A Comprehensive Comparison between Generalized Incidence Calculus and the Dempster-Shafer Theory of Evidence

Citation for published version:
Liu, W & Bundy, A 1994, 'A Comprehensive Comparison between Generalized Incidence Calculus and the Dempster-Shafer Theory of Evidence', *International Journal of Human-Computer Studies*, vol. 40, no. 6. https://doi.org/10.1006/ijhc.1994.1046

Digital Object Identifier (DOI):
10.1006/ijhc.1994.1046

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
International Journal of Human-Computer Studies

Title Page

A comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence

Weiru Liu Alan Bundy

80 South Bridge

Dept. of Artificial Intelligence

University of Edinburgh

Edinburgh EH1 1HN, U.K.

March 1994

Address correspondence to Weiru Liu or Alan Bundy, Dept. of Artificial Intelligence, Univ. of Edinburgh, Edinburgh EH1 1HN, UK.

E-mail:weiru@aisb.ed.ac.uk bundy@aisb.ed.ac.uk

A comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence

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

Abstract

Dealing with uncertainty problems in intelligent systems has attracted a lot of attention in the AI community. Quite a few techniques have been proposed. Among them, the Dempster-Shafer theory of evidence (DS theory) has been widely appreciated. In DS theory, Dempster's combination rule plays a major role. However, it has been pointed out that the application domains of the rule are rather limited and the application of the theory sometimes gives unexpected results. We have previously explored the problem with Dempster's combination rule and proposed an alternative combination mechanism in generalized incidence calculus. In this paper we give a comprehensive comparison between generalized incidence calculus and the Dempster-Shafer theory of evidence. We first prove that these two theories have the same ability in representing evidence and combining DS-independent evidence. We then show that the new approach can deal with some dependent situation while Dempster's combination rule cannot. Various examples in the paper show the ways of using generalized incidence calculus in expert systems.

1 Introduction

The management of uncertainty within knowledge and evidence involves three main tasks: representing, propagating and combining 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 critically discussed by several researchers (Halpern and Fagin (1992), Hunter (1987), Lemmer (1986), Pearl (1988, 1990), Voorbraak (1991), Zadeh (1984, 1986)). They showed that in some situations Dempster's rule gives counterintuitive results. The other people (Shafer (1982, 1990), Smets (1988), Ruspini *et al* (1990)) disagreed and argued that the counterintuitive results are caused by the misapplication of the rule. In Liu (1994), we considered this problem from a very different perspective. We argued that the counterintuitive results are caused by failing to show explicitly the condition of combination defined by Dempster in his original paper (Dempster (1967)).

In this paper, we present a theory for reasoning under uncertainty, generalized incidence calculus, which can be taken as an alternative of the DS theory. Incidence calculus is a probabilistic logic introduced by Bundy (1985). This theory is proved to be equivalent to DS theory in representing evidence in some cases (Correa de Silva and Bundy (1990)). In Liu (1994), the original incidence calculus has been generalized to model a wider range of evidence and an alternative combination rule is proposed to combine multiple pieces of evidence. In this paper, we discuss the relations between generalized incidence calculus and DS theory. This comparison shows that these two theories are equivalent in representing evidence and combining distinct pieces of evidence (or DS-independent pieces of evidence in Voorbraak (1991)) but not equivalent in combining dependent evidence. Therefore, generalized incidence calculus is possible to be applied to deal with those cases which DS theory fails to cope with.

The paper is organized as follows. In section 2, we first introduce the basics of incidence calculus, generalized incidence calculus and the main features of the theory. We then introduce the new combination rule in generalized incidence calculus. Section 3 describes the Dempster-Shafer theory of evidence. Section 4 is about the comparison between DS theory and generalized incidence calculus. We show that they have the same ability to represent evidence (information). We also prove that Dempster's combination rule is subsumed by the new combination rule. Several examples are given to demonstrate the features of these two theories. Finally, in section 5, we conclude the paper.

2 Incidence Calculus and Generalized Incidence Calculus

Incidence calculus (Bundy (1985), (1992)) is a probabilistic logic for reasoning under uncertainty. In contrast to other numerical approaches, in incidence calculus probabilities are associated with a set of possible worlds rather than formulae directly. The probability of a formula is calculated through the incidence set assigned to the formula.

2.1 Propositional Language

The language we use 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 and closed under the operations of negation ($\neg$), disjunction ($\vee$), conjunction ($\wedge$) and implication ($\rightarrow$).*

Suppose that a proposition set P contains $q_1, q_2, ..., q_n$. $\mathcal{A}t$ is the set of basic elements, each of which is in the form $q'_1 \wedge ... \wedge q'_n$, where q'_j is either q_j or $\neg q_j$ and $q_j \in P$. Any formula φ in the language set $\mathcal{L}(P)$ can be represented as

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

$\delta_j = q'_1 \wedge ... \wedge q'_n$ is called a *basic element* and $\mathcal{A}t$ is called the *basic element set*.

2.2 Incidence Calculus

Definition 2 Possible Worlds: *Each possible world is a primitive object of incidence calculus and can be thought of as a partial interpretation of some logical formulae.*

The probability is represented by a function μ from the set of possible worlds $\mathcal{W}$ to real numbers between 0 and 1. For each possible world $w \in \mathcal{W}$, $\mu(w)$ is known and the probability weight of the whole set $\mathcal{W}$ is 1, that is $\mu(\mathcal{W}) = 1$.

If I is a subset of $\mathcal{W}$, the probability of I is defined to be:

$$\mu(I) = \Sigma_{w \in I} \mu(w) \tag{2}$$

Definition 3 Incidence Calculus Theories: *An incidence calculus theory is a quintuple* $< \mathcal{W}, \mu, P, \mathcal{A}, i >$*, where:*

$\mathcal{W}$ is a finite set of possible worlds.
For all $w \in \mathcal{W}$, $\mu(w)$ is the probability of w and $\mu(\mathcal{W}) = 1$.
$\mathcal{A}t$ is the basic element set of P. $\mathcal{L}(P)$ is the language set of P.
$\mathcal{A}$ 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 conditions:*

$$\begin{aligned} i(\phi_1 \vee \phi_2) &= i(\phi_1) \cup i(\phi_2) \\ i(\phi_1 \wedge \phi_2) &= i(\phi_1) \cap i(\phi_2) \\ i(\neg\phi) &= \mathcal{W} \setminus i(\phi) \\ i(false) = \{\} & \qquad i(true) = \mathcal{W} \end{aligned} \tag{3}$$

These conditions guarantee that incidence calculus is truth functional. That is, the incidence set of a formula can be obtained from the incidence sets of its subformulae.

The property $i(\neg\phi) = \mathcal{W} \setminus i(\phi)$ requires that the elements in $\mathcal{W}$ must be distributed into either $i(\phi)$ or $i(\neg\phi)$. If both $i(\phi)$ and $i(\neg\phi)$ are specified respectively, then $i(\phi) \cup i(\neg\phi)$ should be the whole set $\mathcal{W}$.

The property $i(\phi_1 \vee \phi_2) = i(\phi_1) \cup i(\phi_2)$ says that if $i(\phi_1)$, $i(\phi_2)$ and $i(\phi_1 \vee \phi_2)$ are all known, the possible worlds in $i(\phi_1 \vee \phi_2)$ can be split into two groups (not necessarily disjoint) $i(\phi_1)$ and $i(\phi_2)$. $i(\phi_1 \vee \phi_2)$ carries no more information than the union of $i(\phi_1)$ and $i(\phi_2)$.

If a real situation fails to meet either of the above two properties, incidence calculus theories cannot be used to describe it.

2.3 Generalized incidence calculus

In order for incidence calculus to have the ability to model a situation which an original incidence calculus theory is not suitable to represent, as described above, we generalize the original incidence calculus by dropping some of the conditions on it.

