Theory of Evidence - A Survey of its Mathematical Foundations, Applications and Computational Aspects

Jürg Kohlas, Paul-André Monney *)

Key Words

Uncertainty, Evidence, Dempster-Shafer Theory, Belief Functions

Abstract

The mathematical theory of evidence has been introduced by Glenn Shafer in 1976 as a new approach to the representation of uncertainty. This theory can be represented under several distinct but more or less equivalent forms. Probabilistic interpretations of evidence theory have their roots in Arthur Dempster's multivalued mappings of probability spaces. This leads to random set and more generally to random filter models of evidence. In this probabilistic view evidence is seen as more or less probable arguments for certain hypotheses and they can be used to support those hypotheses to certain degrees. These degrees of support are in fact the reliabilities with which the hypotheses can be derived from the evidence. Alternatively, the mathematical theory of evidence can be founded axiomatically on the notion of belief functions or on the allocation of belief masses to subsets of a frame of discernment. These approaches aim to present evidence theory as an extension of probability theory. Evidence theory has been used to represent uncertainty in expert systems, especially in the domain of diagnostics. It can be applied to decision analysis and it gives a new perspective for statistical analysis. Among its further applications are image processing, project planing and scheduling and risk analysis. The computational problems of evidence theory are well understood and even though the problem is complex, efficient methods are available.

1. Origins

The term "Evidence Theory" was coined by Glenn Shafer in his book "A Mathematical Theory of Evidence", published by Princeton University Press in 1976. This work was initiated by a course on statistical inference taught by Arthur Dempster at Harvard University. Dempster developed there a theory of lower and upper probabilities in an attempt to reconcile Bayesian statistics with Fisher's fiducial argument (Dempster 1967, 1968). As Shafer states in the preface of his book "... It offers a reinterpretation of Dempster's work, a reinterpretation that identifies his "lower probabilities" as epistemic probabilities or degrees of belief, takes the rule for combining such degrees of belief as fundamental, and abandons the idea that they arise as lower bounds over classes of Bayesian probabilities." The rule mentioned in this statement became known as Dempster's rule.

*) *Institute for Automation and Operations Research, University of Fribourg, CH-1700 Fribourg (Switzerland).* *Research partly supported by Grants No. 21-30186.90 and 21-32660.91 of the Swiss National Foundation for Scientific Research.*

"Evidence" is an appealing term and it is no surprise that the "Theory of Evidence" found much interest among knowledge engineers who have to model uncertain information. After a lot of ad hoc attempts to model uncertainty in expert systems, a serious theory seemed finally available to treat one of the basic problems of the field. However, as the theory is not really compatible with the simple paradigms of early rule based expert systems, misinterpretations and oversimplistic misuses of the theory caused much deception. Only slowly begins the true nature and the real meaning of evidence theory to emerge and to be understood. And this will hopefully be the starting point for fruitful and appropriate applications to the modeling of uncertainty, decision analysis and knowledge engineering in general.

Evidence is a notion which probably can never be fully captured by a single formal theory. In this survey however "Theory of Evidence" will be understood in a narrow sense as the theory introduced by Dempster and Shafer and variants thereof. This particular theory is also often called Dempster-Shafer theory. It is clear today that this theory can be given various different, but essentially equivalent mathematical forms. Some of them are based on probability theory, others are axiomatic theories, a priori without a reference to probability theory. According to this distinction we classify the approaches broadly into *probabilistic* approaches and *non probabilistic* ones. The former try to integrate evidence theory into the framework of classical probability theory, whereas the latter deliberately go beyond classical probability. May be it would be more correct to call them rather non-standard probability theories than non-probabilistic ones. In section 2 a survey of these approaches is given. Despite the differences in approach and interpretation, all of them lead essentially to mathematically equivalent theories – at least in the finite case. That is they share the same basic theorems and the same computational procedures apply.

An important pragmatic consideration regarding the usefulness of evidence theory is the question, whether there are indeed important practical applications which can be described in a natural and useful way by the structures provided by the theory of evidence. This question is addressed in section 3. From the point of view of computation, evidence theory is complex in the usual technical sense. Nevertheless there exist methods which greatly improve the efficiency of computations. Also, based on these methods, software packages begin to appear which permit to apply evidence theory to practical problems. This subject is discussed in section 4.

