The Combination of Different Pieces of Evidence Using Incidence Calculus

Weiru Liu Alan Bundy
Dept. of Artificial Intelligence
Univ. of Edinburgh
Edinburgh EH1 1HN, U.K.

Abstract

Combining multiple sources of information is a major and difficult task in the management of uncertainty. Dempster's combination rule is one of the attractive approaches. However, many researchers have pointed out that the application domains of the rule are rather limited and it sometimes gives unexpected results. In this paper, we have further explored the nature of combination and achieved the following main results. 1). The condition of combination in Dempster's original combination framework is more strict than that required by Dempster's combination rule in Dempster-Shafer theory of evidence. 2). Some counterintuitive results of using Dempster's combination rule shown in some papers are caused by the overlooking (or ignorance) of different independence conditions required by Dempster's original combination framework and Dempster's combination rule. 3). In Dempster's combination rule, combinations are performed at the target information level. This rule itself does not provide a combination mechanism at the original information level so that it is not able to combine the overlapped information. 4). An alternative approach to the combination of different pieces of evidence by using incidence calculus is proposed. In this approach different pieces of evidence are combined at both the original information level and the target information level rather than only at the target information level. 5). In this approach, we can combine not only independent pieces of evidence but also dependent pieces of evidence. 6). This new approach turns out to be consistent with traditional probability theory. It is more powerful than Dempster's combination rule at combining dependent evidence.

1 Introduction

The management of uncertainty within knowledge and evidence includes three main tasks: the representation, propagation and combination of evidence. The combination of different pieces of evidence is the most difficult task in many cases. Up to date, several approaches have been proposed to represent uncertain information, and the corresponding combination mechanisms have been established. Among these approaches, the Dempster-Shafer theory of evidence is quite popular. But the problems in applying this theory, particularly applying Dempster's combination rule have been discussed intensively by many researchers [Hunter 87, Lemmer 86, Pearl 88, 90, Zadeh 84, 86]. Several authors showed that in some situations Dempster's rule gives counterintuitive results. Several other authors [Shafer 82, 90, Smets 88, Ruspini etal 90] disagreed with this criticism and argued that the counterintuitive results are caused by the misapplication of the rule. The discussion on this problem has lead to the topic of how to understand belief functions[1] – functions which assign a number between 0 and 1 to every subset of a given set.

One view of belief functions is that the theory of belief functions is the generalized probability theory. Another view is that a belief function is an alternative way of representing evidence.

Recently, Halpern and Fagin [Halpern and Fagin 92] further explored the nature of belief functions, and clarified these two views in some detail. They indicated "*it seems that all the*

[1] The formal definition of belief functions is given in section 3.1.

examples showing the counterintuitive nature of the rule of combination arise from an attempt to combine two beliefs that are really being viewed as generalized probabilities. If we view beliefs as a generalized probability, then it makes sense to **update** *beliefs but not* **combine** *them. On the other hand, if we view beliefs as a representation of evidence, then it makes sense to combine them, but not update them.*" But the ideal solution should be that the *"same conclusions will be reached no matter which viewpoint is taken"*. In their paper, they concluded that 1) we have to accept the two views of belief functions. 2) thinking of beliefs as generalized probabilities gives more reasonable results in some cases where thinking of beliefs as a representation of evidence doesn't. 3) the examples in [Black 87, Hunter 87, Lemmer 86, Pearl 88] regarding the counterintuitive nature of belief functions can all be explained in terms of a confusion of these two views. That is, the belief functions in those examples can only be interpreted in terms of generalized probabilities.

Several questions have to be answered. Why different viewpoints may result in different solutions? What is the key cause of such problems? Is it because of the different viewpoints of belief functions, the weakness of the rule itself, or the misapplication of the rule?

With respect to these questions we are going to investigate Dempster's combination rule from a different perspective. We will examine both Dempster's original motivation on combination in his original paper and Dempster's combination rule named by Shafer in DS theory. In the rest of the paper, in order to distinguish them we use **Dempster's combination framework** to name the description of combination in Dempster's original paper and **Dempster's combination rule** to stand for the rule commonly used.

Actually, when people refer to DS theory and mention Dempster's combination rule, they all implicitly mean the combination formula[2] in Shafer's book. The applications and problem examinations are focused on this formula which was abstracted from [Dempster 67]. The spirit of the rule can be stated as: if there are two belief functions bel_1 and bel_2, defined by two distincted pieces of evidence, on the same space S, then their joint impact on the space can be represented by $bel_1 \oplus bel_2$, where $\oplus$ means using Dempster's combination rule. The condition of applying $\oplus$ to two belief functions is usually mentioned as independence in many sources, such as in [Shafer 82, P.325]. This more or less gives the condition of using the rule. In other words, the information carried by one belief function tells us nothing about the information carried by another belief function. Some examples given in several articles seem to satisfy this requirement but give counterintuitive results.

In order to explain why Dempster's combination rule is unapplicable in some cases, we go back to explore Dempster's original paper which is the basis of Shafer's work and to see what we can find. Based on Dempster's paper [Dempster 67] we can simply state his idea of combination as follows[3]: suppose there are two pieces of evidence which are given in the form of two probability spaces[4] (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2). Further suppose there is another space S and some kind of mapping relations from space X_1 and X_2 to S. The relation between one space and another space says that the truth of some elements in the former space suggests the possibility of truth of some elements in the latter space. Given the probability of truth of some elements in spaces X_1 and X_2, we are interested in knowing the impact of the evidence on the space S (we may think S contains answers to our questions or the possible values of a variable).

[2] See definition in section 3.1
[3] See detailed analysis in Section 3.3
[4] See definition in Section 2.2

Facing to this problem, Dempster suggested that we can get the joint probability space (X, χ, μ) $= (X_1 \otimes X_2, \chi_1 \otimes \chi_2, \mu_1 \otimes \mu_2)$ out of the two original spaces as well as the joint mapping relation from X to S first and then propagate the effect of probability distribution μ to S.

The condition of obtaining the product space is that "**the sources (if we treat a space and its probability distribution as a source) are assumed independent. ... Opinions of different people based on overlapping experiences could not be regarded as independent sources. Different measurements by different observations on different equipments would often be regarded as independent**" [Dempster 67]. If we refer to the levels containing spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) as the **original information level**, and space S as the **target information level**, then Dempster's condition of independence is assumed at the original information level. This requirement is called *DS-Independent*[5] in [Voorbraak 91].

Under such a requirement, Shafer's simplified procedure can be explained as propagating two probabilistic distributions from X_1, X_2 to S separately, then combining them on the space S. Therefore the condition of applying Dempster's renowned combination rule is that the two pieces of information (in the form of belief functions) are independent (or distinct) on the same domain *i.e.* at the target information level. Clearly if two pieces of information are DS-Independent, then they must be independent in the normal sense and Dempster's combination rule is applicable. The other way around, two independent belief functions may not be DS-Independent if they are rooted at the original information level, that is, if they are in the form of providing evidence rather than in the form of giving an effect on a specific domain (or space). This can be seen in those examples given in the later sections of this paper. There is ignorance throughout the literature that different independence conditions are required by Dempster in his original framework and by Shafer when he gave Dempster's combination rule in his book. A rare exception is Voorbraak who touched upon this point in giving his DS-Independent definition, but even he failed to make this point explicitly.

Therefore by contrast to these two opinions, we argue that the key cause of giving counterintuitive results in using Dempster's combination rule is the **overlooking (or ignorance)** of the condition of combination given in Dempster's original paper. That is, the condition of combination given by Dempster in his framework is not quite the same as what Shafer (in his book), as well as many people, explained. The former is more strict.

The conclusion we get from the above analysis and section 3 is that those counterintuitive examples given in some articles [Black 87, Hunter 87, Lemmer 86, Pearl 88, 90, Voorbraak 91] are caused by such ignorance. In the sense of DS-Independence required by Dempster's combination framework, those examples don't satisfy this requirement, so Dempster's combination framework is not applicable. However if we accept that those examples satisfy the independent requirement needed by Dempster's combination rule so Dempster's combination rule is applicable, but the combined results are counterintuitive. From the former point of view, they are caused by the misapplication of the framework, from the latter point of view they are caused by the weakness of the combination formula. Neither of them is able to deal with those cases. Based on such a discussion, those belief functions, which can only be viewed as generalized probabilities, are precisely the cases which fail to satisfy the requirement of DS-Independence. So Dempster's combination rule is not suitable to cope with them.

[5] See definition in section 3.3

The aim of our research is twofold: arguing the independence requirement among several belief functions when using Dempster's combination rule and proposing an alternative combination mechanism to overcome the weakness of Dempster's combination rule. In this new approach, on the one hand we absorb the combination spirit of Dempster's combination framework, that is, multiple sources of information are described at the original information level and the joint source needs to be constructed before we do the combination. On the other hand we adopt a different methodology which goes deeper in showing the nature of combination than Dempster's combination rule, that is we prefer that the effect of the original information should be seen in the new combination mechanism.

In order to reach this goal, we have to employ an alternative theory of dealing with uncertainty which is similar in some sense to DS theory. The alternative theory should have the ability to carry out set operations at the original information level.

In this paper, we present a new mechanism for the combination of different pieces of evidence by using *incidence calculus* [Bundy 85, 92]. The important feature of incidence calculus is that probabilities are not directly associated with formulae, rather incidences are directly associated with some formulae. The incidence set of a formula, containing a set of possible worlds each of which is associated with a probability, is a set in which this formula is true. So incidence calculus forms an incidence set first for a formula and then calculates its probability (or upper and lower bounds on its probability). It has well defined set operations on the original information space and suits the requirement of carrying out the combination at both the original and target information level. As a consequence, we can deal with multiple-sources of information no matter whether they are dependent or not.

The paper is organized as follows. Section 2 introduces the relevant terminology of the propositional language and the probability structure which will be used later. Section 3 describes Dempster-Shafer theory of evidence and analyzes problems in Dempster's combination rule in greater detail. In section 4, we will briefly introduce the basics of incidence calculus and the main features of the theory which will be used in the subsequent sections. Following this we will discuss the way of representing incomplete information and modelling problems. Section 5 describes the new combination rule in incidence calculus which can handle both dependent and independent pieces of information. Section 6 is about the comparison between DS theory and incidence calculus. We will show that they have the same ability to represent evidence (information). We will also prove that Dempster's combination rule is covered by this new combination rule. Several examples will be given to demonstrate the features of the new rule in section 7. Finally, in section 8, we will summarize the paper and discuss further work.

2 Propositional Language and Probability Structure

2.1 Propositional Language

The language we are using in this paper is a finite propositional language.

Definition 1: Propositional Language:
$\mathcal{L}(P)$ is the propositional language formed from P, where P is a finite set of propositions. $\mathcal{L}(P)$ is the smallest set containing the truth values and the members of P. It is closed under the operations of negation ($\neg$), disjunction ($\vee$), conjunction ($\wedge$) and implication ($\rightarrow$).

Suppose a proposition set P contains $p_1, p_2, ..., p_n$, $\mathcal{A}t$ is the set of *basic elements*, each of which is in the form $p'_1 \wedge ... \wedge p'_n$, where p'_i is either p_i or $\neg p_i$ and $p_i \in P$. Any formula φ in the language set $\mathcal{L}(P)$ can be represented as

$$\varphi = \delta_1 \vee ... \vee \delta_k \quad where \;\; \delta_i \in \mathcal{A}t. \tag{1}$$

2.2 Probability Structure

In the discussion, we will use the formalization about probability structures given by Fagin and Halpern [Fagin and Halpern 1989a].

Definition 2: Probability Space

A probability space (X, χ, μ) has:

X: a sample space *usually containing all the possible worlds;*

χ: a σ-algebra containing some subsets of X, which is defined as containing X and closed under complementation and countable union.

μ: a probability measure which gives $\mu : \chi \to [0, 1]$ and it has the following features:

P1. *$\mu(X_i) \geq 0$ for all $X_i \in \chi$;*

P2. *$\mu(X) = 1$;*

P3. *$\mu(\cup_{j=1}^{\infty} X_j) = \Sigma_{j=1}^{\infty} \mu(X_j)$, if the X_j's are pairwise disjoint members of χ.*

Propagating a probability distribution from a probability space to a language set is done through a mapping π. π is a mapping which associates with each $x \in X$ a truth assignment $\pi(x) : P \to \{\mathbf{true}, \mathbf{false}\}$. Here P is a finite set of primitive propositions as defined in definition 1 and $\mathcal{A}t$ is its corresponding basic element set. We say that p, an element of P, is true at x if $\pi(x)(p)$=**true**; otherwise we say that p is false at x.

If (X, χ, μ) is a probability space and π is such a mapping then a tuple $M = (X, \chi, \mu, \pi)$ is called a *probability structure*. In this way, we can associate with each state x in X a unique basic element of $\mathcal{A}t$ describing the truth values of the primitive propositions in x. That is $\pi : s \to p'_1 \wedge p'_2 \wedge ... \wedge p'_n = \delta_i$ where if $\pi(x)(p_i) = \mathbf{true}$ then $p'_i = p_i$, otherwise $p'_i = \neg p_i$. we use $\pi[x]$ to denote this element. Thus for any subset $X_i = \{x_{i1}, ..., x_{il}\}$ of X, we define $\pi[X_i] = \phi_{X_i}$ where $\phi_{X_i} = \vee_j \pi[x_{ij}]$.

For any formula ϕ in $\mathcal{L}(P)$, ϕ^π is defined as a subset of X containing all the states where ϕ is true, that is $\phi^\pi = \{x \mid \pi(x)(\phi) = \mathbf{true}\}$. Further it is defined that $W_M(\phi)$ is the *weight* or *probability* of ϕ in M (a probability structure), which is calculated from $\mu(\phi^\pi)$. If ϕ^π is measurable, that is when $\mu(\phi^\pi)$ exists, we can talk about the probability of formula ϕ, otherwise we can only calculate the lower and upper bounds on the probability of ϕ. In general, if ϕ^π is not measurable, then we define $W_M(\phi)$=$\mu_*(\phi^\pi)$[6] , which is the inner measure of ϕ in M. In addition, we define that $true^\pi = X$ and $false^\pi = \{\}$.

A subset χ' of χ is called a *basis* of χ if it contains non-empty and disjoint elements, and if χ consists precisely of countable unions of members of χ'. For any finite χ there is a basis of χ and it follows that

[6] μ_* is a inner measure induced by μ.

$$\mu_*(A) = sup\{\mu(X) \mid X \subseteq A \text{ and } X \in \chi\}$$

$$\Sigma_{X_i \in \chi'} \mu(X_i) = 1$$

If χ is finite, then it must have a basis and the basis is unique. In the following, we suppose that we only consider finite probability structures.

3 Dempster-Shafer Theory of Evidence

Even though we are using probabilistic terminology to state DS theory, we do not mean to reject the view of belief functions as representing evidence. This topic will be further discussed in Section 3.2.