A mapping function $i' : \mathcal{A} \to 2^{\mathcal{W}}$ maps each formula ϕ in $\mathcal{A}$ to a subset of $\mathcal{W}$. $\mathcal{W}$ is interpreted as a set consisting of possible answers to a question. We still call $\mathcal{W}$ a set of possible worlds in this paper. $w \in i'(\phi)$ means that if w is the answer to the question carried by $\mathcal{W}$, then formula ϕ is true. We also require that $i'(false) = \{\}$ and $i'(true) = \mathcal{W}$. For a possible world $w \in \mathcal{W}$, if $w \notin i'(\phi)$, it doesn't necessarily mean that $w \in i'(\neg\phi)$. So if both $i'(\phi)$ and $i'(\neg\phi)$ are known, $i'(\phi) \cup i'(\neg\phi)$ may be just a subset of $\mathcal{W}$. This is explained as that the current information is uncommitted as to whether w supports ϕ or $\neg\phi$. This phenomenon is usually called *ignorance*. A mechanism which can model this phenomenon is said to have the ability to represent ignorance.

Moreover, if $i'(\phi)$, $i'(\psi)$ and $i'(\phi \vee \psi)$ are all specified, it is likely that $i'(\phi) \cup i'(\psi) \subset i'(\phi \vee \psi)$ is valid. For instance, suppose that there are ten delegates elected to attend a meeting. The meeting will be held some day next week on which all the delegates are asked to give their preferences. The meeting will be held on the day which is preferred by most of the delegates.

Suppose that delegates 1 to 4, denoted as a_1, ..., a_4, prefer *mon*, delegate 5, a_5, prefers *mon* or *tues*, the rest prefer *tues*. Then a mapping function i' could be defined as

$$i'(q_1) = \{a_1, a_2, a_3, a_4\}$$

$$i'(q_2) = \{a_6, a_7, a_8, a_9, a_{10}\}$$

and

$$i'(q_1 \vee q_2) = \{a_1, ..., a_5, ..., a_{10}\}$$

where q_1 stands for 'The meeting is held on Monday', q_2 for 'The meeting is on Tuesday'. Obviously, we have $i'(q_1) \cup i'(q_2) \subset i'(q_1 \vee q_2)$ because a_5 cannot be put into either $i'(q_1)$ or $i'(q_2)$.

A mapping function i' which has the ability to represent the above two phenomena is called a generalized incidence function.

For any two formulae ϕ, ψ in $\mathcal{A}$, if $i'(\phi), i'(\psi)$ and $i'(\phi \wedge \psi)$ are all known, then it can be proved that $i'(\phi \wedge \psi) = i'(\phi) \cap i'(\psi)$.

If we use $\wedge(\mathcal{A})$ to denote the language set which contains $\mathcal{A}$ and all the possible conjunctions of its elements, then a generalized incidence function can be extended to any formula in this set by defining $i'(\wedge \phi_j) = \cap_j i'(\phi_j)$, if $\wedge_j \phi_j$ is not given initially. Therefore, the domain of i', the set of axioms $\mathcal{A}$, can always be extended to a set which is closed under the operator $\wedge$.

Thus, whenever we have a set of axioms $\mathcal{A}$ on which a generalized incidence function i' is defined, this set of axioms can always be extended to another set which is closed under the operator $\wedge$. In the following, we assume that the set of axioms $\mathcal{A}$ is already extended and is closed under $\wedge$.

In particular, if $i'(\wedge_j \phi_j) = \{\}$, it doesn't matter whether this formula is in $\wedge(\mathcal{A})$ as this formula has no effect on further inferences. However if $\wedge_j \phi_j = \perp$, then $i'(\wedge_j \phi_j) = \cap_j i'(\phi_j)$ must be empty; otherwise the information for constructing the function i' is contradictory.

In the following, we use i to stand for a generalized incidence function, and from now on we will refer to it simply as an incidence function. Where any confusion could arise we will make clear the distinction between the original and generalized incidence functions.

Definition 4 Generalized Incidence Calculus Theories: *$< \mathcal{W}, \mu, P, \mathcal{A}, i >$ is called a generalized incidence calculus theory if the incidence function i satisfies the conditions*

$$i(false) = \{\} \qquad i(true) = \mathcal{W}$$

$$i(\phi_1 \wedge \phi_2) = i(\phi_1) \cap i(\phi_2) \quad for \quad \phi_i \in \mathcal{A}$$

where $\mathcal{W}$, μ and P are as defined in Definition 3 and $\mathcal{A}$ is closed under $\wedge$.

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

$$i_*(\phi) = \bigcup_{\psi \in \mathcal{A}, \psi \to \phi} i(\psi) \qquad i^*(\phi) = \mathcal{W} \setminus i_*(\neg\phi) \tag{4}$$

Where $\psi \to \phi$ means that for $\psi \in \mathcal{A}$ formula $\psi \to \psi$ is valid. This notation is used in the rest of the paper.

The lower bound represents the set of possible worlds which make ϕ true and the upper bound represents the set of possible worlds which fail to make $\neg\phi$ true. Function $p_*(\phi) = \mu(i_*(\phi))$ gives the degree of our belief in ϕ and function $p^*(\phi) = \mu(i^*(\phi))$ represents the degree we fail to believe in $\neg\phi$. For a formula ϕ in $\mathcal{A}$, if $p_*(\phi) = p^*(\phi)$, then $p(\phi)$ is defined as $p_*(\phi)$ and is called the probability of this formula. In this case, for any ϕ in $\mathcal{A}$, let $p(\phi \mid \psi)$ be the conditional probability of ϕ given ψ, we define

$$p(\phi \mid \psi) = \frac{p(\phi \wedge \psi)}{p(\psi)} \tag{5}$$

2.4 Basic Incidence Assignment

For a formula ϕ, suppose that ϕ is in its disjunctive normal form $\delta_1 \vee ... \vee \delta_l$. We define a subset A of $\mathcal{A}t$ as $A = \{\delta_1, ..., \delta_l\}$ and denote formula ϕ as ϕ_A. ϕ_A means that the disjunction of the elements in A is the disjunctive normal form of formula ϕ. $i(\phi_A)$ contains those possible worlds which make ϕ_A true and those possible worlds which make a ϕ_B true for $B \subset A$. So some of these possible worlds may only make ϕ_A true without making any of ϕ_B (for $B \subset A$) true.

For instance, suppose that we have two propositions q_1 and q_2 in P, then there are four basic elements in $\mathcal{A}t$ as $\delta_1 = q_1 \wedge q_2$, $\delta_2 = q_1 \wedge \neg q_2$, $\delta_3 = \neg q_1 \wedge q_2$ and $\delta_4 = \neg q_1 \wedge \neg q_2$.

If we are given that $i(\phi_{\{\delta_1\}}) = \{w_1\}$, $i(\phi_{\{\delta_1,\delta_2\}}) = \{w_1, w_2\}$ and $i(\phi_{\{\delta_1,\delta_3\}}) = \{w_1, w_3\}$, then w_2 makes only $\phi_{\{\delta_1,\delta_2\}} = q_1$ true without making $q_1 \wedge q_2$ true. Similarly w_3 makes q_2 true without making $q_1 \wedge q_2$ true.

In general, the subset of $i(\phi_A)$ which contains the possible worlds making only ϕ_A true without making any of ϕ_B ($B \subset A$) true is denoted as $ii(\phi_A)$ and the notation ii is called the **basic incidence assignment**. In order to show the relation between i and ii, we look at an example first. 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.* occurs 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

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

and the $\mathcal{A}t$ of P is $\mathcal{A}t = \{rainy \wedge windy, rainy \wedge \neg windy, \neg rainy \wedge windy, \neg rainy \wedge \neg windy\}$. A function ii could be naturally defined as:

$$ii(rainy \wedge windy) = \{fri, mon\}$$
$$ii(rainy) = \{sat, sun\}$$
$$ii(windy) = \{wed\}$$

For any other formula ϕ except *true*, $ii(\phi)$ is empty. It is easy to see that from ii, the incidence function can be recovered as:

$$i(rainy \wedge windy) = ii(rainy \wedge windy)$$

$$i(rainy) = ii(rainy) \cup ii(rainy \wedge windy)$$

$$i(windy) = ii(windy) \cup ii(rainy \wedge windy)$$

Definition 5 Basic Incidence Assignment: *Given a set of axioms $\mathcal{A}$, a mapping function ii: $\mathcal{A} \rightarrow \mathcal{W}$ is called a basic incidence assignment if ii satisfies the following conditions:*

$$\begin{array}{ll} ii(\phi) \neq \{\} & \phi \in \mathcal{A} \\ ii(\phi) \cap ii(\psi) = \{\} & \phi \neq \psi \\ ii(false) = \{\} & \\ ii(true) = \mathcal{W} \setminus \bigcup_j ii(\phi_j) & \end{array}$$

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

Here (and in the rest of the paper) $\phi \neq \psi$ means that either $\phi \rightarrow \psi$ is not valid or $\psi \rightarrow \phi$ is not valid.