Today there are not many comprehensive and introductory presentations and surveys of evidence theory. Beside the book of Shafer (1976 a), which is still a recommended reading, there are recent surveys by Smets, 1988, Shafer, 1990 and 1992. In addition there are some recent publications covering besides evidence theory other approaches to uncertainty: There is a volume edited by G. Shafer and J. Pearl (1990), there are the proceedings of the Conferences on Uncertainty in AI (for example Bonissone et al. (1991), D'Ambrosio et al. (1991) and Dubois et al. (1992)), the proceedings of the Conferences on Information Processing and Management of Uncertainty (IPMU), the latest one edited by Bouchon-Meunier et al. (1991) and the proceedings on the European Conference on Symbolic and Quantitative Approaches to Uncertainty (ECSQAU), Kruse, Siegel (1991). Recent books on uncertainty including chapters on evidence theory are by Kruse et al. (1991) and Hajek et al., 1992. Finally, let's mention a forthcoming book edited by Fedrizzi et al. (1993).

2. The Theory of Evidence - Differing Approaches and Interpretations

First the probablistic approaches to the Theory of Evidence will be introduced in this section and then the non-probabilistic ones. This will be done in the framework of finite frames, as most of the present publications treat this case. The interesting infinite case will be discussed separately. In each section, the different approaches will be presented more or less in their historical order.

2.1 Probabilistic approaches

(i) *Dempster's Multivalued Mappings (Dempster, 1967).* The following citation (up to changed symbols) is taken from Dempster (1967 a): "Consider a pair of spaces Ω and Θ and a multivalued mapping Γ, which assigns a subset $\Gamma(\omega)$ to every $\omega \in \Omega$. Suppose furthermore, that P is a probability measure on Ω, which assigns probabilities to the members of a class a of subsets of Ω. If P is acceptable for probability judgments about an uncertain outcome $\omega \in \Omega$, and if this uncertain outcome ω is known to correspond to an uncertain outcome $\theta \in \Gamma(\omega)$, what probability judgments may be made about the uncertain outcome $\theta \in \Gamma(\omega)$ ...". In the following discussion we shall suppose that Ω and Θ are finite sets and $\mathcal{A}$ the algebra of all subsets of Ω. The structure $(\Omega, P, \Gamma, \Theta)$ introduced by Dempster can be given different interpretations, one of which will be given here and another one in subsection 2.2 (i).

Suppose that a certain precise question,whose answer is unknown, has to be studied and the elements θ of Θ represent the possible answers to the question. This means that exactly one of the $\theta \in \Theta$ is the correct answer, but it is unknown which one. However there is some information or evidence available relative to this question. This information allows for several, distinct interpretations, depending on some unknown circumstances and these interpretations are represented by the elements ω of Ω. This means that there is exactly one correct interpretation ω in Ω, but again it is unknown which one. Not all interpretations are equally likely and the probabilities $p(\omega)$ describe these different likelihoods. If $\omega \in \Omega$ is the correct interpretation, then the unknown answer θ is known to be in the set $\Gamma(\omega)$. Such a piece of information is called a **hint** (Kohlas, 1990). This model allows to give a specific and clear sense to the further important notions introduced by Dempster (1967).

If H is a subset of Θ, the hypothesis that the unknown answer θ is in H can be considered. In order to judge the hypothesis H in the light of the hint $(\Omega, P, \Gamma, \Theta)$, we may ask, which of the possible interpretations make H necessarily true. These are all interpretations ω having a non empty set $\Gamma(\omega)$ which is contained in H because if such an interpretation is the correct one, then $\theta \in \Gamma(\omega)$ and thus necessarily $\theta \in H$. Define

$$u(H) = \{\omega \in \Omega : \emptyset \neq \Gamma(\omega) \subseteq H\}. \tag{2.1}$$

In the same spirit, we may ask what are the interpretations which would make H not necessarily true, but at least possible. These interpretations are those in

$$v(H) = \{\omega \in \Omega : \Gamma(\omega) \cap H \neq \emptyset\} \tag{2.2}$$

because if $\omega \in v(H)$ is the correct interpretation, then $\theta \in \Gamma(\omega)$ and it is at least possible that θ is in H, but not sure.

Note that $v(\Theta)$ contains all interpretations with $\Gamma(\omega) \neq \emptyset$. Now, because Θ is supposed to contain the true answer, an interpretation ω which is not in $v(\Theta)$ is not really a possible interpretation. Thus, the correct interpretation must be in $v(\Theta)$ and this supplementary information allows to pass to the conditional probability $p(\omega|v(\Theta)) = p(\omega)/P(v(\Theta))$ for the possible interpretations in $v(\Theta)$.