3.1 Basics of D-S Theory

Dempster-Shafer theory of evidence, or as it is usually called: *belief function theory* [Shafer 76, Smets 88], associates degrees of belief with every subset of a space which consists of mutually exclusive and exhaustive explanations for a problem. Such a space is named a *frame of discernment* (or frame), and is normally denoted as Θ. A belief function bel on space Θ is required to obey the following three features:

1.$bel(\emptyset) = 0$
2.$bel(\Theta) = 1$
3.$bel(A_1 \cup ... \cup A_n) \geq \Sigma_i bel(A_i) - \Sigma_{i,j} bel(A_i \cap A_j) + \Sigma_{i,j,k} bel(A_i \cap A_j \cap A_k) - + ...$

A belief function is usually described in the form of function called a *mass function m, or a basic probability assignment* which follows $m(\emptyset) = 0$, and $\Sigma_{A \subseteq \Theta} m(A) = 1$. Given a mass function on a frame of discernment, a corresponding belief function can be calculated. The relation between these two functions is:

$$bel(A) = \Sigma_{B \subseteq A} m(B)$$

Similarly, another function called *plausibility function* is defined as

$$pls(A) = \Sigma_{B \cap A \neq \emptyset} m(B) = 1 - bel(\neg A)$$

A subset A of Θ is called a focal element of belief function bel if $m(A) > 0$. If all the focal elements of bel are single elements of Θ, then the mass function m is a probability distribution. In general mass functions are generalized probability distributions. The difference between a mass function and its belief function is that the degree of belief on a subset A of Θ represents our total belief on the set and all its subsets while the mass value of A is the degree of belief exactly assigned to the set and not any of its subsets.

In [Fagin and Halpern 89b], a specific representation of a belief function on a frame of discernment is given as a tuple (X, bel, π) which is called a *DS structure* where X is a frame and π is the same as in probability structures, and $bel : 2^X \rightarrow [0, 1]$ is a belief function.

It has been proved [Fagin and Halpern 1989b] that if (X, χ, μ, π) is a probability structure, then (X, μ_*, π) is a DS structure where μ_* is the inner measure of μ on space X, that is, μ_* is a belief function on X. Because of the mapping π from X to $\mathcal{L}(P)$ and the definition of $W_M(\phi) = \mu_*(\phi^\pi)$, it is easy to see that μ_* also gives a belief function on space $2^{\mathcal{A}t}$ in the sense of the equivalence between a subset A_i of $\mathcal{A}t$ and a formula ϕ_{A_i} in $\mathcal{L}(P)$. This is explained as:

$$A_i \longleftrightarrow \phi_{A_i} \mid \phi_{A_i} = \delta_{i1} \vee \delta_{i2} \vee ... \vee \delta_{in} \ \textit{where} \ \delta_{ij} \in A_i \ \textit{and} \ A_i \subseteq \mathcal{A}t, \tag{2}$$

That is $\pi[X_i]$ is thought of as a subset of $\mathcal{A}t$ if we consider the mapping from X to set $\mathcal{A}t$ while $\pi[X_i]$ is treated as a formula in $\mathcal{L}(P)$ if we consider the mapping from X to the language set $\mathcal{L}(P)$.

Therefore for a basic element set $\mathcal{A}t$, a subset of A_i in $2^{\mathcal{A}t}$ is treated to be equivalent to a formula $\vee\delta_{ij}$ (where $\delta_{ij} \in A_i$) in $\mathcal{L}(\mathcal{A}t)$. In other words, μ_* is a belief function on space $\mathcal{A}t$ in the definition of $\mu_*(A) = \mu_*(\phi_A^\pi)$ when $A \subseteq \mathcal{A}t$. Under this assumption, we can derive the following definition.

Definition 3: Complete DS structures[7]

A structure $(X, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$ is called a complete DS structure in which a belief function on frame $\mathcal{A}t$ can be derived from the probability structure (X, χ, μ, π). When X is finite, this belief function can be constructed by applying the following steps.

1) Let χ' be the basis of χ, $\chi' = \{X_1, ..., X_n\}$;
2) Find $A_{DS} = \{A_1, ..., A_{n_1}\}$ where for each A_i there is at least one X_j which has $\pi[X_j] = \phi_{A_i}$;
3) Define a function m on $\mathcal{A}t$ and let $m(A_i) = \Sigma_{\pi[X_i]=\phi_{A_i}}\mu(X_i)$. It is easy to see that $\Sigma m(A_i) = 1$ for all $A_i \in A$, so m is a mass function;
4) Define $Bel(B) = \Sigma_{A\subseteq B} m(A)$, so Bel is a belief function on $2^{\mathcal{A}t}$.

In DS theory when two independent belief functions are known on the same frame of discernment, their joint impact on that frame can be obtained by using Dempster's combination rule. Dempster's rule is stated as follows:

$$m(C) = \frac{\Sigma_{A\cap B=C} m_1(A)m_2(B)}{1 - \Sigma_{A'\cap B'=\emptyset} m_1(A')m_2(B')}$$

where m_1 and m_2 are two mass functions representing the two belief functions on the frame and A, B, A', B' are arbitrary subsets of the frame of discernment. The advantage of DS theory is that it narrows the hypothesis space using Dempster's combination rule as evidence accumulation.

It has been proved [Fagin and Halpern 89b] that for any DS structure, there is a finite DS structure, and for every DS structure there is an equivalent probability structure. In the following we only consider finite DS structures and finite probability structures without losing generality. Given two DS structures (Θ, bel_1, π) and (Θ, bel_2, π), the combined DS structure will be $(\Theta, bel_1 \oplus bel_2, \pi)$. This is the direct application of Dempster's combination rule. The combination procedure says that we have two mass functions on frame Θ, after we combined them we can propagate the joint impact to the language set through mapping π.

Similarly considering two complete DS structures $(X_1, \mathcal{X}_1, \mu_1, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_1)$ and $(X_2, \mathcal{X}_2, \mu_2, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_2)$, if we consider using the formula above to combine them, the combination procedure will be as follows: from the two complete DS structures, we can calculate two belief functions bel_1 and bel_2 on frame $\mathcal{A}t$ and then use the combination rule to combine them. The condition of doing such combination is that these two belief functions must be independent. Looking at two probability spaces, the independence between bel_1 and bel_2 does not imply the

[7] A similar structure was given in [Correa da Silva and Bundy 90] which is called *Total Dempster-Shafer Structure*. In that structure instead of using π, they used an incidence mapping i which is produced from the mapping function π. The detailed definition can be found in their paper.

independence between two probability distributions on two spaces. This is the point we are going to explore in the Section 3.3.

3.2 Constructing Complete DS Structures for Any Belief Functions

We have just defined complete DS structures and suggested using such a structure to represent both the source of a message and the belief function generated from the message on a frame of discernment. Some one may argue that such a structure can only be used when we view belief functions as generalized probabilities. In this subsection, we are going to show that complete DS structures are sufficient enough to represent any belief functions on a frame. In particular we can also use it to represent a belief function even we view it as an alternative way of representing evidence, at least it is possible and sensible from the computational point of view.

Usually in DS theory, a belief function may be defined on a frame of discernment without giving the source. That is, we only know $(\Theta, 2^{\Theta}, bel)$. In this case, we define $\mathcal{A}t = \Theta$ and $\pi(a_i)(\delta_i) = true$ where $\mathcal{A}t$ is the basic element set of P, and π is the mapping function from $\mathcal{A}t$ to Θ. It is easy to see that bel also gives the same belief function on $\mathcal{A}t$. Hence we have a DS structure (Θ, bel, π). It have been proved [Fagin and Halpern 89b] that for every DS structure (Θ, bel, π) there is an equivalent probability structure (X, χ, μ, π_1). The equivalence here means that the belief function on $\mathcal{A}t$ given by (Θ, bel, π) through π is the same as that given by (X, χ, μ, π_1) through π_1. Therefore we get a complete DS structure $(X, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_1)$. Replacing $\mathcal{A}t$ by Θ (as they are the same), we get $(X, \chi, \mu, \Theta, 2^{\Theta}, \pi_1)$. In other words the belief function given by $(\Theta, 2^{\Theta}, bel)$ can be calculated from the complete DS structure $(X, \chi, \mu, \Theta, 2^{\Theta}, \pi_1)$.

From the above discussion we had the conclusion that every belief function on a frame can be equivalently represented by a complete DS structure on that frame no matter which view of beliefs we take as long as we concern with the calculation procedure. So complete DS structures are sufficient to be used to denote any belief functions (and their sources if we know) in DS theory.

3.3 Problems with Dempster's Combination Rule

In Dempster's original paper, a probability space (X, χ, μ) denotes a piece of information where X is a space, χ is a $\sigma-$algebra of X and μ is a probability measure on χ. The relation among the elements of space X and another space S is given by a multivalued mapping Γ. A multivalued mapping, usually denoted as Γ, is a function which takes every element in a space and maps it to a non-empty subset of another space. From the probability distribution μ on space X, the probability measure on space S is calculated as upper and lower probabilities on all subsets of S. In the case that space S contains a set of propositions, and an element s in Γx means that s is true at x, that is, $\pi(x)(s) = true$ when $s \in \Gamma x$, then Γ is exactly the same as a mapping π. So in general Γ and S represent a wider range of mappings and spaces than π and P. In order to keep consistency in using terminology, in the following we will use P and π to replace S and Γ and use $\pi(x)$ to denote the subset in P where $\pi(x) = \{p_i \mid \pi(x)(p_i) = true\}$. Such replacement will not affect our discussion about the independence requirement among the original information sources.

Suppose n pieces of information are known, *i.e.* (X_i, χ_i, μ_i) for $i = 1, ..., n$, which all have mapping relations (π_i) with another set P, and they are independent, Dempster suggested that the combined source (X, χ, μ) and π are defined in Equation (3).

$$
\begin{aligned}
X &= X_1 \times X_2 \times ... \times X_n \\
\chi &= \chi_1 \times \chi_2 \times ... \times \chi_n \\
\mu &= \mu_1 \times \mu_2 \times ... \times \mu_n \\
\pi(x) &= \pi_1(x) \cap \pi_2(x) \cap ... \cap \pi_n(x)
\end{aligned} \tag{3}
$$

The fourth formula can be restated and explained as

$$\pi(x) = \pi_1'(x) \cap \pi_2'(x) \cap ... \cap \pi_n'(x)$$

where $\pi_i'(x) = \pi_i(x_i)$ when $x \in X_1 \otimes ... \otimes X_{i-1} \otimes \{x_i\} \otimes ... \otimes X_n$.

The meaning behind this set of formulae is that from n independent sources we can get the joint source which denotes the message carried by all separated sources and establish different mapping relations from the joint source to the target space P. Different mapping relations are further unified to get the joint mapping function π and using π the joint probability distribution μ is propagated to P.

The definition of π reflects that $x = (x_1, x_2, ..., x_n) \in X$ is consistent with $p_i \in P$ if and only if p_i belongs to all $\pi_i(x_i)$ simultaneously.

The intuitive meaning of this procedure can be shown in Figure 1.

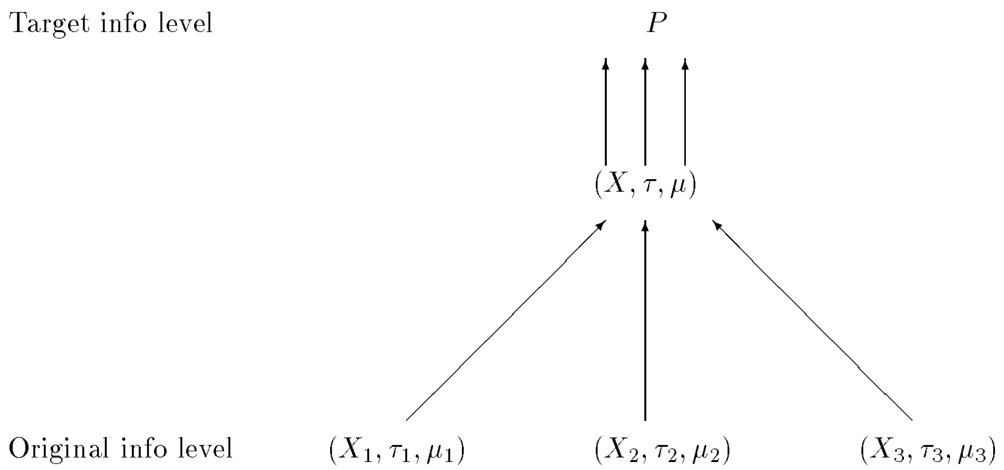

Figure 1 Combination — Propagation

Because neither upper probabilities, nor lower probabilities have a simple product rule of combination, Dempster created another function q, which is called communality functions in DS theory, and q was given as follows:

for a subset $T \subseteq P$, let

$$T' = \{x \in X, \pi(x) \supseteq T\} \quad T_i' = \{x_i \in X_i, \pi_i(x_i) \supseteq T\} \tag{4}$$

and let

$$q(T) = \mu(T') \quad q_i(T) = \mu_i(T_i') \tag{5}$$

Dempster got

$$q(T) = q_1(T) \times q_2(T) \times ... \times q_n(T) \tag{6}$$

In this way Dempster's combination procedure can be explained as propagating different probability distributions μ_i, for $i = 1, ..., n$, from different sources to P first and then producing a unified function q. All the lower and upper probabilities can be calculated on P by using q.

Intuitively Figure 2 demonstrates how those formulae work.

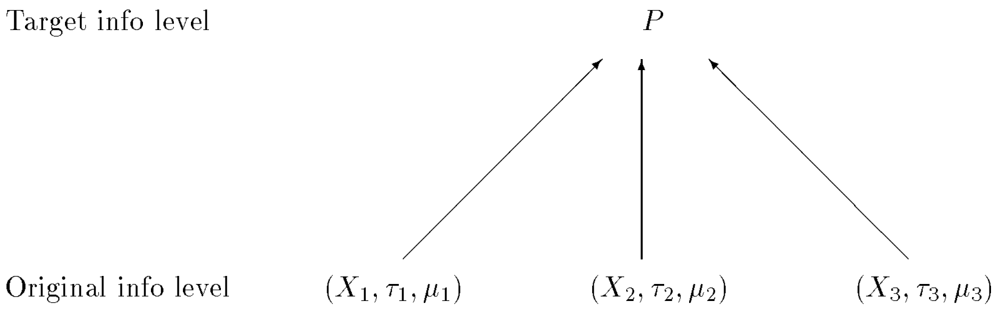

Figure 2 Propagation — Combination

Figure 1 and 2 suggest that there are two ways to combine n different sources:

1) Combining them at the original information level by producing a joint space and a single probability distribution on the space. This should consider the different mappings from the joint space to the target information space, unify these mappings into one mapping and propagate the joint probability distribution to the target information level.

2) Propagating different pieces of evidence at the original information level to the target information level and then combining them.

Dempster assumed implicitly that the results obtained in the above two ways are the same under the condition that n sources are *statistically independent*. When simplifying from Dempster's combination framework to Dempster's combination rule, Shafer adopted the second approach under the condition of independence between the original information sources. But in the simplified combination rule (*i.e.* Dempster's combination rule) the original sources are hidden so that the requirement of independence among the original sources is automatically replaced by the requirement of independence among the different belief functions on the same domain, that is, on the target space P. The invisible of original sources in the simplified combination rule makes it difficult to judge the dependent relations among the belief functions which in turn causes counterintuitive results in many cases.

Actually Dempster's combination framework can be described as: if there are n sources of information which are in the form of complete DS structures $(X_i, \chi_i, \mu_i, S, 2^S, \pi_i)$ and these n sources are independent then the combined result is given in a complete DS structure $(X, \chi, \mu, S, 2^S, \pi)$. What should be the mathematical description of statistically independent required by Dempster's combination framework? The way of describing and judging dependent relations among the original probability spaces is shown by Shafer and Tversky [Shafer and Tversky 1985] and Voorbraak [Voorbraak 91] as follows.