Proposition 1 *Given a set of axioms $\mathcal{A}$ with a basic incidence assignment ii, then the function i defined by equation (6) is an incidence function on $\mathcal{A}$.*

$$i(\phi) = \bigcup_{\phi_j \in \mathcal{A}, \phi_j \rightarrow \phi} ii(\phi_j) \tag{6}$$

Proposition 2 *Given a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$, there exists a basic incidence assignment ii on $\mathcal{A}$ from which the incidence function i in the theory can be derived.*

The proofs of these propositions are given in Liu, Bundy and Robertson (1993). These two propositions tell us that a basic incidence assignment and its generalized incidence function can be recovered from each other.

From the above analysis, we find that whenever we need to get the probabilities on formulae, we always get the incidences for the formulae first. This is called the *indirect encoding of numerical uncertainty values*. Just because of this, it is therefore possible to propose an alternative combination rule in incidence calculus which combines evidence on the basis of possible worlds.

2.5 The Combination Rule in Incidence Calculus

Definition 6 Combination Rule: *Suppose there are two generalized incidence calculus theories $< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$, $< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$, then the joint impact of information carried by the two theories is represented by a quintuple: $< \mathcal{W} \setminus \mathcal{W}_0, \mu', P, \mathcal{A}, i >$ where*

$$\begin{array}{l} \mathcal{W}_0 = \bigcup\{i_1(\phi) \cap i_2(\psi) \mid (\phi \wedge \psi \rightarrow \perp), \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2\} \\ \mathcal{A} = \{\varphi \mid \varphi = \phi \wedge \psi \mid \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, \varphi \neq \perp\} \\ i(\varphi) = \bigcup\{i_1(\phi) \cap i_2(\psi) \mid (\phi \wedge \psi \rightarrow \varphi), \varphi \in \mathcal{A}, \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2, \phi \wedge \psi \neq \perp\} \end{array}$$

for any $w \in \mathcal{W} \setminus \mathcal{W}_0$

$$\mu'(w) = \frac{\mu(w)}{1-\Sigma_{w' \in \mathcal{W}_0}\mu(w')}$$

and let

$$i(false) = \{\} \qquad i(true) = \mathcal{W} \setminus \mathcal{W}_0$$

Here $\perp$ means false.

The 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.

$\mathcal{W}_0$ is a subset of $\mathcal{W}$ reflecting the conflict of two pieces of information and the conflict weight is $\Sigma_{w' \in \mathcal{W}_0}\mu(w')$. If the conflict weight is 1 then these two pieces of information are completely contradictory with each other and they cannot be combined using the rule.

When $\mathcal{A} = \{\}$, these two observations are irrelevant to each other and their combined result tells us nothing.

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

It is proved in Liu (1994) that the combined result is a generalized incidence calculus theory.

The crucial issue in applying the rule to two generalized incidence calculus theories is that these two theories are based on the same set of possible worlds, but based on different sets of axioms and incidence functions. The combination procedure unifies two sets of axioms into one set and two incidence functions into one incidence function. In this way, generalized incidence calculus is expected to be used to combine dependent evidence.

In general, the relations between two generalized incidence calculus theories (provided by two pieces of evidence) can be divided into the following three categories.

1). Two sets of possible worlds in the two generalized incidence calculus theories are the same. In this case, the Combination Rule above is applied to combine the two generalized incidence calculus theories.

2). Two sets of possible worlds in the two generalized incidence calculus theories are different and they are DS-independent[1]. In this case, it is possible to transform the two generalized incidence calculus theories into new forms so that two new generalized incidence calculus theories are based on the same set. Then the Combination Rule is applied on them. This is described in Theorem 1 below.

3). Two sets of possible worlds in the two generalized incidence calculus theories are different but not DS-independent. At the moment, we don't have a framework to deal with this situation in general. It has to be done individually. For a case in this category, if it is possible to find a common set of possible worlds in some way to replace the two existing sets of possible worlds, the Combination Rule is applicable. However, when it is not possible to find a common set of possible worlds to replace the two existing sets of possible worlds, generalized incidence calculus cannot cope with the case. Example 2.1 below demonstrates this situation.

It needs to be pointed out that if two sets of possible worlds in a case are different but they are both derived from a well-defined set, this case is put into the first category as shown in Example 4.4.

[1] The mathematical description of DS-independent can be found in Voorbraak (1991), Liu (1994), or alternatively in Shafer (1981), Shafer and Tversky (1985). This definition can be simply explained as, given (X_1, μ_1) and (X_2, μ_2), two spaces with their probability distributions, if either $\mu_1 \otimes \mu_2$ cannot be taken as the probability distribution on $X_1 \otimes X_2$, for instance, the two probability distributions are not probabilistically independent, or these two spaces are constructed from a pre-defined space (X, μ), then these two spaces are not DS-independent.

As cases in category 2 can be transformed into cases in category 1, category 2 is regarded as an extension of category 1.

Theorem 1 *Suppose we have two generalized incidence calculus theories,* $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$ *and* $< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$*, where* $(\mathcal{W}_1, \mu_1)$ *and* $(\mathcal{W}_2, \mu_2)$ *are DS-independent.Applying the Combination Rule to them we get* $< \mathcal{W}_3, \mu_3, P, \mathcal{A}_3, i_3 >$ *which is a generalized incidence calculus theory, where*

$$\mathcal{W}_0 = \bigcup\{i_1(\phi) \otimes i_2(\psi) \mid (\phi \wedge \psi \rightarrow \perp), \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2\}$$
$$\mathcal{W}_3 = \mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$$
$$\mathcal{A}_3 = \{\varphi \mid \varphi = \phi \wedge \psi, \phi \in \mathcal{A}_1, \psi \in \mathcal{A}_2,\ \varphi \neq \perp\}$$
$$i_3(\varphi) = \bigcup\{i_1(\phi) \otimes i_2(\psi) \mid (\phi \wedge \psi \rightarrow \varphi), \phi \wedge \psi \neq \perp\}$$

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

$$\mu_3(< w_{1l}, w_{2j} >) = \frac{\mu_1(w_{1l})\mu_2(w_{2j})}{1 - \Sigma_{< w_{1t}, w_{2m} > \in \mathcal{W}_0} \mu_1(w_{1t})\mu_2(w_{2m})}$$

Here $w_{1l}, w_{1t} \in \mathcal{W}_1$ *and* $w_{2j}, w_{2m} \in \mathcal{W}_2$

$\otimes$ means set production. $\mathcal{W}_0$ is the subset of $\mathcal{W}_1 \otimes \mathcal{W}_2$ which supports inconsistent conjunctions. The proof of this theorem is given in Liu (1994).

For any formula φ in $\mathcal{L}(P)$, our belief in φ is

$$p_*(\varphi) = \Sigma_{< w_{1l}, w_{2j} > \in i_3(\varphi)} \mu_3(< w_{1l}, w_{2j} >)$$

In the following, we say that two generalized incidence calculus theories are DS-independent if their sets of possible worlds (together with their probability distributions) are DS-independent.

For the joint product of spaces $\mathcal{W}_1$ and $\mathcal{W}_2$, an element $< w_{1l}, w_{2j} >$ in $\mathcal{W}_1 \otimes \mathcal{W}_2 \setminus \mathcal{W}_0$ tells us that possible worlds w_{1l} and w_{2j} support a formula simultaneously. Since $< w_{1l}, w_{2j} >$ in $\mathcal{W}_1 \otimes \mathcal{W}_2$ and $< w_{2j}, w_{1l} >$ in $\mathcal{W}_2 \otimes \mathcal{W}_1$ have the same meaning, we treat $\mathcal{W}_1 \otimes \mathcal{W}_2$ and $\mathcal{W}_2 \otimes \mathcal{W}_1$ as the same set. So the Combination Rule is both commutative and associative and the result of combining several generalized incidence calculus theories is unique irrespective of the sequence in which they are combined.