As the correct interpretation is unknown, it is not possible to confirm whether it is in $u(H)$ or $v(H)$, that is whether H is true or only possible. But it is at least possible to compute the probabilities that the correct interpretation is in $u(H)$ or $v(H)$:

$$sp(H) = P(u(H)|v(\Theta)) = P(u(H))/P(v(\Theta)), \tag{2.3}$$
$$pl(H) = P(v(H)|v(\Theta)) = P(v(H))/P(v(\Theta)). \tag{2.4}$$

Dempster (1967) called $sp(H)$ the *lower probability* of H and he pointed out that $sp(H)$ can be regarded as the minimal amount of probability which can be transfered from Ω to outcomes $\theta \in H$. Similarly, he called $pl(H)$ the *upper probability* of H, which can be regarded as the largest possible amount of probability, which can be transfered to outcomes $\theta \in H$. This point of view will be reconsidered below (point (iv)). Alternatively, because the ω in $u(H)$ can all be regarded as "arguments" in favour of the hypothesis H, arguments however, which are not sure to hold but only more or less probable, $sp(H)$ can be regarded as the probability with which H can be infered or proved from the available information $(\Omega, P, \Gamma, \Theta)$. From this point of view $sp(H)$ can be regarded as the *degree of support* of the hypothesis H by the hint $(\Omega, P, \Gamma, \Theta)$. Similarly, all the interpretations ω in the set $v(H)$ can be regarded as "arguments" for the possibility of H. Then $pl(H)$ can be considered as the *degree of possibility* or *plausibility* of H (the degree of possibility is related to a corresponding but not identical notion in the theory of possibility (Dubois, Prade, 1985)). This point of view leads in a natural way to a theory of the reliability of reasoning with unreliable arguments where proofs are treated as systems whose components may fail and therefore have only a limited reliability (Pearl, 1988; Provan, 1990 a; Kohlas, 1991 a, see also point (iii) below).

Considering all possible hypotheses $H \subseteq \Theta$, sp and pl become functions from the power set of 2^Θ to $[0,1]$ and as such they are called support (also belief) and plausibility functions. The following theorem collects a number of fundamental properties of these functions.

Theorem 2.1 If sp and pl are defined by (2.3) and (2.4) relative to a hint $(\Omega, P, \Gamma, \Theta)$, then

$$sp(\emptyset) = pl(\emptyset) = 0; \; sp(\Theta) = pl(\Theta) = 1 \tag{2.5}$$

$$sp(H) \leq pl(H) \text{ for all } H \subseteq \Theta \tag{2.6}$$

$$sp(H) = 1 - pl(H^c), \; pl(H) = 1 - sp(H^c), \tag{2.7}$$

and

$$sp(H) \geq \sum \{(-1)^{|I|+1} sp(\cap_{i \in I} H_i) : \emptyset \neq I \subseteq \{1, \ldots, n\}\} \tag{2.8}$$

for all $n \geq 1$ and sets H, H_i in Θ such that $H \supseteq H_i$ and

$$pl(H) \leq \sum \{(-1)^{|I|+1} pl(\cup_{i \in I} H_i) : \emptyset \neq I \subseteq \{1, \ldots, n\}\} \tag{2.9}$$

for all $n \geq 1$ and sets H, H_i in Θ such that $H \subseteq H_i$.

A proof of these properties may be found for example in Shafer (1976 a). The inequality (2.8) says that sp is monotone of order ∞ and (2.9) states that pl is alternating of order ∞. Finally, (2.8) together with (2.5) says that sp is a Choquet capacity of order ∞.

It is possible that one has two or more hints or sources of information $\mathcal{H}_i = (\Omega_i, P_i, \Gamma_i, \Theta)$, $i = 1, \ldots, m$ relative to the same question. Then this information has to be combined in order to integrate all the available information. As exactly one interpretation $\omega_i \in \Omega_i$ is the correct one for every hint, there is exactly one correct combined interpretation $(\omega_1, \ldots, \omega_m)$. This combined interpretation restricts the correct answer θ to the question considered into the subset $\Gamma(\omega_1, \ldots, \omega_m) = \Gamma_1(\omega_1) \cap \ldots \cap \Gamma_m(\omega_m)$ of Θ. Thus $\Omega_1 \times \ldots \times \Omega_m$ is the set of all combined interpretations to be considered. They have a joint probability $p_{1 \ldots m}(\omega_1, \ldots, \omega_m)$, which has as marginal probabilities $p_i(\omega_i), i = 1, \ldots, m$. This consideration leads then to the combined hint

$$\mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m = (\Omega_1 \times \ldots \times \Omega_m, P_{1\ldots m}, \Gamma, \Theta). \tag{2.10}$$

In many cases one may assume that the interpretations of the distinct hints are stochastically independent, such that $P_{1\ldots m}$ is the product measure on $\Omega_1 \times \ldots \times \Omega_m$. For this case of independent hints or sources of information, (2.10) is called *Dempster's rule of combination* (seemingly different, but equivalent forms of Dempster's rule are given below). The combined interpretations $(\omega_1, \ldots, \omega_m)$ for which $\Gamma_1(\omega_1) \cap \ldots \cap \Gamma_m(\omega_m)$ is empty are called contradictory. If all combined interpretations are contradictory, the hints are not combinable, they are in complete contradiction.