Suppose X stands for an infallible encoded message, where the code is randomly taken from the subsets of list $C_1, C_2, ..., C_n$ and the chance that code C_i is used is p_i (suppose both the list and the associated chances are known). We use a probability space[8] (X, χ, μ) to denote such a piece of information, where X denotes both the message and the space we are concerning, χ is a σ-algebra of the space X and $C_1, ..., C_n$ form its basis, μ is a probability distribution which is defined as $\mu(C_i) = p_i(C_i)$. A complete DS structure is obtained as $(X, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$. If two such complete DS structures are known with two corresponding probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) representing the bodies of evidence where $C_1, ..., C_n$ and $D_1, ..., D_m$ form the bases for χ_1 and χ_2 respectively, that is $\chi_1' = \{C_1, ..., C_n\}$ and $\chi_2' = \{D_1, ..., D_m\}$, then the combination of these two bodies of evidence can be represented by $((X_1, X_2), (\chi_1, \chi_2), (\mu_1, \mu_2), \mathcal{A}t, 2^{\mathcal{A}t}, (\pi_1, \pi_2))$, where (X_1, X_2) denotes both the conjunction of the encoded messages and their joint space, (χ_1, χ_2) denotes the combined σ-algebra of space (X_1, X_2), (μ_1, μ_2) is the probability distribution on the joint space, and (π_1, π_2) is the unified mapping from the joint space to the language set.

Definition 4: DS-Independent [restated from [Voorbraak 91]] *Two complete DS structures* $(X_1, \chi_1, \mu_1, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_1)$ *and* $(X_2, \chi_2, \mu_2, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_2)$ *are called DS-Independent if the two corresponding probability spaces are DS-independent, that is if they satisfy the condition*

$$\mu_\chi(C_i \mid D_j) = \mu_1(C_i) \qquad \mu_\chi(D_j \mid C_i) = \mu_2(D_j)$$

for all codes C_i *and* D_j,

where μ_χ *is the (a priori) probability measure on* χ *which is the* σ*-algebra of the joint space* X*. Spaces* X_1 *and* X_2 *are constructed from the joint space* X*. If the joint space is* $X_1 \otimes X_2$*, then* $\mu_\chi(c_i)$ *is an abbreviation for* $\mu_\chi(\{(c_i, d_j) \mid 1 \leq j \leq m\})$.

According to this definition, only when two complete DS structures are DS-Independent can their combined structure be in the form of $((X_1 \otimes X_2), (\chi_1 \otimes \chi_2), (\mu_1 \otimes \mu_2), \mathcal{A}t, 2^{\mathcal{A}t}, (\pi_1 \cap \pi_2))$ which follows the spirit of Dempster's combination framework. For a case in which multiple sources of information are not DS-Independent, the joint probability space of these pieces of evidence cannot be simply treated as a set product of several individual probability spaces. Rather, every single probability space in a complete DS structure is constructed out of a well defined probability space from a specific perspective. In such a situation, the combination should be carried out on this well defined space by combining different mappings from this space to the language set, and these mappings are given in different complete DS structures.

The discussion on the dependence at the original information level has nothing to do with the mapping π and S , so this condition is also true in the more general cases, that is for multivalued mapping Γ and S.

3.4 An Example

In the following we will examine an example which shows the importance of considering independent relations among the original information sources. If we only require the independence among the belief functions on a target space without considering the relations among the original sources, we will get counterintuitive results.

Here we look at an example given in [Halpern and Fagin92] (which is original from [Hunter 87]). The example is stated as: *Suppose that we have 100 agents, each holding a lottery ticket, numbered 00 to 99. Suppose that agent* a_1 *holds ticket number 17. Assume that the lottery is*

[8] Note that in Voorbraak's paper, he used a pair (X, c), $c = \{c_1, c_2, ..., c_n, P_c\}$, to denote such a message.

fair, so, a priori, the probability that a given agent will win is 1/100. We are then told that the first digit of the winning ticket is 1, the problem is to determine the probability agent 1 will win.

Using DS theory to model this example, two mass functions on space $\{a_1, a_2, ..., a_{100}\}$ are $m_1(a_i) = 1/100$ for $i = 1, ..., 100$ from the priori probability and $m_2(a_1) = 1/10, m_2(S - \{a_1\}) = 9/10$ from the new information that 'the first digit of the winning ticket is 1 '. The combined result is $m(a_1) = 1/892, m(a_i) = 9/892$ for $i = 2, ..., 100$. This result is inconsistent with the straightforward reasoning in the sense of probability which gives the probability that agent 1 will win as 1/10.

Halpern and Fagin argued that there are two objections to the use of the Dempster's combination rule. 1). Considering the meaning in the story, it is hard to think the second mass function (or belief function) is independent of the belief that the lottery is fair. "*In fact, the second mass function is a direct consequence of our belief that the lottery is fair*". 2). In the two views of belief functions, "*the real problem is that we are trying to use the rule of combination with a belief function that is meant to represent a generalized probability*". They gave an alternative way to deal with this example.

Here we consider this example in Dempster's combination framework. Suppose the target space S is $S = \{a_1, a_2, ..., a_{100}\}$, and two original sources are $(X_1, \chi_1, \mu_1, S, 2^S, \Gamma_1)$ where $X_1 = \{00, 01, ..., 99\}$, $\chi_1 = X_1$, $\mu_1(x) = 1/100$ when $x \in \chi_1$, $\Gamma_1(x) = a_i$ when agent a_i's number is x and $(X_2, \chi_2, \mu_2, S, 2^S, \Gamma_2)$ where $X_2 = \{10, ..., 19\}$, $\chi_2 = X_2$, $\mu_2(x) = 1/10$ when $x \in \chi_2$, $\Gamma_2(x) = a_i$ when agent a_i's number is x, then the two original information sources (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) are not independent as they are based on the overlapped information – the lottery is fair. More precisely, these two pieces of information are not DS-independent. For any $c_i \in X_1$ and $d_j \in X_2$, we have $\mu(c_i \mid d_j) = 1$ when $c_i = d_j$ and $\mu(c_i \mid d_j) = 0$ when $c_i \neq d_j$ where μ is the priori probability on space $\{00, ..., 99\}$. Therefore $\mu(c_i \mid d_j) \neq \mu_1(c_i)$. In the same way we have $\mu(d_j \mid c_i) \neq \mu_2(d_j)$. So Dempster's combination framework is not applicable here and Dempster's combination rule cannot be used This conclusion is the same as Halpern and Fagin got but from a different perspective. We also need to point out that the joint original space X in this case is not the product set (that is $X_1 \otimes X_2$) rather it is exactly the same as X_1 and its priori probability distribution is $\mu(x) = 1/100$.

The importance of considering relations among the original information sources has been discussed above. The result tells us that it is more natural to consider the combination at both the original information level and the target information level than only at the target information level. However neither Dempster's combination framework nor Dempster's combination rule provides such combination facilities. The nature of combination is reflected in our new combination mechanism in which the combination is performed at both the original information level and the target information level. In particular the combination is carried out after the joint probability space is found - as a set product of several single spaces or a well defined space. The most important feature of incidence calculus, *i.e. indirect encoding of probabilities on the language set*, over other numerical methods makes it possible to generalize Dempster's combination framework. As a result, using this new approach we can deal with not only DS-Independent pieces of evidence but also the overlapped information.

4 Incidence Calculus

Incidence calculus [Bundy 85, 92] is a method for managing uncertainty in numerical way. Different from other numerical approaches, in incidence calculus probabilities are associated with a set of possible worlds rather with formulae directly. The probability of a formula is calculated through the incidence set assigned to the formula.

4.1 Incidence Calculus

Definition 5: Possible Worlds

Each possible world is a primitive object of incidence calculus which can be thought of as a partial interpretation of some logical formulae.

The probability *is represented by a function ϱ from possible worlds to real numbers between 0 and 1.*

If I is a subset of $\mathcal{W}$ of possible worlds then $wp(I)$ is called the *weighted probability* of I, and is defined to be:

$$wp(I) = \Sigma_{w \in I} \varrho(w) \tag{7}$$

Definition 6: Incidence Calculus Theories

An incidence calculus theory is a quintuple $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$, where:

$\mathcal{W}$ is a finite set of possible worlds.

For all $w \in \mathcal{W}$, $\varrho(w)$ is the probability of w and $wp(\mathcal{W}) = 1$.

$\mathcal{A}t$ is the basic element set of P. $\mathcal{L}(P)$ is the language of the theory.

$\mathcal{A}$ is a distinguished set of formulae in $\mathcal{L}(P)$ called the axioms *of the theory.*

i is a function from the axioms $\mathcal{A}$ to $2^{\mathcal{W}}$, the set of subsets of $\mathcal{W}$. $i(\phi)$ is called the incidence *of ϕ. $i(\phi)$ is to be thought of as the set of possible worlds in $\mathcal{W}$ in which ϕ is true,* i.e. *$i(\phi) = \{w \in \mathcal{W} | \; w \models \phi\}$. It must satisfy the following two conditions:*[9]

$$i(\phi_1 \wedge \phi_2) = i(\phi_1) \cap i(\phi_2) \tag{8}$$

$$i(\perp) = \{\}$$

That is $\mathcal{A}$ is closed under the operator $\wedge$. For any two formulae $\phi_1, \phi_2 \in \mathcal{A}$, if $i(\phi_1) \cap i(\phi_2) \neq \{\}$ then $\phi_1 \wedge \phi_2$ must be in $\mathcal{A}$ and $i(\phi_1 \wedge \phi_2) = i(\phi_1) \cap i(\phi_2)$; otherwise when $i(\phi_1) \cap i(\phi_2) = \{\}$, it doesn't matter whether $i(\phi_1 \wedge \phi_2)$ is in $\mathcal{A}$ as this formula has no effect on further inference. Usually we don't include it in $\mathcal{A}$. However if $\phi_1 \wedge \phi_2 = \perp$ then $i(\phi_1 \wedge \phi_2) = i(\phi_1) \cap i(\phi_2)$ must be empty otherwise the information for constructing the function i implies mistakes. In particular, we always let $i(T) = \mathcal{W}$. Here $\perp$ stands for *False* and T means *True*.

[9] In Bundy's original paper about incidence calculus [Bundy 85], more restrictions were given on an incidence function i. Here we only require that i possesses these two features.

It is not usually possible to infer the incidence of all the formulas in $\mathcal{L}(P)$. What we can do is to define both the upper and lower bounds on the incidence using the functions i^* [10] and i_* respectively. For all $\phi \in \mathcal{L}(P)$ these are defined as follows:

$$i^*(\phi) = \mathcal{W} \setminus i_*(\neg\phi) \tag{9}$$

$$i_*(\phi) = \bigcup_{\psi \to \phi = T} \{i(\psi)\} \tag{10}$$

where $\psi \to \phi = T$ iff $i(\psi \to \phi) = \mathcal{W}$ or we can explain $\psi \to \phi = T$ as $\psi \wedge \phi = \psi$.

For any $\phi \in \mathcal{A}$, we have $i_*(\phi) = i(\phi)$.

The probability of a formula, such as ϕ, is represented using the partial function p from formulae to real numbers in the interval between 0 and 1. When $i(\phi)$ is defined, $p(\phi)$ is defined as:

$$p(\phi) = wp(i(\phi)) \tag{11}$$

For any formula ϕ in $\mathcal{L}(P)$, we can only define its lower and upper bounds on the probability using the function p_* and p^* respectively:

$$p_*(\phi) = wp(i_*(\phi))$$

$$p^*(\phi) = wp(i^*(\phi))$$

If ϕ and ψ are formulas, let $p(\phi \mid \psi)$ be the *conditional probability* of ϕ given ψ. We define:

$$p(\phi \mid \psi) = \frac{p(\phi \wedge \psi)}{p(\psi)} \tag{12}$$

More features of incidence calculus were discussed in [Bundy85, 86, 92].

4.2 Representing Ignorance in Incidence Calculus

In the above subsection we claimed that incidence calculus has a special feature over other numerical approaches to managing uncertainty. This distinguished feature is the association of an incidence set with a formula rather a probability with a formula directly. Therefore the calculation of the probability of a formula is performed after calculating its incidence set. We call this procedure an **indirect encoding** of probabilities. In this section, we will further explore the incidence function i. The result of this investigation shows that we can also represent ignorance in incidence calculus. By contrast to DS theory, the lack of knowledge (or the lack of precise information regarding the problem we concern) is described in the form of incidence function i

[10] The original definition for upper bound on an incidence set is

$$i^*(\phi) = \bigcap_{\phi \to \psi = T} \{i(\psi)\}$$

which was given in [Bundy 1985]. Here we follow the definition given by Correa da Silva and Bundy [Correa da Silva and Bundy 1990]. The meaning of this upper bound is quite similar to the plausibility function in DS theory except the former is an incidence measure and the latter is a probability measure.

in incidence calculus. In this section, we will further show that an incidence function i can be calculated from a more basic function **basic incidence assignment** which is denoted as ii.

In the following, we will see how to find the basic incidence assignment from an incidence function i and how to recover an incidence function i from its basic incidence assignment and why an incidence function i has the ability to represent ignorance.

Given an incidence calculus theory $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$, we have

$$i(\phi \wedge \psi) = i(\phi) \cap i(\psi)$$

where $\phi, \psi \in \mathcal{A}$.

This requirement of i leads us to the conclusion that if $\psi \rightarrow \phi = T$ then $i(\psi) \subseteq i(\phi)$. As we assume that P is finite, then $\mathcal{A}t, \mathcal{L}(P) and \mathcal{A}$ are all finite. We also have the assumption that any formula in $\mathcal{L}(P)$ is in the form of

$$\varphi = \delta_1 \vee \delta_2 \vee ... \vee \delta_n \ \ where \qquad \delta_i \in \mathcal{A}t$$

If a subset $\mathcal{A}_0$ of $\mathcal{A}$ is chosen as $\mathcal{A}_0 = \{\psi_1, ..., \psi_n\}$, then $\mathcal{A}_0$ satisfies the condition that

$$\forall \psi_i \in \mathcal{A}_0, \forall \phi \in \mathcal{A}, \ \ if \ \ \phi \neq \psi_i \ \ then \ \ \phi \rightarrow \psi_i \neq T$$

Therefore, $\mathcal{A}_0$ contains the "smallest" formulae in $\mathcal{A}$ and $\mathcal{A}_0$ is not empty. In fact, we can get $\mathcal{A}_0$ using the following procedure. For a formula $\psi_i \in \mathcal{A}$, if $\exists \phi \in \mathcal{A}, \phi \neq \psi_i$ and $\phi \rightarrow \psi_i = T$, then we use ϕ to replace ψ_i and repeat the same procedure until we obtain a formula ϕ_j and we cannot find any formula which makes ϕ_j true, and ϕ_j will be in $\mathcal{A}_0$. For any two formulae $\psi_i, \psi_j \in \mathcal{A}_0$, when $\psi_i \neq \psi_j$ we have

$$i(\psi_i) \cap i(\psi_j) = \{\}$$

In fact if $i(\psi_i) \cap i(\psi_j) = \mathcal{W}_0 \neq \{\}$, then

$$\mathcal{W}_0 = i(\psi_i \wedge \psi_j) \Longrightarrow$$
$$\psi = \psi_i \wedge \psi_j \neq false \in \mathcal{A} \Longrightarrow$$
$$\psi \rightarrow \psi_i = T, \ \ \psi \rightarrow \psi_j = T \Longrightarrow$$
$$\psi_i, \psi_j \notin \mathcal{A}_0$$

Contradictory! So we have $i(\psi_i) \cap i(\psi_j) = \{\}$.