Example 1

We now use an example adopted from Pearl (1988, pp.58) to show the situation in which two generalized incidence calculus theories are based on different sets of possible worlds but these two sets are not DS-independent. The example is as follows.

> There are three prisoners, A, B and C, have been tried for murder, and their verdicts will be read tomorrow. They know only that one of them will be declared guilty and the other two will be set free. The identity of the condemned prisoner is revealed to the very reliable prison guard, but not to the prisoners themselves. In the middle of the night, Prisoner A calls the guard over and makes the following request: 'Please give this letter to one of my friends -to one who is to be released. You and I know that at least one of them will be freed'. Later Prisoner A calls the guard again and asks who received the letter. The guard answers, 'I gave the letter to Prisoner B, he will be released tomorrow'. After this Prisoner A feels that his chance to be guilty has been increased from 1/3 to 1/2. What did he do wrong?

Assume that I_B stands for the proposition 'Prisoner B will be declared innocent' and G_A stands for the proposition 'Prisoner A will be declared guilty'. The task is to compute the probability of G_A given all the information obtained from the Guard.

Solving this problems in formal probability theory, Pearl gets

$$Pr(G_A \mid I_B) = \frac{Pr(I_B \mid G_A)Pr(G_A)}{Pr(I_B)} = \frac{Pr(G_A)}{Pr(I_B)} = \frac{1/3}{2/3} = 1/2 \tag{7}$$

where $Pr(I_B \mid G_A) = 1$ since $G_A \supset I_B$ and $Pr(G_A) = Pr(G_B) = Pr(G_C) = 1/3$ from the prior probability distribution.

Pearl argues that this is a wrong result and the wrong result arises from omitting the *full context* in which the answer was obtained by Prisoner A. He further explains that 'By context we mean the entire range of answers one could possibly obtain, not just the answer actually obtained'. Therefore, Pearl introduces another proposition I'_B, stands for 'The guard said that B will be declared innocent', and he gives that

$$Pr(G_A \mid I'_B) = \frac{Pr(I'_B \mid G_A)Pr(G_A)}{Pr(I'_B)} = \frac{1/2.1/3}{1/2} = 1/3 \tag{8}$$

which he believes is the correct result.

Using incidence calculus to solve this problem, we let $P = \{G_A, G_B, G_C\}$ and G_A stand for the proposition 'Prisoner A is guilty'. Then it is possible to form a set of possible worlds $\mathcal{W}_1 = \{w_1, w_2, w_3\}$ with $\mu_1(w_j) = 1/3$ from the prior probability distribution. w_1 implies A is guilty.

From this information, a generalized incidence calculus theory is formed as $< \mathcal{W}_1, \mu_1, P, P, i_1 >$ where $i_1(G_A) = \{w_1\}$, $i_1(G_B) = \{w_2\}$ and $i_1(G_C) = \{w_3\}$.

After the guard passed the letter to a prisoner, it is possible to form another set of possible worlds $\mathcal{W}_2 = \{L_B, L_C\}$ where L_B means Prisoner B received the letter. $\mu_2(L_B) = \mu_2(L_C) = 1/2$.

So the second generalized incidence calculus theory is constructed as $< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$ where $i_2(G_A \vee G_C) = \{L_B\}$, $i_2(G_A \vee G_B) = \{L_C\}$ and $\mathcal{A}_2 = \{G_A \vee G_C, G_A \vee G_B\}$.

These two theories are based on different sets of possible worlds and they are not DS-independent. If we attempt to solve this example using Theorem 1, we can only get the result as shown in equation (7).

However whether it is possible to construct different generalized incidence calculus theories in order to reflect the *full context* of answers (the meaning of I'_B not I_B) remains open.

3 Dempster-Shafer theory of evidence

3.1 Basics of DS theory

The Dempster-Shafer theory of evidence, or as is sometimes called *belief function theory* (Shafer (1976), (1990), Smets (1988)), associates degrees of belief (*bel*) with every subset of a set. Such a set, S, is required to consist of mutually exclusive and exhaustive explanations for a problem. More precisely, at any time, one and only one element in S is the right answer to a question. The conjunction of any two elements is contradictory. A set S satisfying this requirement is called a *frame of discernment* (or frame).

In propositional logic, given a set of propositions P, P may not be a frame. However $\mathcal{A}t$, the basic element set formed from P defined in Definition 1, contains mutually exclusive and exhaustive answers to a question, so $\mathcal{A}t$ is a frame. In Fagin and Halpern (1989a), an arbitrarily defined frame S is taken to be a subset of some $\mathcal{A}t$ (in fact, given a frame S, it is always possible to define $\mathcal{A}t = S$ for a proper P). In the following, we follow the same idea and use $\mathcal{A}t$ to denote any frame of discernment.

A belief function bel on a frame $\mathcal{A}t$ is required to obey the following three features:

1.$bel(\emptyset) = 0$
2.$bel(\mathcal{A}t) = 1$
3.$bel(A_1 \cup ... \cup A_n) \geq \Sigma_i bel(A_i) - \Sigma_{i,j} bel(A_i \cap A_j) + \Sigma_{i,j,k} bel(A_i \cap A_j \cap A_k)...$
where $A_1, .., A_n$ are subsets of $\mathcal{A}t$.

A belief function bel on a frame of discernment $\mathcal{A}t$ is represented as $(\mathcal{A}t, bel)$, which is called a *DS structure*.

A belief function is usually described in the form of a function m called *mass function*, or *a basic probability assignment*. A mass function m is required to satisfy the following two conditions.

$$m(\emptyset) = 0 \qquad\qquad \Sigma_{A \subseteq \mathcal{A}t} m(A) = 1$$

$\emptyset$ represents the empty set.

Given a mass function on a frame of discernment, a corresponding belief function can be calculated as:

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

Given a belief function bel, its mass function m can be recovered by the equation:

$$m(A) = \Sigma_{B \subseteq A} (-1)^{|B|} bel(B)$$

The difference between a mass function and its belief function is that the degree of belief on a subset A of $\mathcal{A}t$ 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.

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

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

A subset A of $\mathcal{A}t$ is called a focal element of the belief function bel if $m(A) > 0$. We use A_{DS} to denote the set containing all the focal elements of a belief function.

3.2 Dempster's combination rule

In DS theory when two belief functions (given in the form of mass functions m_1 and m_2) are derived from two distinct (or DS-independent) pieces of evidence on the same frame of discernment $\mathcal{A}t$, their joint impact, carried by another belief function bel (or its mass function m) on $\mathcal{A}t$, can be obtained by using Dempster's combination rule. Dempster's rule is defined as:

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

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 $\mathcal{A}t$. Usually, $m = m_1 \oplus m_2$ says that m is the combined mass function from m_1 and m_2 and $\oplus$ means that Dempster's combination rule is used.

4 Comparison with DS Theory

In this section we discuss the relationship between DS theory and generalized incidence calculus. We are going to discuss their abilities to represent evidence and compare their abilities in combining evidence. We will prove that 1) they have the same ability in representing evidence, 2) any two pieces of evidence which can be combined using Dempster's combination rule, can also be combined in incidence calculus by applying Theorem 1. So Dempster's combination rule and Theorem 1 are totally equivalent. 3) those dependent cases which can be combined by the new Combination Rule (but not Theorem 1) cannot be combined by Dempster's combination rule.

Therefore we conclude that the new combination rule is superior to Dempster's combination rule.