It is possible that the subsets $\Gamma(\omega)$ are all identical and equal to a subset B of Θ for all interpretations ω. This says simply that the correct answer is for sure in the subset B. Such a hint is called *deterministic* and is simply denoted by B. If $\mathcal{H}$ is any other hint relative to the same question, then $\mathcal{H} \oplus B$ can be formed. This is called conditioning of the hint $\mathcal{H}$. The following theorem holds:

Theorem 2.2 (see for example Shafer, 1976 a). Let $sp, sp(.|B)$ and $pl, pl(.|B)$ denote the support and plausibility functions of $\mathcal{H}$ and $\mathcal{H} \oplus B$ respectively. Then

$$sp(H|B) = (sp(H \cup B^c) - sp(B^c))/(1 - sp(B^c), \tag{2.11}$$
$$pl(H|B) = pl(H \cap B)/pl(B). \tag{2.12}$$

This is called Dempster's rule of conditioning. In particular, it is possible that $B = \Theta$. Then the deterministic hint is called *vacuous*, it carries no information whatsoever concerning the question considered. For a vacuous hint $sp(H) = 0$ for all $H \neq \Theta$ and $pl(H) = 1$ for all $H \neq \emptyset$. This is a perfect representation of complete ignorance. Clearly we have $\mathcal{H} \oplus \Theta = \mathcal{H}$.

If for all ω the sets $\Gamma(\omega)$ are singletons, then the hint is called *precise* or *Bayesian* (by Shafer, 1976 a). It is then essentially a random variable. In this case $sp = pl$ and sp becomes in fact a probability measure on Θ and Dempster's rule of conditioning reduces to the usual definition of conditional probability. This is generalized in the following theorem.

Theorem 2.3 (see for example Shafer, 1976 a). Let $\mathcal{H}_1$ be a precise hint, $\mathcal{H}_2$ an arbitrary hint. Then $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ is also a precise hint. Denote by p, p_1 the probabilities induced by $\mathcal{H}$ and $\mathcal{H}_1$, and by pl_2 the plausibility function relative to $\mathcal{H}_2$. Then, for all $\theta \in \Theta$,

$$\begin{aligned} p(\{\theta\}) &= k\ p_1(\{\theta\})pl(\{\theta\}), \\ k^{-1} &= \sum_{\theta \in \Theta} p_1(\{\theta\})pl(\{\theta\}). \end{aligned} \tag{2.13}$$

This shows that little information is needed in this case from the second hint $\mathcal{H}_2$. The plausibilities of the singletons are sufficient, in fact it is even sufficient to know only the relative values of these plausibilities. This is a quite remarkable result, which permits to link evidence theory with the usual Bayesian analysis, a subject which cannot be pursued here.

The following theorem shows how to obtain support and plausibility functions for combined hints in the general case.

Theorem 2.4 (Kohlas, 1990, Vakili, 1993). Let $\mathcal{H}_i = (\Omega_i, P_i, \Gamma_i, \Theta), i = 1, \ldots, m$ be independent hints relative to the same question with support and plausibility function sp_i, pl_i, and let sp and pl denote the support and plausibility functions of $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m$. Then, for all subsets H of Θ,

$$sp(H) = c \sum \{sp_1(H^c \cup (\cup_{i=2}^m \Gamma_i(\omega_i)^c)) \prod_{i=2}^m P_i(\omega_i) : (\omega_2, \ldots \omega_m) \in \Omega_2 \times \ldots \times \Omega_m\} - (c-1), \quad (2.14)$$

$$pl(H) = c \sum \{pl_1(H \cap (\cap_{i=2}^m \Gamma_i(\omega_i))) \prod_{i=2}^m P_i(\omega_i) : (\omega_2, \ldots \omega_m) \in \Omega_2 \times \ldots \times \Omega_m\}, \quad (2.15)$$

where

$$c^{-1} = \sum \{pl_1(\cap_{i=2}^m \Gamma_i(\omega_i)) \prod_{i=2}^m P_i(\omega_i) : (\omega_2, \ldots \omega_m) \in \Omega_2 \times \ldots \times \Omega_m\}. \quad (2.16)$$

Dempster (1967 a) introduced another set function which has no particular interesting interpretation, but which has a simple transformation under Dempster's rule of combination. The function

$$q(H) = \sum \{p(\omega) : \Gamma(\omega) \supseteq H\} \quad (2.17)$$

is called the *commonality function*. When hints are combined, this function can simply be multiplied:

Theorem 2.5 (see Dempster, 1967 a, Shafer, 1976 a). Let $\mathcal{H}_i = (\Omega_i, P_i, \Gamma_i, \Theta), i = 1, \ldots, m$ be independent hints relative to the same question with commonality functions q_i, and let $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m$ with commonality function q. Then for all subsets H of Θ,

$$q(H) = q_1(H) \cdots q_m(H). \quad (2.18)$$

This nice property facilitates especially theoretical studies of Dempster's rule. It will be seen below that from the commonality function both the support and the plausibility functions can be obtained and vice versa.

