information Sciences 25 (1981) 1-19.

37. SMETS Ph. Combining non distinct pieces of evidence. Proc.NAFIP86, New Orleans, 544-548, 1986.

38. SMETS Ph. Belief functions and Bayes theorem. Artificial Intelligence and Expert Systems. ed. CHORAFAS D. and ROWE A., AMK, Berlin, 1987.

39. SMETS Ph. Bayes' theorem generalized for belief functions. Proc. ECAI-86, vol II., (1986) 169-171.ECCAI

40. SMETS Ph. Belief functions. in SMETS Ph., MAMDANI A., DUBOIS D. and PRADE H. Non standard logics for automated reasoning. Academic Press, London (1988) 253-286.

41. SMETS Ph. Construucting the pignistic probability function in a context of uncertainty. Proc. Fifth Workshop on Uncertainty in AI, Windsor, Canada, (1989) 319-326.

42. SMITH C.A.B. Consistency in statistical inference and decision. J. Roy. Statist. Soc. B23 (1961) 1-37.

43. STRATT T.M. Continuous belief functions for evidential reasoning, in: Proceedings Fourth National Conf. on Artificial Intelligence, Austin, Tx (1984) 308-313.

44. WEBER S. A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms. Fuzzy Sets and Systems, 11 (1983) 115-134.

45. YAGER R.R. On the Dempster-Shafer framework and new combination rules. Technical report MII-504, Machine Intelligence Institute, Iona College, 1985.

46. ZADEH L. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1 (1978) 3-28.

47. ZADEH L. A mathematical theory of evidence (book review) AI Magazine 5(3) (1984) 81-83.

ACKNOWLEDGEMENTS:

The author is indebted to L. Zadeh, G. Shafer, D. Dubois, H. Prade, R. Yager, I.Vanner and G. Cornfield for their useful comments and suggestions.

2. MATHEMATICAL PROPERTIES OF BELIEF FUNCTIONS.

Let ˝ be a finite set of elementary propositions, i.e. the conjunction of any 2 distinct propositions of ˝ is a contradiction. Let ∏ be the boolean algebra of propositions derived from ˝, i.e. ∏ contains the conjunctions, disjunctions and negations of any set of propositions of ˝. Let π be the tautology relative to ∏ and Ø the contradiction relative to ∏. Ø is also called the vacuous proposition of ∏. By definition Ø is not an elementary proposition of ˝. Negation of any proposition A of ∏ symbolized by ¬A or « is taken relatively to ˝. One writes œ and ¢ for the disjunction and the conjunction connectives, A‘B for 'proposition A implies proposition B'. Note that Ø‘A for all A in ∏.

The ® symbol is used with the following meanings.
A®˝ means that A is an elementary proposition of ˝,
A®∏ means that A is a proposition of ∏ and
for B®∏, A®B means that A is an elementary proposition implying B.
Thus Ø®∏ is true but Ø®˝ and Ø®B are false as Ø is not an elementary proposition.

For any A®∏, ¿@@A@@¿ is the number of elementary propositions B®˝ such that B®A.

All the presentation could have been done using sets, unions, intersections, and inclusions. Our choice reflects a personal preference supported by the feeling that the natural domain of a belief is more the truth of a proposition than the belonging to a set.

A **basic belief assignment m** is a function m:∏Ÿ[0,1] such that:

„ m(A) = 1
A‘π

The sum is taken on all A that imply π, i.e. all propositions A of ∏. Any A®∏ such m(A)>0 is called a focal proposition of ∏.

A **belief function bel** is a function bel:∏Ÿ[0,1] such that:

bel(A) = „ m(B)
B‘A
B≠Ø

The sum is taken on all B®∏ that imply A without implying «. It satisfies the following inequalities:

1) bel(π) = 1 - m(Ø) ≤ 1

2) for every n>0 and every collection A_1,A_2...A_n®∏,

bel(œA_i) ≥ „bel(A_i) - „bel(A_i¢A_j)...+$(-1)^{-1}$bel(A_1¢A_2...¢A_n) (1)
i i i>j

A **plausibility function pl** is a function pl:∏Ÿ[0,1] such that:

pl(A) = „ m(B)
B¢A≠Ø

The sum is taken on all B®∏ that do not contradict A, i.e. all those B that do not imply «. A equivalent definition is:

pl(A) = bel(π) - bel(«) = 1 - m(Ø) - bel(«)

In particular, pl(Ø)=0.