4.1 Comparison I: Representing Evidence

Given a DS structure $(\mathcal{A}t, bel)$ and a generalized incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ where the frame $\mathcal{A}t$ is the basic element set of P, we say that this DS structure is equivalent to the generalized incidence calculus theory if for any $A \subseteq \mathcal{A}t$, $bel(A) = p_*(\phi_A)$. Here ϕ_A is defined as

$$\phi_A = \vee\delta_j \ \ where \ \ \delta_j \in A$$

That is if we use DS theory to describe the degree of belief on $\mathcal{A}t$, then we consider $\mathcal{A}t$ to be a frame, but if we use incidence calculus to describe the degree of belief on $\mathcal{A}t$, then we consider $\mathcal{A}t$ to be a collection of basic elements. Therefore, a subset A of $\mathcal{A}t$ in $2^{\mathcal{A}t}$ is treated to be equivalent to the formula $\vee\delta_j$ (where $\delta_j \in A$) in $\mathcal{L}(\mathcal{A}t)$.

Theorem 2 *For any DS structure $(\mathcal{A}t, bel)$, there is an equivalent generalized incidence calculus theory. For any subset A of $\mathcal{A}t$ and its corresponding formula ϕ_A in $\mathcal{L}(\mathcal{A}t)$, $bel(A)$ in the DS theory is equal to $p_*(\phi_A)$ in the generalized incidence calculus theory. That is*

$$bel(A) = p_*(\phi_A)$$

PROOF

Given a DS structure $(\mathcal{A}t, bel)$, suppose $A_{DS} = \{A_1, ...A_n\}$ is the focal element set of belief function bel and m is its mass function, then $\Sigma m(A_j) = 1$.

1) create a set of possible worlds $\mathcal{W} = \{w_1, ...w_n\}$ and let $\mu(w_j) = m(A_j)$.

2) let a subset $\mathcal{A}$ of $\mathcal{A}t$ be $\{\phi_{A_j} \mid A_j \in A_{DS}\}$;

3) define basic incidence assignment ii as $ii(\phi_{A_j}) = \{w_j\}$;

4) define incidence function i from ii as $i(\phi_A) = \{ii(\phi_{A_j}) \mid \phi_{A_j} \in \mathcal{A}$ and $\phi_{A_j} \rightarrow \phi_A$ is valid$\}$

Then $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$ is a generalized incidence calculus theory (it is easy to prove that i has the features of Definition 4 in Section 2).

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

$$\begin{aligned} p_*(\phi_A) &= \mu(i_*(\phi_A)) \\ &= \mu(\textstyle\bigcup_{\phi_{A_j} \in \mathcal{A}, \phi_{A_j} \rightarrow \phi_A} i(\phi_{A_j})) \\ &= \mu(\textstyle\bigcup_{\phi_{A_j} \in \mathcal{A}, \phi_{A_j} \rightarrow \phi_A} (\bigcup_{\phi_{A_l} \in \mathcal{A}, \phi_{A_l} \rightarrow \phi_{A_j}} ii(\phi_{A_l}))) \end{aligned}$$

$$
\begin{aligned}
&= \mu(\bigcup_{\phi_{A_l} \in \mathcal{A}, \phi_{A_l} \to \phi_A} ii(\phi_{A_l})) \\
&= \Sigma_{\phi_{A_l} \in \mathcal{A}, \phi_{A_l} \to \phi_A} \mu(ii(\phi_{A_l})) \\
&= \Sigma_{\phi_{A_l} \in \mathcal{A}, \phi_{A_l} \to \phi_A} \mu(\{w_l\}) \\
&= \Sigma_{A_l \in A_{DS}, A_l \subseteq A} m(A_l) \\
&= bel(A)
\end{aligned}
$$

Then the belief function $bel(A)$ is exactly the same as $p_*(\phi_A)$.
□

So $pls(A) = 1 - bel(\neg A) = 1 - p_*(\neg\phi_A) = \mu(\mathcal{W} \setminus i_*(\neg\phi_A)) = p^*(A)$.

This theorem tells us that the belief function on frame $\mathcal{A}t$ given by a 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. Therefore, any belief function can be obtained as a lower bound from a generalized incidence calculus theory.

Example 2

The example used here is originally from Fagin and Halpern (1989b) and simplified by Correa da Silva and Bundy (1990) as follows.

> A person has four coats: two are blue and single-breasted, one is grey and double-breasted and one is grey and single-breasted. To choose which colour of coat this person is going to wear, one tosses a (fair) coin. Once the colour is chosen, to choose which specific coat to wear the person uses a mysterious nondeterministic procedure which we don't know anything about. What is the probability of the person wearing a single-breasted coat?

We solve this problem by using DS theory first and then deal with it in generalized incidence calculus.

DS structure: Let $P = \{g, d\}$ where g stands for "the coat is grey" and d stands for "the coat is double-breasted", then we have

$$\mathcal{A}t = \{g \wedge d, \neg g \wedge d, g \wedge \neg d, \neg g \wedge \neg d\}$$

which is a frame. The element $\neg g \wedge d$ in this frame is false because there is no coat which is not grey but double-breasted. So the real frame of discernment is reduced to be

$$\mathcal{A}t = \{g \wedge d, g \wedge \neg d, \neg g \wedge \neg d\}$$

According to the story that one tosses a (fair) coin to decide which colour to choose, we can define a mass function on the frame $\mathcal{A}t$ as

$$m(\{\neg g \wedge \neg d\}) = 0.5 \qquad m(\{g \wedge \neg d, g \wedge d\}) = 0.5$$

with the focal element set A_{DS} as

$$A_{DS} = \{\{\neg g \wedge \neg d\}, \{g \wedge \neg d, g \wedge d\}\}$$

Therefore, we have a DS structure $(\mathcal{A}t, bel)$. The degree of belief on $\neg d$ is $bel(\neg d) = m(\neg g \wedge \neg d) = 0.5$ and the degree of plausibility is 1.

The degrees of belief and plausibility say that the probability of the person wearing a single-breasted coat lies somewhere between 0.5 to 1 which cannot be measured in a single number.

Generalized incidence calculus theory: Based on the story we could have two possible worlds: w_1 for blue and single-breasted coats and w_2 for grey coats. The probability of each of the possible worlds is 0.5.

Given a set of propositions P and its basic element set $\mathcal{A}t$ as defined in DS structure, we know that w_1 supports formula $\neg g \wedge \neg d$ and w_2 makes the formula $(g \wedge \neg d) \vee (g \wedge d)$ true. So we define $i(\neg g \wedge \neg d) = \{w_1\}$ and $i(g) = \{w_2\}$. Then $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ is a generalized incidence calculus theory.

From this generalized incidence calculus theory, we have that

$$i_*(\neg d) = i(\neg g \wedge \neg d)$$

$$i^*(\neg d) = \mathcal{W} \setminus i_*(d) = \mathcal{W}$$

so

$$p_*(\neg d) = 0.5 \qquad p^*(\neg d) = 1$$

which is identical to the result from DS theory.

Theorem 3 *For any generalized incidence calculus theory* $< \mathcal{W}, \mu, P, \mathcal{A}, i >$*, there is an equivalent DS structure* $(\mathcal{A}t, bel)$.

PROOF

Suppose $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ is a generalized incidence calculus theory and ii is the corresponding basic incidence assignment,

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

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 $m(A_j) = \mu(ii(\phi_{A_j}))$ where $A_j \in A_{DS}$. Then $\Sigma_{A_j} m(A_j) = 1$.

So bel: $bel(A) = \Sigma_{B \subseteq A} m(B)$ gives a belief function on $\mathcal{A}t$ and we obtain a DS structure $(\mathcal{A}t, bel)$.

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

$$\begin{aligned} p_*(\phi_A) &= \mu(i_*(\phi_A)) \\ &= \mu(\textstyle\bigcup_{\phi_B \in \mathcal{A}, \phi_B \rightarrow \phi_A} ii(\phi_B)) \\ &= \Sigma_{\phi_B \in \mathcal{A}, \phi_B \rightarrow \phi_A} \mu(ii(\phi_B)) \\ &= \Sigma_{B \subseteq A, B \in A_{DS}} m(B) \\ &= bel(A) \end{aligned}$$

Therefore, $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ and $(\mathcal{A}t, bel)$ are equivalent.

□

Example 3

This example demonstrates the procedure of producing a DS structure based on a given generalized incidence calculus theory as indicated in Theorem 2. This weather forecasting example continues the story in Section 2.