For any formula ϕ_i in $\mathcal{A} \backslash \mathcal{A}_0$, there are $\psi_{i1}, ..., \psi_{il} \in \mathcal{A}_0$ where $\psi_{ij} \rightarrow \phi_i = T$. So $i(\psi_{ij}) \subseteq i(\phi_i)$ and $(\bigcup_j i(\psi_{ij})) \subseteq i(\phi_i)$.

From a function i, we can obtain another function ii using the following procedure:

Step 1: for any formula $\psi \in \mathcal{A}_0$, define $ii(\psi) = i(\psi)$.

Step 2: define a subset $\mathcal{A}_1$ of $\mathcal{A}$ as $\mathcal{A}_0$ and update $\mathcal{A}$ as $\mathcal{A} \setminus \mathcal{A}_1$.

Step 3: chose a formula ϕ in $\mathcal{A}$ which satisfies the requirement that there are $\psi_{i1}, ..., \psi_{il} \in \mathcal{A}_0$ where $\psi_{ij} \rightarrow \phi_i = T$ and for any $\phi_j \in \mathcal{A}$, $\phi_j \neq \phi$, then $\phi_j \rightarrow \phi \neq T$.

Step 4: update $\mathcal{A}_1$ as $\mathcal{A}_1 \cup \{\phi\}$ and delete ϕ from $\mathcal{A}$. If $\mathcal{A}$ is empty then terminate the procedure otherwise go to step 3.

We call the function ii the **basic incidence assignment**. A possible world w in $ii(\phi)$ means that w makes formula ϕ true but doesn't make any subformula of ϕ true. Following this explanation we immediately have

$$ii(\phi_i) \cap ii(\phi_j) = \{\} \qquad where \phi_i \neq \phi_j$$

We can have actually the following inference procedure.

$$\begin{gathered} w \in ii(\phi_i) \cap ii(\phi_j) \Longrightarrow \\ w \in i(\phi_i) \ and \ w \in i(\phi_j) \Longrightarrow \\ w \in i(\phi_i) \cap i(\phi_j) \Longrightarrow \\ w \in i(\phi_i \wedge \phi_j) \Longrightarrow \\ w \in i(\phi) \ \exists \phi \neq \perp \wedge \phi = \phi_i \wedge \phi_j \Longrightarrow \\ w \notin i(\phi_i) \setminus i(\phi) \ and \ w \notin i(\phi_j) \setminus i(\phi) \Longrightarrow \\ w \notin ii(\phi_i) \cap ii(\phi_j) \end{gathered}$$

Conflict.

Further defining $ii(T) = \mathcal{W} \setminus \cup_j ii(\phi_j)$, if $ii(T) \neq \{\}$ then $ii(T)$ represents those possible worlds which only make T true. This is also an alternative way to represent ignorance. That is, based on the current information we don't know which formula $ii(T)$ makes true except T. Given a basic incidence assignment ii, it is easy to calculate the incidence set of any formula in $\mathcal{A}$ as

$$i(\phi) = \bigcup_{\phi_j \to \phi = T} ii(\phi_j)$$

Therefore

$$p(\phi) = wp(i(\phi)) = \Sigma_{\phi_j \to \phi = T} wp(ii(\phi_j))$$

When a set of axioms is fixed, an incidence function i and its basic incidence assignment are unique to each other. If we replace $wp(i(\phi))$ by $bel(\phi)$ and substitute $wp(ii(\phi_j))$ with $m(\phi_j)$, then the above mathematical equation will be

$$bel(\phi) = \Sigma_{\phi_j \to \phi = T} m(\phi_j)$$

which is very similar to the relationship between a belief function and its mass function.

So it would be interesting to examine the formal relations among DS theory and incidence calculus in representing evidence theoretically. This will be discussed in Subsection 6.1.

4.3 Modeling a Problem in Terms of Incidence Calculus

Modelling a problem in incidence calculus can be done as follows:

1) form a set P consisting of propositions we are interested in.

2) form a set $\mathcal{W}$ consisting of all the possible worlds and determine a probability distribution on it.

3) define function i between sets P and $\mathcal{W}$ which gives the interrelations among their elements.

In order to see how this works in practice, we consider a simple example which is adopted from [Bundy 92].

Suppose there are two propositions, $P = \{rainy, windy\}$, and seven possible worlds, $\mathcal{W} = \{sun, mon, tues, wed, thus, fri, sat\}$. Assume that each possible world is equally probable, *i.e.* occur 1/7 of the time. Through a piece of evidence, we learn that four possible worlds *fri, sat, sun, mon* make *rainy* true, and three possible worlds *mon, wed, fri* make *windy* true. Therefore the incidence sets of these two propositions are:

$$i(rainy) = \{fri, sat, sun, mon\}$$

$$i(windy) = \{mon, wed, fri\}$$

As $i(rainy \wedge windy) = i(rainy) \cap i(windy)$, we also have $i(rainy \wedge windy) = \{fri, mon\}$. So the set of axioms $\mathcal{A}$ is $\mathcal{A} = \{rainy, windy, rainy \wedge windy\}$. The corresponding incidence calculus theory is

$$< \mathcal{W}, \varrho, P, \mathcal{A}, i >$$

and the $\mathcal{A}t$ of P is $\mathcal{A}t = \{rainy \wedge windy, rainy \wedge \neg windy, \neg rainy \wedge windy, \neg rainy \wedge \neg windy\}$.

From this we can calculate the upper and lower bound on the incidence sets of all other formulae in the language set of $\mathcal{L}(P)$. For instance:

$$i_*(\neg rainy) = \{\}$$

$$i^*(\neg rainy) = \{tues, wed, thus\}$$

The inference mechanism of incidence calculus begins with the assumption that some incidence sets have been assigned to the axioms. But in some cases, an uncertainty inference problem assigns probabilities to the axioms rather than incidence sets. It is then necessary to re-discover the incidence sets of the axioms which respect the assignment of probabilities and correlations. This topic has also been discussed in [Bundy 92] and [Liu and Bundy 92]. In the following, we always assume that we can define the initial incidence sets for axioms for a given problem.

Suppose we have already had an incidence calculus theory, $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$, for a given problem, if a new piece of information regarding this problem is known, then it may have one of the following effects:

1) This piece of information gives a new probability distribution on the set of possible worlds to replace the old probability distribution, then the new incidence calculus theory will be created to substitute the old one and the further inference will be made upon the new incidence calculus theory.

Considering the 100 agents problem here, we can first form an incidence calculus theory as $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ where $\mathcal{W} = \{00, ..., 99\}$, $\varrho(w) = 1/100$, $P = \{a_1, ..., a_{100}\}$, $\mathcal{A} = P$ and

$i(a_i) = \{w\}$ when a_i's number is w. Here a_i stands for a proposition a_i *will win*. When we are told that *the first digit of the winning ticket is 1 later*, the probability distribution on $\mathcal{W}$ will be changed as $\varrho_1(w) = 1/10$ when w is in $\{10, ..., 19\}$ and $\varrho_1(w) = 0$ otherwise. Therefore the new incidence calculus theory is $< \mathcal{W}, \varrho_1, P, \mathcal{A}, i >$. It is then easy to know that the probability that a_1 will win is $1/10$.

2) This piece of evidence specifies a new incidence function from sets $\mathcal{W}$ to $\mathcal{L}(P)$ without changing the set of possible worlds and its probability distribution. Then a new incidence calculus theory is formed. Both the new and old incidence calculus theories will make impacts on $\mathcal{L}(P)$. So it is necessary to consider how to obtain their joint impact.

Considering the weather example again, we have, first of all, an incidence calculus theory as $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$. If a new piece of information tells us that $i_1(rainy) = \{fri, sat, sun\}$ and $i_1(windy) = \{wed, fri\}$, then another incidence calculus theory $< \mathcal{W}, \varrho, P, \mathcal{A}_1, i_1 >$ is formed which gives an alternative interrelation among the elements of the two sets. We need to consider the joint impact of both the old and new information on the formula set. That is we must combine the two pieces of information. For a particular formula in $\mathcal{L}(P)$, if we have $i(\phi) = \mathcal{W}_1$ and $i_1(\phi) = \mathcal{W}_2$ respectively, then it is natural to infer that $i \cap i_1(\phi) = \mathcal{W}_1 \cap \mathcal{W}_2$. More generally if $i(\phi) = \mathcal{W}_1$ and $i_1(\psi) = \mathcal{W}_2$ then $i \cap i_1(\phi \wedge \psi) = \mathcal{W}_1 \cap \mathcal{W}_2$. This is the basic idea of giving a combination mechanism in incidence calculus which will be further discussed in greater detail in the next section.

3) This piece of information defines a new incidence calculus theory different from the above two cases. Like situation 2) both the new and old incidence calculus theories will make impacts on $\mathcal{L}(P)$, so it is necessary to consider how to obtain their joint impact. If the old one is $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ and the new one is $< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1 >$, then we form two probability spaces $(\mathcal{W}, \mathcal{W}, \varrho)$ and $(\mathcal{W}_1, \mathcal{W}_1, \varrho_1)$. However different from situation 2), these two probability spaces are not the same.

In this case our purpose is to find the joint space of these two probability spaces and to modify two incidence functions from the joint space to the language set. When these two space are DS-independent, their joint space will be the set product. If they are not DS-independent, then the approaches to constructing the joint space vary from problem to problem. After the joint space is fixed and the new incidence functions are established, the principle in situation 2) can be used.

In summary, apart from some very simple cases shown in situation 1), usually when the new pieces of information are obtained it is necessary to combine them with the existing information. The corresponding combination mechanism is essential to play such a role in producing the final effect of all the information. Currently incidence calculus doesn't have such a facility to cope with this problem. So it is important to propose a combination mechanism in incidence calculus to combine multiple pieces of evidence and to compare it with Dempster's combination rule.

5 Combining Different Pieces of Evidence

This section describes an alternative approach to the combination of different pieces of evidence in incidence calculus, which can solve the problems in applying Dempster's combination rule.

5.1 Relations among Multiple Pieces of Evidence

It is not easy to define a combination mechanism to deal with both dependent and independent evidence without getting a clear picture about the relations among multiple pieces of evidence. In the following we will examine three cases first in order to explore the nature of the relation among multiple sources of information.

CASE 1: Suppose we have two pieces of evidence which define two probability structures $(X_1, \chi_1, \mu_1, \pi_1)$ and $(X_2, \chi_2, \mu_2, \pi_2)$. Two complete DS structures can be formed. If the two corresponding probability spaces are DS-independent, then these two pieces of evidence can be combined using Dempster's combination rule. Suppose the two mass functions produced from them are m_1 and m_2, applying Dempster's combination rule we can obtain their joint impact on $2^{\mathcal{A}t}$.

In fact in such a situation, because the two probability spaces (X_1, χ_1, μ_1) and (X_2, χ_2, μ_2) are statistically independent, that an element $x_1 \in X_1$ makes a subset S_1 of S true does not affect whether an element $x_2 \in X_2$ makes S_1 true. So it is possible to extend π_1 and π_2 as the new mapping relations between the joint set of $X_1 \otimes X_2$ and the space S. Using the extended mappings π'_1 and π'_2, another two complete DS structures can be formed as

$$(X_1 \otimes X_2, \chi_1 \otimes \chi_2, \mu_1 \otimes \mu_2, \mathcal{A}t, 2^{\mathcal{A}t}, \pi'_1)$$

$$(X_1 \otimes X_2, \chi_1 \otimes \chi_2, \mu_1 \otimes \mu_2, \mathcal{A}t, 2^{\mathcal{A}t}, \pi'_2)$$

where $\pi'_1[< x_1, x_2 >] = \pi_1[x_1]$ and $\pi'_2[< x_1, x_2 >] = \pi_2[x_1]$. That is the two pieces of evidence provide two mapping relations from the joint space $X_1 \otimes X_2$ to $2^{\mathcal{A}t}$.

CASE 2: (from [Smets and Hsia 90]) Assume P is a set of propositions $\{Bi, Pe, Fl\}$ where Bi for Bird, Pe for Penguin and Fl for Fly. In common sense we can form two rules $Bi \rightarrow Fl$ with belief .9 and $Pe \rightarrow \neg Fl$ with belief .95. When we learn that $Tweety$ is a bird, we can conclude that $m(Fl) = .9$. When we also learn that $Tweety$ is in fact a penguin, we can also conclude that $m'(\neg Fl) = .95$. Using Dempster's rule to combine m and m' on frame $\Theta = \{Fl, \neg Fl\}$, we can eventually obtain $m''(Fl) = .31, m''(\neg Fl) = .66, m''(\Theta) = .03$.

Obviously the intuitive result should be $m''(\neg Fl) = .95$. So Dempster's rule fails to deal with this case because of the dependence of evidence – the first and second mass functions are all (or indirectly) dependent on an object $Tweety$. The fact that $Tweety$ is a penguin should in some way block the inference of $Bi \rightarrow Fl$.

Suppose that observation X provides the information that $Tweety$ is a bird and observation Y provides the information that $Tweety$ is a penguin. Let W be $\{Tweety\}$, X and Y tell us that $\mathcal{W}$ supports statements Bi and Pe respectively. So we have $\{Tweety\} \wedge \{Tweety\}$ makes $Bi \wedge Pe$ true. That is $\{Tweety\}$ makes Pe true. In this way we can get the correct result. An alternative way to solve this problem in DS theory was discussed in [Smets and Hsia 90]. Formally in such a situation we could construct a set of possible worlds W concerning this problem, and we have the result that if a subset W_1 makes Y true then it must also make the observation X true and Y implies X. The more general situation is the information carried by X and Y may partially affect each other.

CASE 3: Another kind of problem arising from applying Dempster's combination rule is the so called 'sample space problem' [Lemmer 86 , Voorbraak 91]. It was argued that if two observations come from the different aspects of the same sample space, the two mass functions yielded from the observations could not be combined by Dempster's combination rule.

Here we look at an example given by Voorbraak [Voorbraak 91].

There are 100 balls in an urn which are labelled as shown in Table 1.

Label	*Number of Balls*
axy	4
ax	4
ay	16
a	16
bxy	10
bx	10
by	20
b	20

Table 1. 100 balls and their labels

Suppose X and Y are separate observations. The information carried by them is:

X: Drawing a ball from the urn and the ball has label x;

Y: Drawing a ball from the urn and the ball has label y.

Let a space be $\Theta = \{a, b\}$. Drawing a ball from the urn, among several labels of the ball, the labels of the ball make one and only one element of Θ true at each time, so Θ is a frame of discernment. Then the two observations X and Y give two pieces of evidence in the form of mass functions on Θ as:

$$m_X(a) = 2/7, \;\; m_X(b) = 5/7$$

$$m_Y(a) = 2/5, \;\; m_Y(b) = 3/5$$

where $m_X(\{a\})$ $(m_Y(\{a\}))$ is the mass value given by observation X (Y) which represents the possibility of a ball having label a when the ball is observed having label x (y).

The result of applying Dempster's combination rule to the above two mass functions is $Bel_X \oplus Bel_Y(b) = 15/19$.

While in probability theory, the probability that a ball has both label x and y is

$$p(x \wedge y) = 0.14 = 0.28 \times 0.5 = p(x)p(y)$$

Therefore, we have $p(b \mid x \wedge y) = 5/7$. Obviously the results obtained in DS theory and in probability theory are not the same. See the detailed analysis of the example in [Voorbraak 91].

Similar to the previous example in some sense, observations X and Y are governed by the same set of possible worlds $\mathcal{W}$ which contains 100 balls and a known probability distribution on $\mathcal{W}$ even though there are no explicit implications between X and Y. Each of these observations specifies some relations from $\mathcal{W}$ to proposition set $P = \{a, b\}$.