(ii) *Random Sets.* The mapping $\Gamma : \Omega \to 2^\Theta$ can be interpreted as a random set. The probabilities

$$m(H) = \sum \{p(\omega) : \Gamma(\omega) = H\}, \text{ for all } H \in 2^\Theta \quad (2.19)$$

define a probability distribution over 2^Θ. The probability measure m defines a random set in Θ and can be used instead of $(\Omega, P, \Gamma, \Theta)$ to represent evidence relative to a given question. It is easy to see that

$$\begin{aligned} sp(H) &= \sum \{m(B) : \emptyset \neq B \subseteq H\}/(1 - m(\emptyset)), \\ pl(H) &= \sum \{m(B) : B \cap H \neq \emptyset\}/(1 - m(\emptyset)). \end{aligned} \quad (2.20)$$

Theorem 2.6 (Shafer, 1976 a) Let $\mathcal{H}_i = (\Omega_i, P_i, \Gamma_i, \Theta), i = 1, \ldots, m$ be independent hints relative to the same question with commonality functions q_i, and let m denote the random set defined by the hint $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m$. Then

$$1 - m(\emptyset) = \sum \{(-1)^{|B|+1} \prod_{i=1}^{m} q_i(B) : \emptyset \neq B \subseteq H\}. \tag{2.21}$$

Furthermore, Dempster's rule of combination can also be expressed in terms of m. Indeed, let $m_i, i = 1, \ldots, r$, specify r random sets, describing r independent sources of information relative to the same question. Then Dempster's rule of combination leads to a random set described by m which is defined by

$$m(H) = \sum \{m_1(B_1) \cdots m_r(B_r) : B_1 \cap \ldots \cap B_r = H\}. \tag{2.22}$$

This corresponds to an intersection of the m random sets. The random set point of view has been discussed by Nguyen (1978), Goodman, Nguyen (1985), Hestir et al. (1991) in particular also in the framework of infinite frames Θ (see subsection 2.3 below).

(iii) *Propositional Logic with Uncertain Arguments.* A quite different approach to evidence theory has been proposed by Provan (1990 a), based on former work of Laskey, Lehner (1988), which noted a relation between evidence theory and the well known Assumption Based Truth Maintenance (ATMS) technique in AI (De Kleer, 1986). In this framework knowledge and information is described by a set of clauses $\Sigma = \{\xi_1, \ldots, \xi_r\}$ over some set of propositions N. A clause is a disjunction $x_1 \vee \ldots \vee x_t$ of literals and a literal is a proposition p or its negation $\neg p$. There is a subset of propositions $A = \{a_1, \ldots, a_s\}$ of N called assumptions, which are uncertain, but for which probabilities of truth $q_i = p(a_i)$ are given. Such assumptions serve for example to define uncertain rules or implications like $(p \wedge a_i \rightarrow q) = (\neg p \vee \neg a_i \vee q)$. If the assumption a_i holds, then the implication $p \rightarrow q$ is valid, otherwise this implication does not hold.

An interpretation of assumptions is an assignment of truth-values (0 or 1) to every assumption $a_1, \ldots, a_s$. Let's denote such an interpretation by x, which is a Boolean vector. Any interpretation x in $B_s = \{0,1\}^s$ has a determined probability (assuming independence)

$$p(x) = \prod \{q_i : x_i = 1\} \prod \{(1 - q_i) : x_i = 0\}. \tag{2.23}$$

An interpretation x defines also a conjunction $a(x) = y_1 \wedge \ldots \wedge y_s$ of literals of assumptions, where $y_i = a_i$, if $x_i = 1$ and $y_i = \neg a_i$, if $x_i = 0$. Given an interpretation x and the knowledge described by Σ, it may be that $a(x)$ and Σ are incompatible, which means that the system $a(x), \Sigma$ is not satisfiable. Such interpretations are called contradictory (relative to Σ), they are not really possible interpretations. Denote the set of contradictory interpretations by S_0.