> Assume we know that on *fri, sat, sun, mon* it will rain and on *mon, wed, fri* it will be windy. The question we are interested in is on which days it will not rain.

Generalized incidence calculus theory: Let a set of possible worlds $\mathcal{W}$ be $\{sun, mon, tues, wed, thus, fri, sat\}$ and they have equal probability *i.e.* $\mu(w_i) = 1/7$ and let $P = \{rainy, windy\}$. The incidence function defined out of the above story is

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

the basic incidence assignment ii is

$$ii(rainy \wedge windy) = \{fri, mon\}$$
$$ii(rainy) = \{sat, sun\}$$
$$ii(windy) = \{wed\}$$
$$ii(\mathcal{A}t) = \{tues, thur\}$$

and the basic element set $\mathcal{A}t$ is

$$\mathcal{A}t = \{rainy \wedge windy, rainy \wedge \neg windy, \neg rainy \wedge windy, \neg rainy \wedge \neg windy\}$$

Therefore the generalized incidence calculus theory is

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

where $\mathcal{A} = \{rainy, windy, rainy \wedge windy\}$.

From this theory, we have

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

so

$$p_*(\neg rainy) = 0 \qquad p^*(\neg rainy) = 3/7$$

That is we cannot be sure on which day it will not rain but possibly on Tuesday, Wednesday and Thursday.

DS structure: For frame $\mathcal{A}t$ as defined above, we can derive a mass function m on it based on Theorem 2 as

$$m(rainy \wedge windy) = 2/7$$
$$m(rainy) = 2/7$$
$$m(windy) = 1/7$$
$$m(\mathcal{A}t) = 2/7$$

So we have $bel(\neg rainy) = 0$ and $pls(rainy) = 3/7$. The DS structure $(\mathcal{A}t, bel)$ gives the same result as incidence calculus.

A similar result has also been achieved in Correa de Silva and Bundy (1990). In their paper, it is proved that any original incidence calculus theory is equivalent to a Total Dempster-Shafer probability structure[2], and any Total Dempster-Shafer probability structure is equivalent to an original incidence calculus theory. In this paper, we have generalized incidence calculus theories and shown generalized incidence calculus theories are totally equivalent to DS structures.

[2] A Total Dempster-Shafer probability structure is used (Correa de Silva and Bundy (1990)) to represent a belief function on a frame in order to maintain the probability space from which the belief function is derived.

4.2 Comparison II: Combining DS-Independent evidence

For any two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$, if we assume that the two belief functions are derived from two DS-independent pieces of evidence, then these two belief functions can be combined using Dempster's combination rule. From these two DS structures, two generalized incidence calculus theories can also be produced, and their combination leads to the third generalized incidence calculus theory using Theorem 1. What we need to prove in such a situation is that the combined result of the two DS structures turns out to be equivalent to the combined generalized incidence calculus theory.

Theorem 4 *Suppose $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ are two DS structures and bel_1 and bel_2 are obtained from the two DS-independent pieces of evidence and assume that the combined DS structure is $(\mathcal{A}t, bel)$. Further let $< \mathcal{W}_1, \mu_1, \mathcal{A}t, \mathcal{A}_1, i_1 >$ and $< \mathcal{W}_2, \mu_2, \mathcal{A}t, \mathcal{A}_2, i_2 >$ be the two generalized incidence calculus theories produced from these DS structures, and $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$ be the combined generalized incidence calculus theory, then $(\mathcal{A}t, bel)$ is equivalent to $< \mathcal{W}, \mu, \mathcal{A}t, \mathcal{A}, i >$. That is, for any subset A of $\mathcal{A}t$,*

$$bel(A) = p_*(\phi_A)$$

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 $\mu(\mathcal{W}_0)$ in the combined generalized incidence calculus theory. In part two we need to prove that $bel(A) = p_*(\phi_A)$ for any $A \subseteq \mathcal{A}t$.

Because bel_1 and bel_2 are derived from two DS-independent pieces of evidence, $(\mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mu_2)$ are DS-independent. So Theorem 1 is used to combine these two derived generalized incidence calculus theories.

PROOF

Suppose the two focal element sets in these two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ are

$$A_{DS} = \{A_1, A_2, ..., A_n\} \ \Sigma m_1(A_l) = 1$$
$$B_{DS} = \{B_1, B_2, ..., B_m\} \ \Sigma m_2(B_j) = 1$$

The combined DS structure is $(\mathcal{A}t, bel)$ with bel defined as $bel_1 \oplus bel_2$.

Furthermore the two sets of axioms in the corresponding two generalized incidence calculus theories $< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 > < \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$ are:

$$\mathcal{A}_1 = \{\phi_{A_1}, \phi_{A_2}, ..., \phi_{A_n}\}, ii_1(\phi_{A_l}) = \{w_{1l}\}, \mu_1(w_{1l}) = m_1(A_l)$$
$$\mathcal{A}_2 = \{\psi_{B_1}, \psi_{B_2}, ..., \psi_{B_m}\}, ii_2(\phi_{B_j}) = \{w_{2j}\}, \mu_2(w_{2j}) = m_2(B_j)$$

Part One

Part one proves $k = \mu(\mathcal{W}_0)$ where k is the weight of the conflict between these two DS structures, and $\mathcal{W}_0$, which is defined in Section 2, is the conflict set in the combined generalized incidence calculus theory.

Suppose $m = m_1 \oplus m_2$, for any $A_l \cap B_j = \{\}$ $(A_l \in A_{DS}, B_j \in B_{DS})$, $m_1(A_l)m_2(B_j)$ will be a part of k. That is $k = k' + m_1(A_l)m_2(B_j)$.

For ϕ_{A_l} from A_l and ψ_{B_j} from B_j $(\phi_{A_l} \in \mathcal{A}_1, \psi_{B_j} \in \mathcal{A}_2)$, we have $\phi_{A_l} \wedge \psi_{B_j} \rightarrow \perp$ is valid. So

$$
\begin{aligned}
\mu(\mathcal{W}_0) &= \mu(\bigcup_{\phi_{A_l} \wedge \psi_{B_j} \to \perp} i_1(\phi_{A_l}) \otimes i_2(\psi_{B_j})) \\
&= \mu(\bigcup_{\phi_{A_l} \wedge \psi_{B_j} \to \perp} (\bigcup_{\phi_{A'_l} \to \phi_{A_l}} ii_1(\phi_{A'_l})) \otimes (\bigcup_{\psi_{B'_j} \to \psi_{B_j}} ii_2(\psi_{B'_j}))) \\
&= \mu(\bigcup_{\phi_{A_l} \wedge \psi_{B_j} \to \perp} (\bigcup_{\phi_{A'_l} \wedge \psi_{B'_j} \to \phi_{A_l} \wedge \psi_{B_j}} ii_1(\phi_{A'_l}) \otimes ii_2(\psi_{B'_j}))) \\
&= \mu(\bigcup_{\phi_{A'_l} \wedge \psi_{B'_j} \to \perp} ii_1(\phi_{A'_l}) \otimes ii_2(\psi_{B'_j})) \\
&= \Sigma(\mu_1(ii_1(\phi_{A'_l}))\mu_2(ii_2(\psi_{B'_j})) \mid \phi_{A'_l} \wedge \psi_{B'_j} \to \perp \text{ is valid}) \\
&= \Sigma(\mu_1(w_{1l'})\mu_2(w_{2j'}) \mid \phi_{A'_l} \wedge \psi_{B'_j} \to \perp \text{ is valid}) \\
&= \Sigma(m_1(A'_l)m_2(B'_j) \mid A'_l \cap B'_j = \{\}) \\
&= k
\end{aligned}
$$

Part Two

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

For $A_l \in A_{DS}$ and $B_j \in B_{DS}$, if $A_l \cap B_j \subseteq C$, then $m_1(A_i)m_2(B_j)$ is a part of $bel(C)$.