In summary, from the above three cases we can see that giving two observations X and Y, no matter whether the two probability spaces produced from X and Y are DS-independent or not, it is always possible to restate the effect of X and Y in means of defining different mapping relations from a unified space (or set of possible worlds) to the target space. So considering how to combine two mapping relations from one space X to another space S as the basis of providing an alternative combination rule seems both reasonable and possible. This discussion leads us to the definition of the combination rule in incidence calculus.

5.2 The Combination Rule in Incidence Calculus

As we discussed before, if we perform the combination by only considering the information carried by observations X and Y, then the result might be wrong as shown in cases 2 and 3, even though sometimes we could not see the relations between two pieces of information explicitly as in case 3. The more natural way of considering and doing such combinations is to trace the original information source which provides the basis for observations X and Y. They establish two relationships between the propositional set and the unified space (or more generally between the language set and the unified space). Therefore the nature of the combination is to combine these two (or more than two) relation specifications into one relation and then propagate the probability from the space to the language set. The most important step in this new combination mechanism is to form a unified probability space and then carry out the combination on it. For situations in which several pieces of evidence are not DS-Independent, we assume that we can trace this unified space. Therefore, what we need to do is to construct a unified probability space for DS-independent pieces of evidence.

Suppose we have two incidence calculus theories:

$< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2 >$

where $(\mathcal{W}_1, \mathcal{W}_1, \varrho_1)$ and $(\mathcal{W}_2, \mathcal{W}_2, \varrho_2)$ are two DS-independent probability spaces which carry two pieces of information. All of them have contributions to the problem solving, that is they all have effect on space $\mathcal{A}t$. So we need to combine them in order to get their joint impact on space $\mathcal{A}t$.

The basic principle of combination is if w_{1i} makes ϕ true, and w_{2j} makes ψ true, then $< w_{1i}, w_{2j} >$ makes $\phi \wedge \psi$ true. In other words, If w_{1i} is in the incidence set of formulae ϕ in theory one, w_{2j} is in the incidence set of formulae ψ in theory two, then we will conclude that the pair $< w_{1i}, w_{2j} >$ will be in the incidence set of formulae $\phi \wedge \psi$ in the combined theory.

Because these two incidence calculus theories are given independently, there may be conflicts between them. When $(\phi \wedge \psi)$ is false, we have to rule the pair $< w_{1i}, w_{2i} >$ out of the joint set of possible worlds. If we keep it in the joint set of possible worlds, its probability must be 0 because it makes $\perp$ true. Alternatively if we could confirm that the pair $< w_{1i}, w_{2j} >$ should be in the joint set of possible worlds and it does make some formulas true later on, then we have to trace the previous incidence calculus theories which must have been ill-defined. In this case we need to re-define those incidence calculus theories, and re-combine them again.

In principle, when two incidence calculus theories are combined, $\mathcal{W}_1 \otimes \mathcal{W}_2$ will be the joint set of possible worlds, and its probability distribution will be $\varrho'(< w_{1i}, w_{2j} >) = \varrho_1(w_{1i}) \times \varrho_2(w_{2j})$. However as we discussed above, some pairs of $\mathcal{W}$ should be ruled out from the joint set of possible worlds. Suppose the set we take out from $\mathcal{W}$ is $\mathcal{W}_0$ the elements of which make $\perp$ true. The adjusted probability distribution should be given as:

$$\varrho(< w_{1i}, w_{2j} >) = \frac{\varrho_1(w_{1i})\varrho_2(w_{2j})}{1 - \Sigma_{w' \in \mathcal{W}_0}\varrho_1(w'_{1i})\varrho_2(w'_{2j})}$$

Here $w' = (w'_{1i}, w'_{2j})$. It is easy to prove that $\Sigma_{w \in W \setminus W_0}\varrho(w) = 1$.

For an incidence calculus theory $< W_i, \varrho_i, P, A_i, i_i >, (i = 1, 2)$, if $ii_i(T) \neq \{\}$ then $ii_i(T)$ reflects our ignorance (we cannot provide more precise information at the current stage). This ignorance might be changed or modified to produce a meaningful result when more information is available. So adding T to $\mathcal{A}_i$ is essential when we consider combining it with another incidence calculus theory while they are DS-independent; otherwise it is not necessary to do so even in combining two dependent incidence calculus theories.

Definition 7

Suppose we have two incidence calculus theories

$$< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1 > \textit{ and } < \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2 >$$

and their probability spaces are DS-independent, then another two incidence calculus theories can be constructed from them as: $< \mathcal{W}_3, \varrho_3, P, \mathcal{A}_1, i'_1 >$ *and* $< \mathcal{W}_3, \varrho_3, P, \mathcal{A}_2, i'_2 >$

where

$$\mathcal{W}_0 = \bigcup_{\phi \wedge \psi = \perp} i_1(\phi) \otimes i_2(\psi)$$

$$\mathcal{W}_3 = \mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$$

$$i'_1(\phi \in \mathcal{A}_1) = (i_1(\phi) \otimes \mathcal{W}_2) \setminus \mathcal{W}_0$$

$$i'_2(\psi \in \mathcal{A}_2) = (\mathcal{W}_1 \otimes i_2(\psi)) \setminus \mathcal{W}_0$$

the new probability distribution on $\mathcal{W}_3$ *is:*

$$\varrho_3(w) = \varrho_3(< w_{1i}, w_{2j} >) = \frac{\varrho_1(w_{1i})\varrho_2(w_{2j})}{1 - \Sigma_{w' \in \mathcal{W}_0}\varrho_1(w'_{1i})\varrho_2(w'_{2j})} \tag{13}$$

Where $\perp$ means false, $w =< w_{1i}, w_{2j} >$, $w' =< w'_{1i}, w'_{2j} >$, $\otimes$ means a set product, and $\Sigma_{w' \in \mathcal{W}_0}\varrho_1(w'_{1i})\varrho_2(w'_{2j})$ is the weight of the conflict between two theories. If the conflict part is 1 then these two pieces of information are completely conflict with each other and they cannot be combined.

In general for any two pieces of evidence on the two sets of possible worlds, we could always produce a common set of possible worlds and a probability distribution on it. In summary, no matter in which case (DS-Independent or dependent) we can always assume that a unified space and its probability distribution are known based on two or more observations. These observations establish different mapping relations between the space and the language set of propositions. The purposes of the combination are to obtain the unified mapping relation and to propagate the probability using the unified relation. These are performed by the rule below.

Combination Rule

Suppose there are two incidence calculus theories $< \mathcal{W}, \varrho, P, \mathcal{A}_1, i_1 >$, $< \mathcal{W}, \varrho, P, \mathcal{A}_2, i_2 >$, where $\mathcal{W}$ is a set of possible worlds and ϱ is a probability distribution on $\mathcal{W}$. Given two incidence functions i_1 and i_2 from two observations X and Y, then the joint impact of information carried by the two theories is represented by a quintuple as: $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ where

$$\mathcal{A} = \{\varphi \mid \varphi = \phi \wedge \psi, \textit{where } \phi \in \mathcal{A}_1 \wedge \psi \in \mathcal{A}_2 \wedge \varphi \neq \perp\}$$

$$i(\varphi) = \bigcup_{(\phi \wedge \psi \rightarrow \varphi) = T} i_1(\phi) \cap i_2(\psi) \qquad \varphi \in \mathcal{A}$$

and let

$$i(\perp) = \{\} \qquad i(T) = \mathcal{W}$$

Here we need to prove that $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ is also an incidence calculus theory, as stated in the following theorem.

Theorem 1 *Suppose we have two incidence calculus theories $< \mathcal{W}, \varrho, P, \mathcal{A}_1, i_1 >$, and $< \mathcal{W}, \varrho, P, \mathcal{A}_2, i_2 >$, if by applying the above combination rule to them, we get $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ then it is an incidence calculus theory.*

PROOF

Because the set of possible worlds and its probability distribution are not changed during the combination we only need to prove that i has the feature defined in section 4.1.

For any $w \in \mathcal{W}$, if $w \in i(\varphi_1 \wedge \varphi_2)$, then we have

$$\begin{aligned}
& w \in i(\varphi_1 \wedge \varphi_2) \\
\Longleftrightarrow\ & (\exists \varphi_0)(\varphi_0 \rightarrow \varphi_1 \wedge \varphi_2 = T) \wedge (w \in i(\varphi_0)) \\
\Longleftrightarrow\ & (\exists \varphi_0)(\varphi_0 \rightarrow \varphi_1 = T) \wedge (\varphi_0 \rightarrow \varphi_2 = T) \wedge (w \in i(\varphi_0)) \\
\Longleftrightarrow\ & (\exists \varphi_0)(i(\varphi_0) \subseteq i(\varphi_1)) \wedge (i(\varphi_0) \subseteq i(\varphi_2)) \wedge (w \in i(\varphi_0)) \\
\Longleftrightarrow\ & (\exists \varphi_0)(i(\varphi_0) \subseteq i(\varphi_1) \cap i(\varphi_2)) \wedge (w \in i(\varphi_0)) \\
\Longleftrightarrow\ & w \in i(\varphi_1) \cap i(\varphi_2)
\end{aligned}$$

So

$$i(\varphi_1 \wedge \varphi_2) = i(\varphi_1) \cap i(\varphi_2)$$

Function i has all the features of Equation (8) in Section 4 and it is an incidence function.

$< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ is an incidence calculus theory.

END

When $\mathcal{A} = \{\}$ and $i(\perp) = \{\}$ (before we assign $i(\perp) = \{\}$ artificially) then these two observations are irrelevant with each other and their combined result tells us nothing.

When $i(\perp) \neq \{\}$ (before we assign $i(\perp) = \{\}$ artificially) these two observations imply mistakes. We need to re-define the incidence calculus theories to cancel the effect of mistakes. This phenomenon only appears when we try to combine two dependent pieces of information as for the independent situation we have already taken the subset, which makes $\perp$ true, out of the joint space before we combine them.

When $\mathcal{A} \neq \{\}, \forall \phi \in \mathcal{A}, i(\phi) = \{\}$ and $i(\perp) = \{\}$, these two observations repel each other. In other words, only one of them is held at each time.

The intuitive meaning of the Combination Rule is shown in Figure 3.

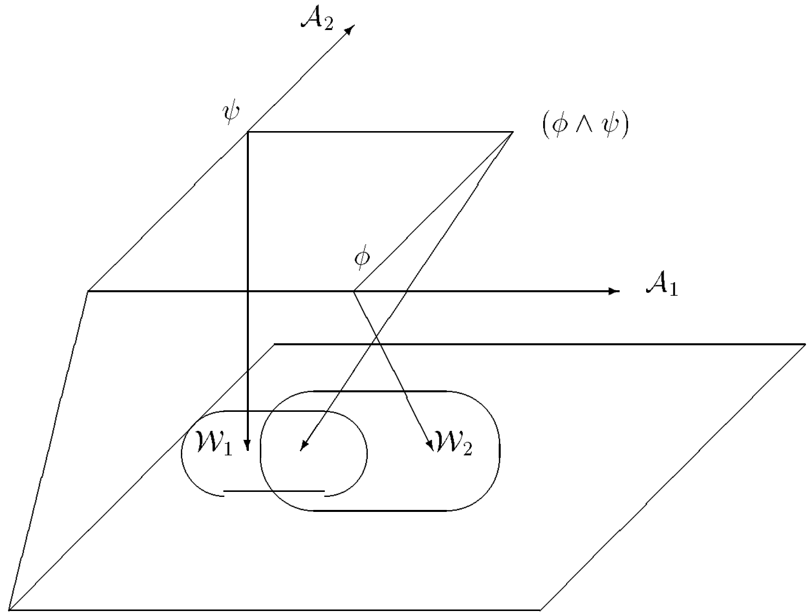

Figure 3. Combining two incidence calculus theories

> The lower level space represents the unified space of two incidence calculus theories. $\mathcal{A}_1$ and $\mathcal{A}_2$ are the axioms of the two theories, and ϕ, ψ are two elements of $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively. The higher level space contains the conjunction of elements of $\mathcal{A}_1$ and $\mathcal{A}_2$. A vector from a formula (such as ϕ) to a subset of possible worlds (such as $\mathcal{W}_1$) means this subset is the incidence set of the formula. The incidence set of $\phi \wedge \psi$ is $\mathcal{W}_1 \cap \mathcal{W}_2$ in this example.

The theoretical explanation of this combination rule is that if observation X says that $\mathcal{W}_{1i}$ makes statement ϕ true, and observation Y says that $\mathcal{W}_{2j}$ makes statement ψ true, then $\mathcal{W}_{1i} \cap \mathcal{W}_{2j}$ should make statement $(\phi \wedge \psi)$ true when we know that both X and Y hold. Therefore, all the conflicts between two theories are caused by ill-defined axioms or incidence functions. In other words, we suppose that the probability distribution on the set of possible worlds $\mathcal{W}$ is correct and consistent. When two incidence calculus theories are specified reasonably, there should be no conflict at all.

Corollary 1 *Suppose we have two incidence calculus theories,* $< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1 >$ *and* $< \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2 >$*, where* $(\mathcal{W}_1, \mathcal{W}_1, \varrho_1)$ *and* $(\mathcal{W}_2, \mathcal{W}_2, \varrho_2)$ *are two DS-Independent probability spaces. Applying the Combination Rule to them we get* $< \mathcal{W}_3, \varrho_3, P, \mathcal{A}_3, i_3 >$ *which is an incidence calculus theory.*

where $\mathcal{W}_3$ *and* ϱ_3 *are the same as in definition 7,*

$$\mathcal{A}_3 = \{\varphi \mid \varphi = \phi \wedge \psi, \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, and\ \varphi \neq \perp\}$$

and

$$\begin{aligned} i_3(\varphi) &= \bigcup_{(\phi\wedge\psi\rightarrow\varphi)=T} (i'_1(\phi) \setminus \mathcal{W}_0) \cap (i'_2(\psi) \setminus \mathcal{W}_0) \\ &= \bigcup_{(\phi\wedge\psi\rightarrow\varphi)=T} (i_1(\phi) \otimes i_2(\psi)) \setminus \mathcal{W}_0 \end{aligned}$$

It is easy to prove this corollary by using definition 7 and theorem 1. For any formulae φ in $\mathcal{L}(P)$, if we know $i_3(\varphi)$, then the probability of φ is

$$p(\varphi) = \Sigma_{w\in i_3(\varphi)}\varrho_3(w)$$

The Combination Rule is both commutative and associative because $A \cap B$ and $B \cap A$ are the same when A and B are two subsets of a space. For the joint product of spaces X_1 and X_2 an element $< x_{1i}, x_{2j} >$ in $X_1 \otimes X_2$ means that both possible worlds x_{1i} and x_{2j} are chosen to support a formula. An element $< x_{2j}, x_{1i} >$ in $X_2 \otimes X_1$ implies the same meaning as $< x_{1i}, x_{2j} >$. Therefore we treat $X_1 \otimes X_2$ as the same as $X_2 \otimes X_1$. So the result of combining several incidence calculus theories is unique irrespective of the sequence in which they are combined. The relations between this Combination Rule and Dempster's combination rule will be discussed in Section 6.

6 Comparison with DS Theory

In this section we are going to discuss the relations between DS theory and incidence calculus. We want to discuss their ability in presenting evidence and compare their ability in combining evidence. We will prove that 1) they have the same ability in presenting evidence and 2) any two pieces of evidence which can be combined in DS theory, can also be combined in incidence calculus by applying the Combination Rule we proposed and that they obtain the same results. We will also show that the new combination rule can combine dependent evidence as well.