A hypothesis can be specified by any propositional formula h containing propositions of N. For certain interpretations x it may be that h is a logical consequence of $a(x)$ and the knowledge Σ, which is expressed as $a(x), \Sigma \models h$. These are the interpretations which support the hypothesis h. Denote this set of interpretations by S_h. Note that by the usual conventions of propositional logic $S_0 \subseteq S_h$, so that the real supports of the hypothesis h are the interpretations in $S_h - S_0$. This set corresponds to $u(H)$ in (2.1) Then, in analogy to (2.3), we may define the degree of support of h induced by the knowledge Σ and the probabilities q_i of the assumptions as

$$sp(h) = (P(S_h) - P(S_0))/(1 - P(S_0)). \tag{2.24}$$

This framework renders accessible the powerful methods of propositional logic and also of reliability theory of binary systems (Barlow, Proschan, 1975, Kohlas, 1987 and 1991 a) to evidence theory. This leads also to alternative computational approaches to evidence theory (see section 4).

(iv) *Evidence Versus Partially Known Probabilities.* Dempster (1967) introduced the concept of compatible probability measures P such that $sp(H) \leq P(H) \leq pl(H)$ for every subset H of Θ. Ever since sp and pl have often been looked at as lower and upper bounds for some unknown probability distribution by many authors. This has had unfortunate effects because it led to intriguing misinterpretations of evidence theory. The question treated by evidence theory is whether H may be true or not, or more precisely, to what degree an evidence supports or discards a hypothesis H. It is quite another question to ask what could be the probability of a hypothesis H. In the latter case the unknown is not whether H is true or not, but the value of its probability. So, let us emphasize that although approaches using partially specified probabilities may be very useful in many circumstances, evidence theory as understood here, does not belong to this class of methods.

The difference may be examplified by an approach proposed by Fagin and Halpern (1989) (see also Jaffray, 1992 b). These authors remark that when specifying probabilities of some events in Θ, it may not always be possible to define probabilities on the whole algebra 2^Θ. The available information makes it only possible to specify probabilities P on a subalgebra $\mathcal{A}$ of 2^Θ. However the inner and outer probabilities P_* and P^* of P represent lower and upper limits for any extension P' of P on the whole algebra $2^\Theta : P_*(H) \leq P'(H) \leq P^*(H)$ for any subset H of Θ. This is a typical approach of specifying partially known probability. The fact that the inner measure P_* and the outer measure P^* have exactly the same properties as support and plausibility functions (Fagin, Halpern (1989, 1991) show that they satisfy the assertions in theorem 2.1, in particular (2.8) and (2.9) respectively) is rather a coincidence in this context (see however section 2.3 (iii) below). This becomes clear, if conditioning on an event $B \subseteq \Theta$ is considered. In the context considered here, if it becomes known that B holds, then all probability measures P' in the family $\mathcal{F}$ of all probability measures on 2^Θ defined by $P_*(H) \leq P'(H) \leq P^*(H)$ must be conditioned on B in the usual way of probability theory, such that the new bounds are

$$P_*(H|B) = inf\ \{P'(H|B) : P' \in \mathcal{F}\} = inf\ \{P'(H \cap B)/P'(B) : P' \in \mathcal{F}\},$$
$$P^*(H|B) = sup\ \{P'(H|B) : P' \in \mathcal{F}\} = sup\ \{P'(H \cap B)/P'(B) : P' \in \mathcal{F}\}.$$

Fagin, Halpern (1991) show that

$$\begin{aligned} P_*(H|B) &= P_*(H \cap B)/(P_*(H \cap B) + P_*(H^c \cap B)) \\ P^*(H|B) &= P^*(H \cap B)/(P^*(H \cap B) + P^*(H^c \cap B)) \end{aligned} \tag{2.25}$$

which is different from Dempster's rule of conditioning (see theorem 2.2). And this is no enigma, because P_* and P^* are not degrees of support and plausibility for H respectively, but lower and upper limits of the only partially known probability of H, which is quite another thing.

Before concluding this subsection, let's remark that there is still another interesting probabilistic model of evidence theory proposed by Ruspini (1987), which we cannot discuss here due to space limitations.

2.2 Nonprobabilistic approaches

The above approaches to the theory of evidence are all developed as applications of probability theory. But the theory of evidence can also stand by its own as an axiomatic mathematical theory without reference to probability theory. In this subsection, two such theories are presented. But, as Shafer (1992) says, "In order to use the theory, however, we need something more. At the very least, we need canonical examples with which to compare and calibrate actual evidence. Ideally, these canonical examples should motivate the axioms and rules ..." Whereas there may be canonical examples which do not rely on probability theory, most of them do. And here the theory is then again tied to probability theory.