For ϕ_{A_l} from A_l and ψ_{B_j} from B_j, we have $\phi_{A_l} \wedge \psi_{B_j} \to \varphi_C$ is valid. So

$$
\begin{aligned}
p_*(\varphi_C) &= \mu(i_*(\varphi_C)) \\
&= \mu(\bigcup_{\phi_{A_l} \wedge \psi_{B_j} \to \varphi_C} i_1(\phi_{A_l}) \otimes i_2(\psi_{B_j})) \\
&= \Sigma_{\phi_{A_l} \wedge \psi_{B_j} \to \varphi_C}(\mu_1(i_1(\phi_{A_l}))\mu_2(i_2(\psi_{B_j}))) / (1-k) \\
&= \Sigma_{\phi_{A_l} \wedge \psi_{B_j} \to \varphi_C}(\mu_1(\bigcup_{\phi_{A'_l} \to \phi_{A_l}} ii_1(\phi_{A'_l}))\mu_2(\bigcup_{\psi_{B'_j} \to \psi_{B_j}} ii_2(\psi_{B'_j}))) / (1-k) \\
&= \Sigma_{\phi_{A_l} \wedge \psi_{B_j} \to \varphi_C}(\mu_1(\bigcup_{\phi_{A'_l} \to \phi_{A_l}} \{w_{1l'}\})\mu_2(\bigcup_{\psi_{B'_j} \to \psi_{B_j}} \{w_{2j'}\})) / (1-k) \\
&= \Sigma_{\phi_{A'_l} \wedge \psi_{B'_j} = \varphi_{C'}, \varphi_{C'} \to \varphi_C}(\mu_1(\{w_{1l'}\})\mu_2(\{w_{2j'}\})) / (1-k) \\
&= \Sigma_{C' \subseteq C, A'_l \cap B'_j = C'}(m_1(A'_l)m_2(B'_j)) / (1-k) \\
&= \Sigma_{C' \subseteq C} m(C') \\
&= bel(C)
\end{aligned}
$$

□

Example 4 Combining two DS-independent pieces of evidence using both Dempster's combination rule and Theorem 1 in generalized incidence calculus.

Using Dempster's combination rule

Assume that we have two DS structures $(\mathcal{A}t, bel_1)$ and $(\mathcal{A}t, bel_2)$ with the following additional information.

$$\mathcal{A}t = \{a, b, c, d\}$$

$$A_{DS} = \{\{a, b, c\}, \mathcal{A}t\}$$

$$B_{DS} = \{\{c, d\}, \mathcal{A}t\}$$

$$m_1(\{a, b, c\}) = 0.7, m_1(\mathcal{A}t) = 0.3$$

$$m_2(\{c, d\}) = 0.6, m_2(\mathcal{A}t) = 0.4$$

Table 1: Combination of two DS-independent pieces of evidence

A	$\{a,b,c\}$	$\mathcal{A}t$
m	0.7	0.3
$\{c,d\}$	$\{c\}$	$\{c,d\}$
0.6	0.42	0.18
$\mathcal{A}t$	$\{a,b,c\}$	$\mathcal{A}t$
0.4	0.28	0.12

Combining these two belief functions derived from m_1 and m_2, we get a joint belief function as shown in Table 1.

From this table, it is possible to calculate degrees of belief on any subsets of $\mathcal{A}t$. For instance, for subset $\{a,b,c\}$, we have $bel(\{a,b,c\}) = 0.42 + 0.28 = 0.7$ and $pls(\{a,b,c\}) = 1$.

Using the incidence calculus combination rule

From the two DS structures given above, we are able to form two generalized incidence calculus theories from them as

$$< \mathcal{W}_1, \mu_1, P, \mathcal{A}_1, i_1 >$$

$$< \mathcal{W}_2, \mu_2, P, \mathcal{A}_2, i_2 >$$

with the following additional information

$$\mathcal{W}_1 = \{w_{11}, w_{12}\}, \mu_1(w_{11}) = 0.7, \mu_1(w_{12}) = 0.3$$

$$P := \mathcal{A}t$$

$$\mathcal{A}_1 = \{a \vee b \vee c, \mathcal{A}t\}$$

$$i_1(a \vee b \vee c) = \{w_{11}\}, i_1(\mathcal{A}t) = \mathcal{W}_1$$

$$\mathcal{W}_2 = \{w_{21}, w_{22}\}, \mu_2(w_{21}) = 0.6, \mu_2(w_{22}) = 0.4$$

$$P := \mathcal{A}t$$

$$\mathcal{A}_2 = \{c \vee d, \mathcal{A}t\}$$

$$i_2(c \vee d) = \{w_{21}\}, i_2(\mathcal{A}t) = \mathcal{W}_2$$

As $(\mathcal{W}_1, \mu_1)$ and $(\mathcal{W}_2, \mu_2)$ are DS-independent, Theorem 1 in section 2 is used to combine these two incidence calculus theories as given in Table 2.

The combined generalized incidence calculus theory is $< \mathcal{W}_1 \otimes \mathcal{W}_2, \mu, P, \mathcal{A}, i >$ where $\mu(< w_{1l}, w_{2j} >) = \mu_1(w_{1l})\mu_2(w_{2j})$. From this theory, we are also able to obtain the degree of our belief in any formula. For example, $p_*(a \vee b \vee c) = \mu(i_*(a \vee b \vee c)) = 0.7$ and $p^*(a \vee b \vee c) = 1$ which are the same as we got in DS theory.

Comparing Table 1 with Table 2 we will find that these two structures give the same result (numerically) on any subset (or formula). We will also find that whenever a numerical value (mass value) appears in Table 1, a corresponding incidence set replaces its position in Table 2.

Table 2: Combination of two DS-independent generalized incidence calculus theories

ϕ_A	$a \vee b \vee c$	$true$
$i(\phi_A)$	$\{w_{11}\}$	$\mathcal{W}_1$
$c \vee d$	c	$c \vee d$
$\{w_{21}\}$	$\{w_{11}\} \otimes \{w_{21}\}$	$\mathcal{W}_1 \otimes \{w_{21}\}$
$true$	$a \vee b \vee c$	$true$
$\mathcal{W}_2$	$\{w_{11}\} \otimes \mathcal{W}_2$	$\mathcal{W}_1 \otimes \mathcal{W}_2$

The combination procedure in generalized incidence calculus combines possible worlds instead of numbers. The degree of belief in a formula is calculated based on the incidence set.

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 section we are going to show that we can handle a wider range of information in incidence calculus by applying the new combination rule.

4.3 Comparison III: Combining Dependent Evidence

In this section, we first show an example which can be dealt with using the combination rule in incidence calculus but cannot be dealt with using Dempster's combination rule. We then simply explore the theoretical difference between the two theories and argue why DS theory fails to deal with dependent evidence while incidence calculus succeeds.

Example 5

This example is from Voorbraak (1991). There are 100 labelled balls in an urn as given in Table 3.

Table 3: 100 balls and their labels

Label	Number of Balls	Subset Name in $\mathcal{W}$
axy	4	W_1
ax	4	W_2
ay	16	W_3
a	16	W_4
bxy	10	W_5
bx	10	W_6
by	20	W_7
b	20	W_8

Suppose X and Y are separate observations on drawing a ball from the urn. The information carried by them is:

X: the drawn ball has label x;

Y: the drawn ball has label y.

Based on these two pieces of evidence, we are interested in knowing the degree of our belief that the drawn ball also has label b.

Using Dempster's combination rule:

Let a set of propositions P be $\{a, b\}$. a stands for a proposition 'The drawn ball has label a' and b stands for the proposition 'The drawn ball has label b'. Then the basic element set $\mathcal{A}t$ is the same as P which is a frame. Two mass functions are defined on $\mathcal{A}t$ based on the information carried by the two observations X and Y as:

$$m_X(a) = 2/7, \;\; m_X(b) = 5/7$$
$$m_Y(a) = 2/5, \;\; m_Y(b) = 3/5$$

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

The result of applying Dempster's combination rule to the above two mass functions is $m(b) = m_X \oplus m_Y(b) = 15/19$. So $bel(b) = 15/19$.

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 and the result given in DS theory is wrong. See the detailed analysis of the example in Voorbraak (1991).