6.1 Comparison I: Representing Evidence

In subsection 4.1 we defined a complete DS structure for the representation of a piece of evidence and its effect on frame $\mathcal{A}t$. We start with a complete DS structure and prove that any complete DS structure can be reformed as an incidence calculus theory. The other way around, any incidence calculus theory can also be replaced by a complete DS structure.

Definition 8: Producing an Incidence Calculus Theory from a Complete DS Structure

Suppose $(S, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$ *is a complete DS structure,* χ' *is the basis of* χ*, and* $A_{DS} = \{A_1, ...A_{N1}\}$ *is given in definition 3.*

1) let the set of possible worlds be χ'*, and the probability distribution* ϱ *be* μ*;*

2) let a subset $\mathcal{A}'$ *of* $\mathcal{A}t$ *be* $\{\phi_{A_i} \mid \phi_{A_i} \leftrightarrow A_i, A_i \in A_{DS}\}$*;*

3) define basic incidence assignment ii as $ii(\phi_{A_i}) = \{X_j \mid \pi[X_j] = \phi_{A_i}\}$*;*

4) let $\mathcal{A}$ be $\mathcal{A}'$ and expand $\mathcal{A}$ as follows. For any $\phi_1, \phi_2 \in \mathcal{A}'$, if $\phi_1 \wedge \phi_2 \neq \perp$ and $\phi_1 \wedge \phi_2 \notin \mathcal{A}'$, then define $ii(\phi_1 \wedge \phi_2) = \{\}$ and let $\mathcal{A} := \mathcal{A} \cup \{\phi_1 \wedge \phi_2\}$. Eventually $\mathcal{A}$ contains $\mathcal{A}'$ and is closed under operator $\wedge$.

5) define incidence function i from ii as $i(\phi_i) = \{ii(\phi_j) \mid \phi_j \to \phi_i = T\}$.

Then $< \chi', \mu, \mathcal{A}t, \mathcal{A}, i >$ is an incidence calculus theory (it is easy to prove that i has the features of Definition 6 in Section 4).

This definition leads us to the next corollary.

Corollary 2 *Suppose $(S, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$ is a complete DS structure, χ' is the basis of χ, and m is a mass function on $\mathcal{A}t$ given in definition 3. Let $< \chi', \mu, \mathcal{A}t, \mathcal{A}, i >$ be the corresponding incidence calculus theory, then for any formula $\phi_A \in \mathcal{A}$ we have*

$$p(\phi_A) = wp(i(\phi_A)) = \Sigma_{B \subseteq A} m(B)$$

It is easy to prove this corollary from definition 3 and 8 and the features of the basis of χ. We will use this result many times in proving the next two theorems.

Theorem 2 *If $(S, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$ is a complete DS structure, $(\chi', \mu, \mathcal{A}t, \mathcal{A}, i)$ is the incidence calculus theory produced from definition 8, then for any subset A_i of $2^{\mathcal{A}t}$ and its corresponding formula ϕ_{A_i} in $\mathcal{L}(\mathcal{A}t)$, $Bel(A_i)$ in the DS theory is equal to $wp(i_*(\phi_{A_i}))$ in the incidence calculus theory. That is*

$$Bel(A_i) = p(\phi_{A_i}) = wp(i_*(\phi_{A_i}))$$

PROOF

For any formula ϕ_A in $\mathcal{L}(\mathcal{A}t)$ and its related subset A of $2^{\mathcal{A}t}$, we have

$$\begin{aligned}
& wp(i_*(\phi_A)) = wp(\textstyle\bigcup_{\phi_{A_i} \to \phi_A = T} i(\phi_{A_i})) \\
& = wp(\textstyle\bigcup_{\phi_{A_i} \to \phi_A = T} \{X_j \mid \pi[X_j] \to \phi_{A_i} = T\}) \\
& = wp\{X_j \mid \pi[X_j] \to \phi_A = T\} \\
& = \Sigma_j \mu(X_j \mid \pi[X_j] \to \phi_A = T) \\
& = \Sigma_l m(A_l \mid \pi[X_j] = \phi_{A_l} \; and \; \phi_{A_l} \to \phi_A = T) \\
& = \Sigma_l m(A_l \mid \pi[X_j] = \phi_{A_l} \; and \; A_l \subseteq A) \\
& = Bel(A)
\end{aligned}$$

Then the belief function $Bel(A_\phi)$ is exactly the same as $wp(i_*(\phi))$.

End

This theorem tells us that the belief function on frame $\mathcal{A}t$ produced by a complete DS structure is the same as the lower bound of the probabilities on the formulae if we think of $\mathcal{A}t$ as a basic element set. Because any belief function can be stated in the form of a complete DS structure, we have the conclusion that any belief function can be obtained as a lower bound from an incidence calculus theory.

Definition 9: Producing a Complete DS Structure from an Incidence Calculus Theory

Suppose $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ *is an incidence calculus theory and ii is the corresponding basic incidence assignment,*

1) define a subset $\mathcal{A}_{DS}$ *of* $\mathcal{A}t$ *as* $A_{DS} = \{A \mid ii(\phi_A) \neq \{\}, \phi_A \in \mathcal{A}\}$.

2) if $\bigcup_{\phi_A} ii(\phi_A) \neq \mathcal{W}$, *then* $A_{DS} := A_{DS} \cup \{\mathcal{A}t\}$ *where* $ii(\mathcal{A}t) := \mathcal{W} \setminus \bigcup_{\phi_A} ii(\phi_A)$.

3) define $m(A_i) = wp(ii(\phi_{A_i}))$ *where* $\phi_{A_i} \in A_{DS}$. *Then* $\Sigma_{A_i} m(A_i) = 1$.

4) let $bel(B) = \Sigma_{A_i \subseteq B} m(A_i)$.

So $(\mathcal{A}t, bel)$ gives a belief function on $\mathcal{A}t$. Based on the discussion in subsection 4.2, we have a complete DS structure $(S, \chi, \mu, \mathcal{A}t, 2^{\mathcal{A}t}, \pi)$ which is produced from the incidence calculus theory.

This definition tells us that from an incidence calculus theory we can produce a belief function on the set of basic elements. It is also easy to prove that the lower bound of the probabilities given by the incidence calculus theory is the same as the belief function defined by definition 9. Hence incidence calculus and DS theory have the equivalent ability in representing evidence. The same result has also been proved in [Correa da Silva and Bundy 90].

6.2 Comparison II: Combining DS-independent evidence

For any two complete DS structures, by applying Dempster's combination rule to the two mass functions on $2^{\mathcal{A}t}$, the third mass function and its belief function can be obtained. Obviously, from these two complete DS structures, two incidence calculus theories can also be produced, and their combination can lead to the third incidence calculus theory. What we need to prove in such a situation is that the combined result of the two complete DS structures turns out to be equivalent to the third incidence calculus. The following theorem answers this question.

Theorem 3 *Suppose* $(S_1, \chi_1, \mu_1, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_1)$ *and* $(S_2, \chi_2, \mu_2, \mathcal{A}t, 2^{\mathcal{A}t}, \pi_2)$ *are two complete DS structures and they are DS-Independent,* Bel_1 *and* Bel_2 *are two belief functions on* $2^{\mathcal{A}t}$ *from these two structures respectively and their combined belief function is* Bel. *Further let* $< \chi_1', \mu_1, \mathcal{A}t, \mathcal{A}_{1_{ic}}, i_1 >$ *and* $< \chi_2', \mu_2, \mathcal{A}t, \mathcal{A}_{2_{ic}}, i_2 >$ *be the two incidence calculus theories produced from these DS structures, then the combined incidence calculus theory is equivalent to* $Bel_1 \oplus Bel_2$. *That is, for any subset A of* $\mathcal{A}t$, $Bel(A)$ *is the same as* $wp(i_*(\phi_A))$ *in the combined incidence calculus theory.*

$$Bel(A) = Bel_1 \oplus Bel_2(A) = wp(i_*(\phi_A))$$

Our proof is divided into two parts. In part one we need to prove that the conflict weight k in the combined DS structure is equal to $wp(\mathcal{W}_0)$ in the combined incidence calculus theory. In part two we need to prove that $Bel(A_i) = wp(i_*(\phi_{A_i}))$.

PROOF

Suppose the two bases of two $\sigma-$algebra of S_1 and S_2 are:

$$\chi_1' = \{X_1, X_2, ..., X_n\}$$

$$\chi_2' = \{Y_1, Y_2, ..., Y_m\}$$

the two sets in DS structures produced from χ'_1 and χ'_2 are

$$A = \{A_1, A_2, ..., A_n\} \ \Sigma m_1(A_i) = 1$$
$$B = \{B_1, B_2, ..., B_n\} \ \Sigma m_2(B_j) = 1$$

where A_i and B_j are given in definition 5.

Furthermore the two sets of axioms in the two incidence calculus theories are:

$$A_{1ic} = \{\phi_{A_1}, \phi_{A_2}, ..., \phi_{A_n}\}$$
$$A_{2ic} = \{\psi_{B_1}, \psi_{B_2}, ..., \psi_{B_m}\}$$

Part One

Part one proves $k = wp(\mathcal{W}_0)$ where k is the weight of the conflict between these two DS structures, and $\mathcal{W}_0$, which is defined in section 5.2, is the conflict set in the combined incidence calculus theory.

Step I

Suppose $m = m_1 \oplus m_2$, if $A_i \cap B_j = \{\}$, then $m_1(A_i)m_2(B_j)$ will be a part of k. That is $k = k' + m_1(A_i)m_2(B_j)$.

Because of $A_i \leftrightarrow \phi_{A_i}$ and $B_j \leftrightarrow \psi_{B_j}$, we have $\phi_{A_i} \wedge \psi_{B_j} =\perp$.

According to the definition of the Combination Rule, we have $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}) \subseteq \mathcal{W}_0$.

We further have

$$m_1(A_i) = \Sigma_k \mu_1(X_{ik}) \ \ when \ \ \pi_1[X_{ik}] = \phi_{A_i}$$
$$m_2(B_j) = \Sigma_{k'} \mu_2(Y_{jk'}) \ \ when \ \ \pi_2[Y_{jk'}] = \psi_{B_j}$$

those pairs $(X_{ik}, Y_{jk'})$ are in $\mathcal{W}_0$.

We have

$$m_1(A_i)m_2(B_j) = (\Sigma_k \mu_1(X_{ik}))(\Sigma_{k'} \mu_2(Y_{jk'})) = \Sigma_k \Sigma_{k'} \mu_1(X_{ik}) \mu_2(Y_{jk'})$$

That is $m_1(A_i)m_2(B_j)$ is a part of $wp(\mathcal{W}_0)$. So $wp(\mathcal{W}_0) \geq k$.

Step II

The other way around, if $\phi_{A_i} \wedge \psi_{B_j} =\perp$, then $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}) \subseteq \mathcal{W}_0$ and $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ will be a part of $wp(\mathcal{W}_0)$.

For any $X_k \in i_1(\phi_{A_i})$ and $Y_{k'} \in i_2(\psi_{B_j})$, there must exist ϕ_{A_k} and $\psi_{B_{k'}}$ which make the following equations hold:

$$\pi_1[X_k] = \phi_{A_k}; \ \ \pi_2[Y_{k'}] = \psi_{B_{k'}}$$

and

$$\phi_{A_k} \rightarrow \phi_{A_i}; \ \ \psi_{B_{k'}} \rightarrow \psi_{B_j}$$

so

$$\phi_{A_k} \wedge \psi_{B_{k'}} =\perp; \ \ A_k \cap B_{k'} = \{\}$$

We can divide $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ into different groups under the condition that if $(X_{k1}, Y_{k'1})$ and $(X_{k2}, Y_{k'2})$ are in the same group, then $\pi_1[X_{k1}] = \pi_1[X_{k2}]$ and $\pi_2[Y_{k'1}] = \pi_2[Y_{k'2}]$. Because

of the disjoint feature of any basis, every pair (X_i, Y_j) in $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ must belong to one and only one group. For each group t, the probability weight of the elements in the group is
$\Sigma_l \mu_1(X_{kl})\mu_2(Y_{k'l}) = (\Sigma_l \mu_1(X_{kl}))(\Sigma_l \mu_2(Y_{k'l}))$
$= m_1(A_k)m_2(B_{k'})\ \ (\pi_1[X_{kl}] = \phi_{A_{kl}} \wedge \pi_2[Y_{k'l}] = \psi_{B_{k'l}} \wedge (\phi_{A_{kl}} \wedge \psi_{B_{k'l}}) = \perp)$

For all groups, we have

$\Sigma_t \Sigma_l \mu_1(X_{kl})\mu_2(Y_{k'l})$
$= \Sigma_t m_1(A_{kt})m_2(B_{k't})\ \ (\pi_1[X_{kt}] = \phi_{A_{kt}} \wedge \pi_2[Y_{k't}] = \psi_{B_{k't}} \wedge (\phi_{A_{kt}} \wedge \psi_{B_{k't}}) = \perp)$
$= \Sigma_t m_1(A_{kt})m_2(B_{k't})(A_{kt} \cap B_{k't} = \{\})$
As $\Sigma_t m_1(A_{kt})m_2(B_{k't})(A_{kt} \cap B_{k't} = \{\})$ is a part of k, So $wp(i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}))$ is a part of k. That is $k \geq wp(\mathcal{W}_0)$.
To summarize steps I and II, we have the conclusion that $k = wp(\mathcal{W}_0)$.

Part Two

For any subset C of $2^{\mathcal{A}t}$, and its corresponding formula φ_C, we need to prove that $Bel(C) = wp(i_*(\varphi_C))$.

Step I

Suppose $m = m_1 \oplus m_2$, if $A_i \cap B_j \subseteq C$, then $m_1(A_i)m_2(B_j)$ is a part of $Bel(C)$.
Because of $A_i \leftrightarrow \phi_{A_i}$ and $B_j \leftrightarrow \psi_{B_j}$, we have $\phi_{A_i} \wedge \psi_{B_j} \rightarrow \varphi_C = T$.
According to the definition of the Combination Rule, we have $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}) \subseteq i_*(\varphi_C)$.
We further have

$$m_1(A_i) = \Sigma_k \mu_1(X_{ik})\ \ when\ \ \pi_1[X_{ik}] = \phi_{A_i}$$
$$m_2(B_j) = \Sigma_{k'} \mu_2(Y_{jk'})\ \ when\ \ \pi_2[Y_{jk'}] = \psi_{B_j}$$

those pairs $(X_{ik}, Y_{jk'})$ are in $i_*(\varphi_C)$.
We have

$$m_1(A_i)m_2(B_j) = (\Sigma_k \mu_1(X_{ik}))(\Sigma_{k'} \mu_2(Y_{jk'})) = \Sigma_k \Sigma_{k'} \mu_1(X_{ik})\mu_2(Y_{jk'})$$

That is $m_1(A_i)m_2(B_j)$ is a part of $wp(i_*(\varphi_C))$. So $wp(i_*(\varphi_C)) \geq Bel(C)$.

Step II

The other way around, if $\phi_{A_i} \wedge \psi_{B_j} \rightarrow \varphi_C = T$, then $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}) \subseteq i_*(\varphi_C)$ and $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ is a part of $wp(i_*(\varphi_C))$.