A **commonality function q** is a function q:∏Ÿ[0,1] such that:

q(A) = „ m(AœB)
B‘«

In particular, q(Ø)=1. It satisfies the following inequalities, with b = ¿@@B@@¿@@:

„ $(-1)^{b}$ q(AœB) ≥ 0
B‘«

These 4 functions define each other uniquely. Among other, one has:

m(A) = „ $(-1)^{a-b}$ bel(B)
B‘A
B≠Ø

with a-b = ¿@@@A¢˜@@¿ .

@ m(A) = „ (-1)^{b} q(AœB) (2)

B‘«

bel(A) + m(Ø) = „ (-1)^{b} q(B)

B‘«

with b = ¿@@@B@@¿@.

The **vacuous belief function** is such that:

m(π) @ = 1

Bel(π)@@@@= 1

Bel(A) = 0 for all A ≠ π

q(A) = 1 for all A®∏

Shafer defines these functions differently, the only difference being that he requires a null basic assignment to the contradiction Ø. In order to distinguish the functions defined here and those of Shafer, we use the symbols m, bel, pl and q for those functions as defined here and M, Bel, Pl and Q for the equivalent ones as defined in Shafer.

Shafer's functions M, Bel, Pl and Q obey all the above mentioned rules to which one adds:

M(Ø) @@= 0

what implies:

Bel(Ø) = 0 Bel(π) = 1

Pl(π) @@= 1

In the whole presentation, each time one of the functions m, bel, pl or q is introduced with some supplementary symbols, we will abstain to define each one in relation to the others. This avoids the necessity to define explicitely m_1, pl_1 and q_1 as being the basic probabilty assignment m_1, the plausibility function pl_1 and the commonality function q_1 related to the belief function bel_1. The simple declaration of one of them implies automatically the others, the supplementary symbols being sufficient to know which one are interrelated. The same applies for the functions as defined by Shafer.

Given 2 belief functions bel_1 and bel_2 induced by 2 distinct evidences, the belief function bel that results of their combination is obtained by **Dempster's rule of combination**.

m(A) = „ m_1(AœX) m_2(AœY)

X‘«
Y‘«
X¢Y=Ø

and M(A) = m(A) / (1 - k)

with k = „ m_1(X) m_2(Y) = m(Ø)
X¢Y=Ø

One has also the very useful relations:

q(A) = q_1(A) q_2(A)

and Q(A) = q(A) / (1 - k)

Conditioning on the proposition A is represented by the special case where m_2(A)=1. Let bel_2 represent the belief function induced by a conditioning on A. Then the combination of a belief function bel_1 with that particular conditioning function bel_2 induces the belief function bel such that:

m(X) = „ m_1(XœY)	for all X‘A
Y‘«	
m(X) = 0	for all X¢«≠Ø
bel(X) = bel_1(Xœ«) - bel_1(«)	for all X‘A
O	for all X¢«≠Ø
pl(X) = pl_1(X¢A)	for all X®∏
q(X) = q_1(X)	for all X‘A
0	for all X¢«≠Ø

Bel(X) = (Bel_1(Xœ«) - Bel_1(«)) / (1 - Bel_1(«))

Pl(X) = Pl_1(X¢A) / Pl_1(A)

Q(X) = Q_1(X)/Pl_1(A)	for all X‘A
0	for all X¢«≠Ø

The rules for Shafer's functions are **Dempster's rule of conditioning**.

All proofs for the M, Bel, Pl and Q functions as well as further properties of these can be founded in Shafer [24]. Their extensions to the m, bel, pl and q functions are immediate.

Q1: entailment functionality:

$q_{12}(A)$ is a function of A,q_1 and q_2 only.

Q2: symmetry:

$q_1\$q_2 = q_2\$q_1$

Q3: associativity:

$(q_1\$q_2)\$q_3 = q_1\$(q_2\$q_3)$

Q4: conditioning:

if q_2 is such that $m_2(B)=1$, then

$q_{12}(A) = q_1(A)$ for all A‘B

@@@@= 0 otherwise