(i) *Shafer's Theory of Belief Functions.* One way to construct evidence theory axiomatically is to postulate set functions which satisfy the basic properties of support functions as in theorem 2.1. Given a frame of discernment Θ for a precise question, postulate the existence of a numerical measure of the degree of belief $Bel(H)$ induced by the available information on every hypothesis $H \subseteq \Theta$. Impose then the following conditions as axioms upon the set function Bel (compare with theorem 2.1):

$$\begin{aligned}&(1) Bel(\emptyset) = 0\\ &(2) Bel(\Theta) = 1\\ &(3) Bel(H) \geq \sum \{(-1)^{|I|+1} Bel(\cap_{i \in I} H_i) : \emptyset \neq I \subseteq \{1, \ldots, n\}\}\end{aligned} \tag{2.26}$$

for all $n \geq 1$ and sets H, H_i in Θ such that $H \supseteq H_i$.

Bel is then called a *belief function*, and the name "Theory of Belief Functions" is sometimes used as a synonym to "Theory of Evidence" or "Dempster-Shafer Theory". This is the approach sometimes taken by Shafer (for example Shafer, 1976, 1978). From these axioms follows the following theorem:

Theorem 2.7 (see Shafer, 1976 a) If Bel is a belief function, then

$$m(H) = \sum \{(-1)^{|H-B|} Bel(B) : B \subseteq H\} \tag{2.27}$$

is a set function $m : 2^\Theta \to [0, 1]$ satisfying

$$\begin{aligned}&(1)\ m(\emptyset) = 0\\ &(2)\ m(H) \geq 0 \text{for all } H \subseteq \Theta\\ &(3)\ \sum \{m(H) : H \subseteq \Theta\} = 1\end{aligned} \tag{2.28}$$

and such that

$$Bel(H) = \sum \{m(B) : B \subseteq H\}, \text{ for all } H \subseteq \Theta. \tag{2.29}$$

Conversely, if a function $m : 2^\Theta \to [0, 1]$ satisfies (2.28), then Bel defined by (2.29) is a belief function. This function m reminds of course the random set model of evidence, with the exception that $m(\emptyset)$ is imposed to vanish. The set function m, satisfying (1) to (3) of (2.28) is called a *basic probability assignment* (bpa). It is interpreted as a distribution of a probability mass over the nonempty subsets of Θ, such that $m(B)$ is understood as "the measure of belief that is committed exactly to H" (Shafer, 1976 a). $Bel(H)$ is then the total belief committed to H, which is the sum of all beliefs exactly committed to its subsets (see (2.29)). Sets B for which $m(B) > 0$ are called *focal sets*.

The belief committed to H^c represents the doubt into H, $Dou(H) = Bel(H^c)$ and the less the doubt in H, the more plausible is H, $pl(H) = 1 - Dou(H) = 1 - Bel(H^c)$. The functions Bel, m

and pl are thus all linked together and the knowledge of one of them permits the reconstruction of the other two. Hence one may start with anyone of these functions, either Bel with axioms (2.26) or m with properties (2.28) or also pl with axioms similarly to Bel, except that (2.8) is replaced by (2.9). The function defined by $Bel(H) = 1 - pl(H^c)$ is then a belief function (Shafer, 1976 a). There is furthermore the commonality function

$$q(H) = \sum \{m(B) : B \supseteq H\} \tag{2.30}$$

which is linked to the other functions by the following theorem:

Theorem 2.8 (Shafer, 1976 a). Suppose that Bel is a belief function over Θ, pl is its plausibility function and q is its commonality function. Then for all nonempty subsets H of Θ:

$$\begin{aligned} Bel(H) &= \sum \{(-1)^{|B|} q(B) : B \subseteq H^c\}, \\ pl(H) &= \sum \{(-1)^{|B|+1} q(H) : \emptyset \neq B \subseteq H\}, \\ q(H) &= \sum \{(-1)^{|B|} Bel(B^c) : B \subseteq H\}, \\ &= \sum \{(-1)^{|B|+1} pl(B) : B \subseteq H\}. \end{aligned} \tag{2.31}$$

Dempster's rule of combination is in this framework best defined using the bpa. The definition can be modeled after (2.22), but taking into account that in the bpa $m(\emptyset) = 0$. Thus, if $m_i, i = 1, \ldots, r$ are r bpas for r independent belief functions, then the bpa of the combined belief function is given by

$$\begin{aligned} m(H) &= c \sum \{m_1(B_1) \cdots m_r(B_r) : B_1 \cap \ldots \cap B_r = H\} \text{ for every } H \neq \emptyset, \\ c^{-1} &= 1 - \sum \{m_1(B_1) \cdots m_r(B_r) : B_1 \cap \ldots \cap B_r = \emptyset\}. \end{aligned} \tag{2.32}$$

In contrast to the approaches in the previous subsection, the normalization is done in Dempster's rule of combination whereas above the normalization was done in the definition of support and plausibility. This is no essential difference. Thus (2.18) still holds for $H \neq \emptyset$, only the normalization constant c has to be added on its right hand side.