For any $X_k \in i_1(\phi_{A_i})$ and $Y_{k'} \in i_2(\psi_{B_j})$, there must exist ϕ_{A_k} and $\psi_{B_{k'}}$ which make the following equations hold:

$$\pi_1[X_k] = \phi_{A_k};\ \ \pi_2[Y_{k'}] = \psi_{B_{k'}}$$

and

$$\phi_{A_k} \rightarrow \phi_{A_i};\ \ \psi_{B_{k'}} \rightarrow \psi_{B_j}$$

$$\phi_{A_k} \wedge \psi_{B_{k'}} \rightarrow \varphi_C = T;\ \ A_k \cap B_{k'} \subseteq C$$

We can divide $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ into different groups under the condition that if $(X_{k1}, Y_{k'1})$ and $(X_{k2}, Y_{k'2})$ are in the same group, then $\pi_1[X_{k1}] = \pi_1[X_{k2}]$ and $\pi_2[Y_{k'1}] = \pi_2[Y_{k'2}]$. Because of the disjoint feature of any basis, every pair (X_i, Y_j) in $i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j})$ must belong to one and only one group. For each group t, the probability weight of the elements in the group is

$\Sigma_l \mu_1(X_{kl})\mu_2(Y_{k'l}) = (\Sigma_l \mu_1(X_{kl}))(\Sigma_l \mu_2(Y_{k'l}))$
$= m_1(A_k)m_2(B_{k'})$ $(\pi_1[X_{kl}] = \phi_{A_{kl}} \wedge \pi_2[Y_{k'l}] = \psi_{B_{k'l}} \wedge (\phi_{A_{kl}} \wedge \psi_{B_{k'l}} \rightarrow \varphi_C = T)$

For all groups, we have

$\Sigma_t \Sigma_l \mu_1(X_{kl})\mu_2(Y_{k'l})$
$= \Sigma_t m_1(A_{kt})m_2(B_{k't})$ $(\pi_1[X_{kt}] = \phi_{A_{kt}} \wedge \pi_2[Y_{k't}] = \psi_{B_{k't}} \wedge (\phi_{A_{kt}} \wedge \psi_{B_{k't}}) \rightarrow \varphi_C = T)$
$=\Sigma_t m_1(A_{kt})m_2(B_{k't})(A_{kt} \cap B_{k't} \subseteq C)$

Because $\Sigma_t m_1(A_{kt})m_2(B_{k't})$ is a part of $Bel(C)$, so $wp(i_1(\phi_{A_i}) \otimes i_2(\psi_{B_j}))$ is a part of $Bel(C)$ and $Bel(C) \geq wp(i_*(\varphi_C))$.

From steps I and II we get $wp(i_*(\phi_A)) = Bel(A)$.

END

Now it has been proved that what we can combine using Dempster's combination rule can also be combined in incidence calculus and they obtain the same result. Moreover in the next subsection we are going to show that we can handle a wider range of information in incidence calculus by applying the new Combination Rule.

6.3 Comparison III: Combining Dependent Evidence

As we discussed before, when several pieces of evidence are not DS-independent, both Dempster's combination framework and Dempster's combination rule fail to cope with them. However the alternative combination rule we have proposed in this paper is able to deal with this situation as we will see in the next section through examples. In the following we explore the theoretical difference between the two theories and argue why DS theory fails to deal with dependent evidence while incidence calculus succeeds.

Suppose we have two pieces of evidence which are given in the form of probability spaces (S_1, χ_1, μ_1) and (S_2, χ_2, μ_2), and they are not DS-independent. Even though each of them provides a belief function on a proposition set P separately (assume we know the relations between the two probability spaces and P), the two belief functions should not be combined using Dempster's combination rule, otherwise a wrong result will be obtained. Further suppose the two pieces of evidence are given by two observations X and Y, then the observing objects of X and Y must be related to each other or they are the same because of the assumption that the two pieces of evidence are not DS-independent. Let (S, χ, μ) be the common probability space producing (S_1, χ_1, μ_1) and (S_2, χ_2, μ_2) through X and Y, then S_1 and S_2 are all connected with S. Let $S_1 \cap S_2 = S_{12}$, then $S_1 = S_{11} \cup S_{12}$ and $S_2 = S_{21} \cup S_{12}$. When using DS theory to describe the effect of the two pieces of evidence on P, we have to describe them in a numerical way, *e.g.* using either m functions or bel functions. Assume that we get two mass functions $m_{(S_{11} \cup S_{12})}$ and $m_{(S_{21} \cup S_{12})}$ where $m_{(S_{11} \cup S_{12})}$ means this mass function is obtained based on the information carried by the subset $S_{11} \cup S_{12}$ of S. If we combine them in DS theory as $m_{(S_{11} \cup S_{12})} \otimes m_{(S_{11} \cup S_{12})}$, then the information carried by the subset S_{12} will be counted twice. So DS theory has no ability to combine such overlapped (or dependent) evidence.

However in incidence calculus, instead of describing the effect of evidence in a numerical way, it shows the effect of evidence by establishing the relations between the subset $S_{11} \cup S_{12}$ to

P and the subset $S_{21} \cup S_{12}$ to P separately. During the combination, the biggest information subset used in this procedure is $(S_{11} \cup S_{12}) \cup (S_{21} \cup S_{12}) = S_{11} \cup S_{12} \cup S_{21}$ where the information carried by S_{12} is used only once. For any formula in $\mathcal{L}(P)$, once its incidence set (or the lower and upper bound of its incidences) is known, its probability can be calculated from the incidence set.

In the above discussion we assume that the intersection of S_1 and S_2 are not empty. In fact, as long as S_1 and S_2 are constructed from a common space (which is not the set product of S_1 and S_2), the mass functions produced from them should not be combined in DS theory no matter whether their intersection is empty or not. This will be seen in the next section in the second part of example 1.

Therefore the indirect encoding of probabilities of formulae in incidence calculus makes it possible to combine the overlapped information which is superior to any other numerical approaches to managing uncertainty.

7 Analysing Examples

In this section we are going to explore five examples which cannot be combined by using Dempster's combination rule. These examples can be easily dealt with in incidence calculus.

7.1 Example 1

In Pearl's paper [Pearl 90], the original example is stated as

$r_1 : I(x) \rightarrow Po(x)$; if a person is intelligent, then that person is popular.
$r_2 : F(x) \rightarrow \neg Po(x)$; If a person is fat, then that person is unpopular.

It is also assumed that each of the two rules has a strength m. When one learns that 'Joe is fat', DS theory produces the result that 'Joe is believed to be not intelligent with m^2'. Pearl argued that it is more reasonable to believe (with degree m) that 'Joe is unpopular' rather 'Joe is believed to be not intelligent with m^2'.

When we use incidence calculus to deal with this example, we treat the two rules r_1, r_2 as non-independent (we don't assume that they are given independently). We prefer that r_1 and r_2 are from a well-defined knowledge system (common sense) in which a set of possible worlds is implicitly used to support the rules or the rules are got from a statistical result of a large population. Purely from these two rules we can define an incidence calculus theory as:

$$< W, \varrho, P, \mathcal{A}, i >$$

where $P = \{I, Po, F, ...\}$ and $\mathcal{A} = \{r_1, r_2, r_1 \wedge r_2, T\}$. $i(r_1) = W_1, wp(W_1) = m$, $i(r_2) = W_2, wp(W_2) = m$, and $i(r_1 \wedge r_2) = W_1 \cap W_2$.

When one knows that 'Joe is fat', another incidence calculus is formed as $< \{Joe\}, 1, P, \{F\}, i(F) = \{Joe\} >$. Combining these two incidence calculus theories using the following table, the result is that 'Joe is unpopular with degree m' which is more intuitive.

ϕ	r_1	r_2	$r_1 \wedge r_2$	T
$i(\phi)$	W_1	W_2	$W_1 \cap W_2$	W
F	$F \wedge r_1$	$F \wedge r_2$	$F \wedge r_1 \wedge r_2$	F
$\{Joe\}$	$W_1 \otimes \{Joe\}$	$W_2 \otimes \{Joe\}$	$(W_1 \cap W_2) \otimes \{Joe\}$	$W \otimes \{Joe\}$

Table 2. Combining two incidence calculus theories

So $i_*(\neg Po) = W_2 \otimes \{Joe\} \cup (W_1 \cap W_2) \otimes \{Joe\}$. Therefore $p(\neg Po) = wp(i_*(\neg Po)) = wp(W_2 \otimes \{Joe\}) = wp(W_2) \times wp(\{Joe\}) = m$, that is, 'Joe is unpopular with degree of belief m.

7.2 Example 2

This example is also adopted from [Pearl 90]. Suppose we are given the following two rules:

If A then B with certainty 0.9;
If $\neg A$ then B with certainty 0.7.

If we encode the messages carried by the two rules in terms of belief functions in DS theory, the result will be $bel(B) = 0.63$. Pearl argued that "common sense dictates that even if we do not have any information about A we should still believe in B to a degree at least 0.7". Dealing with this example in incidence calculus theory is somehow similar to the previous example. First of all, an incidence calculus theory based on the two rules is constructed; secondly, another incidence calculus theory is created regarding A. The combination of these two theories tells us that $p(B) = 0.7 + 0.2 \times p'(A)$ where $p'(A)$ is the certainty of A and $0 \leq p'(A) \leq 1$. So $0.7 \leq p(B) \leq 0.9$.

By contrast to DS theory, in incidence calculus we don't make the assumption that several rules are distinct, rather we consider them related to another set of events which is called the set of possible worlds. The relation between the rules and the set of possible worlds is stated in terms of incidence calculus theories. Dempster's rule has no ability to cope with the cases when a set of rules are relevant.

7.3 Example 3

In this example we continue to analyse 'Penguin – Bird – Fly' event as we introduced in CASE 2 in Section 5.

From the evidence that Tweety is a bird, an incidence calculus theory describing this piece of information can be obtained as:

$$< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1 >$$

where $\mathcal{W}_1 = \{Tweety\}, \varrho_1(Tweety) = 1, \mathcal{A}_1 = \{Bi\}, i_1(Bi) = \{Tweety\}$ and $P = \{Bi, Pe, Fl\}$.

When one later learns that Tweety is a Penguin, another incidence calculus theory can be as

$$< \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2 >$$

where $\mathcal{W}_2 = \{Tweety\}, \varrho_2(Tweety) = 1, \mathcal{A}_2 = \{Po\}, i_2(Po) = \{Tweety\}$.

Combining these two theories (they are based on the same set of possible worlds) using the rule we proposed we have the third incidence calculus theory which is exactly the same as the second one.

$$< \mathcal{W}_3, \varrho_3, P, \mathcal{A}_3, i_3 >=< \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2 >$$

The fact that Tweety is a penguin enables us make further inference by using some common sense knowledge – a set of relevant rules. These relevant rules form the fourth incidence calculus theory as

$$< \mathcal{W}_4, \varrho_4, P, \mathcal{A}_4, i_4 >$$

where $r_3 : Po \to Bi$ and $r_4 : Po \to \neg Fl$. $\mathcal{A}_4 = \{r_3, r_4, r_3 \wedge r_4, T\}$, $i_4(r_3) = \mathcal{W}_{4i}, i(r_4) = \mathcal{W}_{4j}$, $i(r_3 \wedge r_4) = \mathcal{W}_{4i} \cap \mathcal{W}_{4j}$, $\varrho(\mathcal{W}_{4i}) = 1$ and $\varrho(\mathcal{W}_{4j}) = .95$. Here $\mathcal{W}_{4k} \subseteq \mathcal{W}_4$ for $k = i, j$.

Using the combination rule to combine theories 3 and 4, the result is $p(Bi) = 1$ and $p(\neg Fl) = .95$ which is correct.

7.4 Example 4

Following the example in case 3. There are 100 balls in an urn which are labelled as shown in Table 1.

$Label$	$Number\ of Balls$	$Subset\ Name\ in\ \mathcal{W}$
axy	4	S_1
ax	4	S_2
ay	16	S_3
a	16	S_4
bxy	10	S_5
bx	10	S_6
by	20	S_7
b	20	S_8

Table 3. 100 balls and their labels

As we assumed that observation X (or Y) denotes that a particular ball drawn from the urn has label x (or y), two pieces of evidence are obtained in the form of probability spaces as:

$$(X_1, X_1, \mu_1) \quad and \quad (X_2, X_2, \mu_2)$$

where

$$X_1 = S_1 \cup S_2 \cup S_5 \cup S_6$$

$$\mu_1(S_1) = 1/7 \qquad \mu_1(S_2) = 1/7$$

$$\mu_1(S_5) = 5/14 \qquad \mu_1(S_6) = 5/14$$

and

$$X_2 = S_1 \cup S_3 \cup S_5 \cup S_7$$

$$\mu_2(S_1) = 2/25 \qquad \mu_2(S_3) = 8/25$$

$$\mu_2(S_5) = 1/5 \qquad \mu_2(S_7) = 2/5$$

It is easy to see these two probability spaces are not DS-independent and they cannot be combined using either Dempster's combination framework or Dempster's combination rule. The intersection of X_1 and X_2 is $S_1 \cup S_5$.

Let us examine the example of the labelled balls in incidence calculus theory and see what we can get.

First of all, we suppose that the set of possible worlds $\mathcal{W}$ contains 100 labelled balls.

$$\mathcal{W} = S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5 \cup S_6 \cup S_7 \cup S_8$$

where S_1 contains 4 possible worlds each of which specifies that a labelled ball may be chosen later, ..., S_8 contains 20 possible worlds and the probability distribution on $\mathcal{W}$ is $\varrho(w) = 1/100$ for any $w \in \mathcal{W}$. We further suppose the set of propositions P contains $\{a, b, x, y\}$ where a means that the chosen ball has label a et al. If observations X and Y state that a particular ball drawn from the urn has label x or y respectively, then we can construct two mapping relations between $\mathcal{W}$ and P in terms of incidence functions separately.

$$i_1(x) = S_1 \cup S_2 \cup S_5 \cup S_6$$

$$i_1(a \wedge x) = S_1 \cup S_2, \quad i_1(b \wedge x) = S_5 \cup S_6$$

$$i_2(y) = S_1 \cup S_3 \cup S_5 \cup S_7$$

$$i_2(a \wedge y) = S_1 \cup S_3, \quad i_2(b \wedge y) = S_5 \cup S_7$$

Thus two incidence calculus theories can be formed as $< \mathcal{W}, \varrho, P, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}, \varrho, P, \mathcal{A}_2, i_2 >$, where $\mathcal{A}_1 = \{x, a \wedge x, b \wedge x\}$ and $\mathcal{A}_2 = \{y, a \wedge y, b \wedge y\}$.

Applying the Combination Rule to these two theories, we can get the third incidence calculus theory $< \mathcal{W}, \varrho, P, \mathcal{A}, i >$ where

$\mathcal{A} = \{x \wedge y, a \wedge x \wedge y, b \wedge x \wedge y, a \wedge b \wedge x \wedge y\}$

and

$$i(b \wedge x \wedge y) = S_5$$

$$i(x \wedge y) = S_1 \cup S_5$$

That is, $p(b \wedge x \wedge y) = 10/100$ and $p(x \wedge y) = 14/100$. According to Equation(12) in Section 2, we have

$$p(b \mid x \wedge y) = \frac{p(b \wedge x \wedge y)}{p(x \wedge y)} = \frac{wp(i(b \wedge x \wedge y))}{wp(i(x \wedge y))} = 5/7$$

Obviously, this result is consistent with what we could get in probability theory as shown in Voorbraak's paper. The advantage of calculating the conditional probability based on the incidences of the related formulas in incidence calculus theory makes it possible to be consistent with probability theory.