Using the incidence calculus combination rule:

Let us examine this example in incidence calculus theory. First of all, we suppose that the set of possible worlds $\mathcal{W}$ contains 100 labelled balls.

$$\mathcal{W} = W_1 \cup W_2 \cup W_3 \cup W_4 \cup W_5 \cup W_6 \cup W_7 \cup W_8$$

where W_1 contains 4 possible worlds each of which specifies a ball with labels xya, ..., W_8 contains 20 possible worlds each of which specifies a ball with label b. The probability distribution on $\mathcal{W}$ is $\mu(w) = 1/100$ 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 etc.

From observations X and Y, it is possible to construct two generalized incidence calculus theories

$$< \mathcal{W}, \mu, P, \mathcal{A}_1, i_1 >$$
$$< \mathcal{W}, \mu, P, \mathcal{A}_2, i_2 >$$

where

$$i_1(x) = W_1 \cup W_2 \cup W_5 \cup W_6$$
$$i_1(a \wedge x) = W_1 \cup W_2, \quad\; i_1(b \wedge x) = W_5 \cup W_6$$

and

$$i_2(y) = W_1 \cup W_3 \cup W_5 \cup W_7$$
$$i_2(a \wedge y) = W_1 \cup W_3, \quad\; i_2(b \wedge y) = W_5 \cup W_7$$

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 proposed in incidence calculus to these two theories, we can get the third incidence calculus theory $< \mathcal{W}, \mu, P, \mathcal{A}, i >$ with $\mathcal{A} = \{x \wedge y, a \wedge x \wedge y, b \wedge x \wedge y, a \wedge b \wedge x \wedge y\}$

$$i(b \wedge x \wedge y) = W_5 \qquad i(x \wedge y) = W_1 \cup W_5$$

$$i(a \wedge x \wedge y) = W_1 \qquad i(a \wedge b \wedge x \wedge y) = \{\}$$

It is easy to prove that for any $\phi \in \mathcal{A}$, $i_*(\phi) = i^*(\phi) = i(\phi)$, so a function p, defined as $p(\phi) = p_*(\phi) = \mu(i(\phi))$, is a probability function on $\mathcal{A}$. Further because $p(b \wedge x \wedge y) = 10/100$ and $p(x \wedge y) = 14/100$, according to Equation (5) in Section 2, we have

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

This result is consistent with what we could get in probability theory.

Now we try to explain theoretically why Dempster's combination rule cannot be used in this case. In fact, the two mass functions are derived from two probability spaces (S_1, μ_1) and (S_2, μ_2) where $S_1 = W_1 \cup W_2 \cup W_5 \cup W_6$, $\mu_1(s) = 1/28$ and $S_2 = W_1 \cup W_3 \cup W_5 \cup W_7$ and $\mu_2(s) = 1/50$. These two probability spaces are defined from the unique space $(\mathcal{W}, \mu)$ and they share the information carried by the subset $W_1 \cup W_5$. Therefore Dempster's combination rule cannot be used to combine the two mass functions derived from the two probability spaces.

In incidence calculus, instead of combining numbers on set $\mathcal{A}t$, we combine two pieces of evidence symbolically at the original information level, *i.e.*, at the probability space level. For the above example, since the two probability spaces are somehow related to the unique space $(\mathcal{W}, \mu)$, we establish two generalized incidence functions from $\mathcal{W}$ to P rather than from S_1 and S_2 to P respectively. Therefore it is possible to cancel the overlapped information carried by the two observations. Because DS theory is a purely numerical uncertainty reasoning mechanism, it is not possible to combine evidence symbolically. So it is not possible to represent and cancel the joint (or overlapped) part of the information provided by two pieces of evidence.

Therefore, we conclude that even though the two theories have the same ability in representing evidence and combining DS-independent information, their theoretical structures are rather different. The essence of incidence calculus, *indirect encoding of probabilities of formulae,* makes it possible to cancel the effect of overlapped information and provide an alternative combination mechanism which combines dependent information in some situations. Although trying to combine dependent information at the probability space level has been considered in Shafer (1982) and in Lingras and Wong (1990), no unique rule was provided in DS theory for general cases because of the theoretical limitation of the theory.

5 Conclusions

In this paper, we described generalized incidence calculus and made a comprehensive comparison between DS theory and generalized incidence calculus on their abilities in the following three aspects: 1) representing evidence; 2) combining DS-independent evidence; 3) dealing with dependent evidence. We conclude that these two theories have the same ability in representing incomplete information and combining DS-independent evidence. However, incidence calculus is superior to DS theory in coping with overlapped information. This difference results from their different theoretical structures. DS theory is a pure numerical approach while incidence calculus possesses both symbolic and numerical features. That is incidence calculus can make an inference either at the symbolic level by producing incidence sets or at the numerical level by calculating lower or upper bounds on probabilities of formulae.

In general, independent relations among multiple sources of evidence can be considered as special cases of dependent situations. As Pearl indicated (Pearl (1992)), "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 combination rule, we have indeed adopted the same idea and made some efforts towards combining dependent evidence. This result would be useful for further research work on either this topic or the relevant topics. It tells us that it is a promising way to cancel the overlapped and duplicated information from several pieces of evidence at the symbolic level rather than at the numerical level.

Apart from its similarities with DS theory, generalized incidence calculus has also shared some similarities with the *propositional probability structure* in Bacchus (1990). Both of the generalized incidence calculus theory and propositional probability structure use possible worlds to define and explain the probability of a formula, but possible worlds in incidence calculus remain separated from formulae while possible worlds are a part of formulae in propositional probability structures in Bacchus (1990). Separating possible worlds from formulae makes incidence calculus possible to carry out inferences both at the symbolic level and the numerical level but the embedding of possible worlds into formulae makes propositional probability structures easy to be extended to sentences in first order logic.

Acknowledgement

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

The first author is being supported by a "Colin and Ethel Gordon Scholarship" of Edinburgh University and an ORS award. The computer facilities used for this research are supported by an SERC Grant GR/H/23610.

References

Bacchus, F. (1990). *Representing and Reasoning with Probabilistic Knowledge*, The MIT Press.

Bundy, A. (1985). Incidence calculus: a mechanism for probability reasoning, *J. of Automated Reasoning*, **1**, 263-283. It is also available as Research Paper 216 in the Dept. of Artificial Intelligence, Univ. of Edinburgh.

Bundy, A. (1992). Incidence Calculus, *The Encyclopedia of AI*, 663-668. It is also available as Research Paper 497 in the Dept. of Artificial Intelligence, Univ. of Edinburgh.

Correa da Silva, F. and Bundy, A. (1990). On some equivalence relations between incidence calculus and Dempster-Shafer theory of evidence, *Proc. of sixth workshop of Uncertainty in Artificial Intelligence*, 378-383. It is also available as Research Paper 470 in the Dept. of Artificial Intelligence, Univ. of Edinburgh.

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

Fagin, R. and Halpern, J. (1989a). Uncertainty, belief and probability, *IJCAI-89*, 1161-1167.

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

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

Hunter, D. (1987). Dempster-Shafer vs. probabilistic logic, *Proc. Third AAAI uncertainty in artificial intelligence workshop*, 22-29.

Lingras, P. and Wong, S.K. (1990). Two perspectives of Dempster-Shafer theory of evidence functions, *Int. J. Man-Machine Studies*, **33**, 117-125.

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

Liu, W., Bundy, A. and Robertson, D. (1993). Recovering incidence functions, 2^{nd} *European conference on symbolic and quantitative approaches to reasoning and uncertainty*, Lecture notes in Computer Science 747, 241-248, Springer. Full paper is available as Research Paper 648 in the Dept. of Artificial Intelligence, Univ. of Edinburgh.

Liu, W. (1994). *Extended incidence calculus and its comparison with related theories*, Forthcoming PhD thesis.

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

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

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

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

Shafer, G. (1976). *A mathematical theory of evidence*, Princeton University Press.

Shafer, G. (1981). Constructive probability, *Synthese*, **48**, 1-60.

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

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

Shafer, G. and Tversky, A. (1985). Languages and Designs for Probability Judgement, *Cognitive Science*, **9**, 309-339.

Smets, P. (1988). Belief functions, *Non-Standard Logics for Automated Reasoning*, (Smets, Mamdani, Dubois and Prade Eds.), 253-286.

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

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

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**, 85-90.