Next, we observe the labels of the 100 balls from another perspective. Instead of having the frame $\Theta = \{a, b\}$, we assume another frame $\Theta_1 = \{x \wedge y, x \wedge \neg y, \neg x \wedge y, \neg x \wedge \neg y\}$. When drawing a ball from the urn, its multiple labels only make one element of Θ true. Further suppose that we have two observations A and B where

A: denotes that a particular ball drawn from the urn has label a;
B: denotes that a particular ball drawn from the urn has label b.

Once again these two observations define two probability spaces denoting two pieces of evidence, and the two pieces of evidence give two mass functions on frame Θ_1 as:

$$m_A(x \wedge y) = 1/10 \qquad m_A(x \wedge \neg y) = 1/10$$
$$m_A(\neg x \wedge y) = 2/5 \qquad m_A(\neg x \wedge \neg y) = 2/5$$
$$m_B(x \wedge y) = 1/6 \qquad m_B(x \wedge \neg y) = 1/6$$
$$m_B(\neg x \wedge y) = 2/6 \qquad m_B(\neg x \wedge \neg y) = 2/6$$

Combining them using Dempster's combination rule we have $m_{A \wedge B}(x \wedge y) = 1/20$. That is, the probability that a ball has both label x and y is $1/20$ (in this case the probability of a ball is exactly the same as the degree of belief in the ball) based on observations A and B. In fact, it is impossible to draw a ball which has both label a and b. So in probability we have $p(x \wedge y \wedge a \wedge b) = 0$ or we have $p(x \wedge y \mid a \wedge b) = \infty$ if we consider the conditional probability because of $p(a \wedge b) = 0$.

When we use incidence calculus to deal with the messages carried by the two observations, two incidence calculus theories are formed first. Applying the rule we defined in this paper to the two incidence calculus theories, we can combine them and get the third incidence calculus theory as[11]:

$$< \mathcal{W}, \varrho, P, \mathcal{A}, i >$$

where $P = \mathcal{A} = \Theta$ and for any formula ϕ in $\mathcal{A}$, we have $i(\phi) = \{\}$. So we have the same result as what we have obtained in probability theory. As we have explained in section 4, in such a situation, the two observations cannot be held at the same time. They repel each other. Whatever the relations between two observations (or two pieces of evidence) are, incidence calculus can be always used to deal with them and give the correct result while DS theory has no ability to deal with them no matter in which way we explain the condition of using Dempster's combination rule.

7.5 Example 5

What we will show next is that the Combination Rule can also be used to deal with the partial implication problems. This example is also from Voorbraak's paper. The original example states that:

Let Θ be the frame $\{A, \neg A\}$, X and Y denote two observations, where

A denotes the proposition "patient M has flu",
X represents the observation that M has a fever $\geq 39°C$ and
Y represents the observation that M has a fever $\geq 38.5°C$.

[11] We leave the details of the example to the reader.

Then two mass functions are established on frame Θ based on these two observations.

$$m_X(A) = 0.6, \quad m_X(\Theta) = 0.4$$

$$m_Y(A) = 0.4, \quad m_Y(\Theta) = 0.6$$

Using Dempster's combination rule to combine them, we get $Bel(A) = 0.76$. But Voorbraak argued that because X implies Y, the effect of observation Y should be merged by X during the combination, and the correct result should be $Bel(A) = 0.6$.

We explore this example again in incidence calculus. In fact we can assume that there is a set of possible worlds $\mathcal{W}$, the subset $\mathcal{W}_1$ of $\mathcal{W}$ makes proposition X_M="M has a fever $\geq 39°C$" true. Obviously, $\mathcal{W}_1$ will make proposition Y_M="M has a fever $\geq 38.5°C$" true as well. Then there will be two incidence calculus theories which contain $i_1(X_M) = \mathcal{W}_1$ and $i_2(Y_M) = \mathcal{W}_1 \cup \mathcal{W}_2$ respectively. Therefore the combination of the two theories will cause the result $i(X_M) = \mathcal{W}_1$ to be included in the new incidence calculus theory. If we accept the description that if proposition X_M is true, then the patient M has flu with probability 0.6, then we eventually have $P(A) = 0.6$. The concrete analysis of the example is given below. In this example we use two production rules $B \rightarrow A(r)$ and $C \rightarrow B(r_1(C))$, where the first rule states that if a patient M has a fever then M has flu with probability r. Similarly the second rule says that if the patient M has a body temperature C then M has a fever with probability $r_1(C)$. These two rules are hidden in Voorbraak's example.

This example is more complicated than the previous one, because in this example we need to construct three sets of possible worlds.

Suppose a proposition set P has three propositions A, B, and C where A denotes that a person has flu, B denotes that a person has a fever and C denotes the temperature a person has. Then rule $B \rightarrow A$ and $C \rightarrow B$ (here we temporarily ignore the rule strengths and we will associate the strengths later using incidence functions) are in a language set $\mathcal{L}(P)$.

Assume the first set of possible worlds is $\mathcal{W}_1$ and its subset w_1 makes rule $B \rightarrow A$ true, and $p(w_1) = r$. Then we have an incidence calculus theory as :

$$< \mathcal{W}_1, \varrho_1, P, \mathcal{A}_1, i_1)$$

where $\mathcal{A}_1 = \{B \rightarrow A\}$ and $i_1(B \rightarrow A) = w_1$.

Assume the second set of possible worlds is $\mathcal{W}_2$ and its subset w_2 makes rule $C \rightarrow B$ true, and $p(w_2) = r_1(C)$. Then we have the second incidence calculus theory as:

$$< \mathcal{W}_2, \varrho_2, P, \mathcal{A}_2, i_2)$$

where $\mathcal{A}_2 = \{C \rightarrow B\}$ and $i_2(C \rightarrow B) = w_2$.

Further suppose the third set of possible worlds is $\mathcal{W}_3$ and

$$\mathcal{W}_3 = \{t_1, 35, ..., 38.5, 39, ..., 50, t_2\}$$

Here we put the temperature below $35°C$ and above $55°C$ into two groups t_1, t_2. Observations X and Y specify the third and fourth incidence calculus theories from $\mathcal{W}_3$ to $\mathcal{L}(P)$ as:

$$< \mathcal{W}_3, \varrho, P, \mathcal{A}_3, i_3 >$$

$$< \mathcal{W}_3, \varrho, P, \mathcal{A}_4, i_4 >$$

where $\mathcal{A}_3 = \{C\}$, $\mathcal{A}_4 = \{C\}$, $i_3(\mathcal{A}_3)=\{39, ..., 50, t_2\}$, $i_4(\mathcal{A}_4) = i_3(\mathcal{A}_3) \cup \{38.5\}$.

In fact, the semantic meaning of this example can be shown as:

$$\begin{tikzcd} X \arrow[dr] & & & \\ & C \arrow[r] & B \arrow[r] & A \\ Y \arrow[ur] & & & \end{tikzcd}$$

This figure shows that from a person has high body temperature we can infer that this person has a fever, and from this person has a fever we can further infer that this person may have flu. Two observations X and Y give two possible temperature values of the person.

Applying the Combination Rule to the last two theories, we can get a combined incidence calculus theory as:

$$< \mathcal{W}_3, \varrho, P, \mathcal{A}_5, i_5 >$$

where $\mathcal{A}_5$=$\mathcal{A}_3$, and $i_5 = i_3$.

Then applying the Combination Rule to incidence calculus theories 1, 2, and 5, we eventually have an incidence calculus theory as:

$$< \mathcal{W}, \varrho, P, \mathcal{A}_6, i >$$

where $\mathcal{W} = \mathcal{W}_1 \otimes \mathcal{W}_2 \otimes \mathcal{W}_3$, and $p(A) = r_1(C) \times r$. In the sense of Voorbraak, when $C \geq 39°C$, we have $p(A) = 0.6$. As proved before because $p(A) = wp(i(A)) = wp(i_*(A)) = Bel(A)$ we can get the same result as Voorbraak in his paper which is different from what we can get in DS theory.

Someone may argue about the relation between the sets $\mathcal{W}_2$ and $\mathcal{W}_3$. In fact we can always let $\mathcal{W}_2$ have two elements w_{11} and w_{12}, and let w_{11} make B true. But the probability distribution on $\mathcal{W}_2$ is a function of the temperature that a person has. For example, assume a person's temperature is $37°C$, then a theory based on the information gives that $i(C) = \{37\}$. This result produces the probability distribution on $\mathcal{W}_2$ as $\varrho(w_{11}) = 0$ and $\varrho(w_{12}) = 1$, since we think temperature $37°C$ is reasonable.

This example also shows us that we can associate a rule strength with an incidence calculus theory as we have done here.

The examples in this section show that the alternative combination rule proposed by using incidence calculus is more general than Dempster's combination rule. In particular, this rule can be applied to deal with the overlapped or relevant information.

8 Conclusion

Pooling a joint impact from multiple sources of information is an important and difficult task in the management of uncertainty. In this paper, we have further explored the well known combination rule in DS theory and revealed the failure of the rule in solving some problems. We have argued that the counterintuitive results of using the rule are caused by the misexplanation of the independence requirements of Dempster's combination framework with the combination rule. We can conclude from this exploration that the combination carried out before or after the propagation of evidence may give absolutely different results in many situations. However Dempster's combination rule in DS theory is applicable under the condition that the result of combination is unique no matter whether it is done before or after the propagation of evidence.

Trying to combine dependent pieces of information using Dempster's combination rule has been mentioned in [Dubois and Prade 86, Kennes 91, Nguyen and Smets 92, Smets 90], but their work all focused on Dempster's combination rule without examining Dempster's original combination framework.

In order to overcome this difficulty, we have proposed an alternative combination mechanism using incidence calculus. The comparison with DS theory has shown that this new combination approach is more powerful than Dempster's combination rule, especially in solving those non DS-independent problems. This result has not yet been shown in any other theories in the management of uncertainty.

In general, independent relations among multiple sources of evidence can be considered as the special cases of dependent situations. As Pearl indicated [Pearl 1992, p:435] "If we have several items of evidence, each depending on the state of nature, these items of evidence should also depend on each other. This kind of dependency is not a nuisance but a necessary bliss; no evidential reasoning would otherwise be possible." In our proposed combination rule, we have indeed adopted the same idea and made some efforts towards combining dependent evidence. Even though we cannot promise that the proposed rule in this paper is perfect to cope with all cases, at least we have revealed the approach to combine them. This result will be useful for further research work either on this topic or in the relevant aspects. It tells us that cancelling the overlapped and duplicated information at the semantic level is a promising way to obtain the result of several pieces of evidence.

The future work will be concerned with applying these combination rules to practical problems. The topics of using incidence calculus to represent default logic and to implement ATMS are also interesting. Producing explanations after a system inference is a side product of a knowledge based system. Usually numerical methods for dealing with uncertainty are weak in giving explanations for the results after several steps of combination or fusion. It is worthwhile to explore this topic in incidence calculus by tracing the incidence sets of formulas.

Acknowledgement

The authors are grateful for useful comments from Paul Brna, Dave Robertson, Flavio S Correa da Silva and Jun Hong and in particular, comments from Prof. Smets and Prof. Pearl.

The first author is a PhD student at the Univ. of Edinburgh, and is being supported by a "Colin and Ethel Gordon Scholarship" of Edinburgh University and an ORS award. The second author is supported by an SERC senior fellowship.

References

[1] Black,P.(1987) Is Shafer general Bayes? in: Proc. of Third AAAI Uncertainty in Artificial Intelligence workshop, 2-9.

[2] Bundy,A. (1985) Incidence calculus: a mechanism for probability reasoning. J. of Automated Reasoning 1, pp.263-283.

[3] Bundy,A. (1986) Correctness Criteria of some algorithms for uncertain reasoning using incidence calculus. J. of Automated Reasoning 2. pp.109-126.

[4] Bundy,A. (1992) Incidence Calculus. The Encyclopedia of AI, pp. 663-668, 1992. It is also available as the Research paper No. 497 in the Dept. of Artificial Intelligence, Univ. of Edinburgh.

[5] Correa da Silva,F. and A.Bundy (1990) On some equivalence relations between incidence calculus and Dempster-Shafer theory of evidence. Proc. of sixth workshop of Uncertainty in Artificial Intelligence, pp.378-383.

[6] Dempster,A.P. (1967) Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. 38, pp.325-339.

[7] Dubois,D. and H.Prade (1986) On the unicity of Dempster rule of combination . I.J. of Intelligence Systems Vol.1:133-142.

[8] Fagin,R. and J. Halpern (1989a) Uncertainty, belief and probability. IJCAI-89, pp.1161-1167.

[9] Fagin,R. and J. Halpern (1989b) Uncertainty, belief and probability. Research Report of IBM, RJ 6191.

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

[11] Hunter,D. (1987) Dempster-Shafer vs. probabilistic logic. in: proc. Third AAAI uncertainty in artificial intelligence workshop, 22-29.

[12] Kennes,R. (1991) Evidential reasoning in a categorical perspective: Conjunction and disjunction of belief functions. in Proc. of 7th uncertainty in artificial intelligence, 174-181.

[13] Lemmer,J.F. (1986) Confidence factors, empiricism and Dempster-Shafer theory of evidence. Uncertainty in Artificial Intelligence 1986, pp.117-125.

[14] Liu,W. and A.Bundy (1992) Recovering the probability space from the lower bound of probabilities on a set of axioms. Forthcoming.

[15] Nguyen,H.T. and Ph. Smets, (1992) On Dynamics of Cautious Belief and Conditional Objects. Technical Report No.TR/IRIDIA/91-13.2. To appear in I.J. of Approx. Reasoning.

[16] Pearl,J. (1988) Probabilistic Reasoning in Intelligence Systems: networks of plausible inference. Morgan Kaufmann Publishers, Inc.

[17] Pearl,J. (1990) Reasoning with belief functions: An analysis of compatibility. International J. of Approximate Reasoning 4, pp.363-389.

[18] Pearl,J. (1992) Rejoinder to Comments on "Reasoning with belief function: An analysis of compatibility". I.J.of Approx. Reasoning 6:425-443.

[19] Ruspini,E.H., Lowrance,J.D. and Strat,T.M. (1990) Understanding Evidential Reasoning. SRI International Technical Note 501.

[20] Shafer,G. (1976) A mathematical theory of evidence. Princeton University Press.

[21] Shafer,G. (1982) Belief functions and parametric models. J. R. Stat. Soc. Ser. B 44: 322-352.

[22] Shafer,G. (1990) Perspectives on the theory and practice of belief functions. Working paper No.218. Univ. of Kansas, School of Business.

[23] Shafer,G. (1987) Belief function and possibility measures. In: J.C.Bezdek, ed., Analysis of Fussy Information 1, CRC Press.

[24] Shafer,G. and A.Tversky (1985) Languages and Designs for Probability Judgement. Cognitive Science 9, p:309-339.

[25] Smets,P. (1988) Belief functions. In: Smets, Mamdani, Dubois and Prade eds., Non-Standard Logics for Automated Reasoning: 213-252.

[26] Smets,Ph. (1990) The Combination of Evidence in the Transferable Belief Model. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.12, No.5.

[27] Smets,P. and Y.T.Hsia (1990) Default Reasoning and the Transferable Belief Model. Proc. of Uncertainty in Artificial Intelligence p:529-537.

[28] Voorbraak,F. (1991) On the justification of Dempster's rule of combination. Artificial Intelligence 48, pp.171-197.

[29] Zadeh,L.A. (1984) A mathematical theory of evidence (book review). AI Mag 5 (3): 81-83.

[30] Zadeh,L.A. (1986) A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. AI Mag. 7, pp.85-90.