(ii) *Smets' transferable belief model.* Another axiomatic approach to belief functions has been introduced by Smets (see for example Smets (1988), Smets, Kennes (1990)). He starts with an allocation $m(H)$ of a total mass of belief 1 to subsets H of the frame of discernment Θ representing the possible answer to the question to be studied. This is a bpa, except that Smets admits that possibly some positive mass of belief goes to $\emptyset$. This represents the possibility that the true answer is not contained in Θ, but is outside Θ, which is called the open world assumption (in contrast to the closed world assumption considered so far). $Bel(H), pl(H), q(H)$ are defined as above, except that in (2.29) the sum extends only over $m(B), B \neq \emptyset$. The difference between open and closed world assumptions is not really essential (see however Smets, 1992). One may always add an element λ to Θ, which signifies "something else" and with $\Theta \cup \{\lambda\}$ we have a closed world. This may even be more prudent, because the empty set is subset of any frame, and if relative to one frame Θ_1 some belief is allocated to $\emptyset$, then this belief is automatically also a belief that something is outside Θ_2, which makes no sense.

The difference in the approach of the transferable belief model comes in the treatment of conditioning and combination. The model first defines *conditioning*: If it becomes known that the truth is in a subset B of Θ, then the allocation m must be changed into a new allocation m' in the following way:

(1) if $H \subseteq B$, then the evidence that the truth is in B does not modify the belief associated to $H : m'(H) = m(H)$

(2) if both $H \cap B \neq \emptyset$, and $H \cap B^c \neq \emptyset$, then the belief allocated to H must be transfered to $H \cap B$, because we learn that the truth is not in $H \cap B^c$

(3) If $H \subseteq B^c$, then the belief allocated to H must be transfered to λ because our initial belief in H now accounts for "something else" relative to B.

This leads to the following definition of the new allocation of belief

$$m'(H) = \sum \{m(H \cup C) : C \subseteq B^c\} \text{ for all } H \subseteq B. \tag{2.33}$$

This corresponds to Dempster's rule of conditioning. The rule of *combination* of two belief function Bel_1 and Bel_2 into a new belief function $Bel_{12} = Bel_1 \oplus Bel_2$ is now defined axiomatically using the rule of conditioning. Here are the first four axioms of Smets (1988):

(1) *compositionality:* Bel_{12} is a function of Bel_1 and Bel_2 only

(2) *symmetry:* $Bel_1 \oplus Bel_2 = Bel_2 \oplus Bel_1$

(3) *associativity:* $(Bel_1 \oplus Bel_2) \oplus Bel_3 = Bel_1 \oplus (Bel_2 \oplus Bel_3)$

(4) *conditioning:* If Bel_2 is such that $m_2(B) = 1$, then Bel_{12} is defined from m_{12} which is itself obtained from (2.33) by replacing m' by m_{12} and m by m_1.

Smets (1988) adds four more axioms, which are rather technical, and proves then (Smets, 1990) that these axioms imply the existence and unicity of Bel_{12} and also the product rule for commonality functions (2.18), hence Dempster's rule of combination (without normalization). For further ideas along similar lines see also Nguyen, Smets (1991) and Klawonn, Smets (1992).

2.3. Infinite frames

Only few efforts have been spent so far to study evidence theory with respect to infinite frames Θ. This problem is however very important both from a mathematical point of view as well as with respect to applications.

(i) *Shafer's allocation of probability:* The first step into this field has been made by Shafer himself (see Shafer, 1976 b, 1979). He starts with an axiomatic definition of belief functions similar as above (subsection 2.2. (i)). More precisely, if $\mathcal{E}$ is a multiplicative class of subsets of Θ (a class closed under finite intersections), then a function Bel on $\mathcal{E}$ is called a belief function if $\emptyset$ and Θ are in $\mathcal{E}$, $Bel(\emptyset) = 0, Bel(\Theta) = 1$ and Bel is monotone of order ∞ on $\mathcal{E}$ (see (2.8)). Now, theorem 2.7 is no more valid, there is no simple bpa as in the finite case. However there is something corresponding to it.

Given a belief function Bel on $\mathcal{E}$, Shafer constructs a set $\mathcal{M}$ of probability masses which has the structure of a complete Boolean algebra (see Halmos, 1969). A Boolean algebra is complete, if it is closed under arbitrary unions (disjunctions, joins) and intersections (conjunctions, meets), not only denombrable ones. Furthermore, a completely additive probability measure μ is defined on $\mathcal{M}$, which means that μ is additive over arbitrary collections of disjoint masses $\mathcal{B} \subseteq \mathcal{M}$. Now, Shafer (1979) proves, using an integral representation theorem of Choquet (1953), that there is a mapping $\rho : \mathcal{E} \to \mathcal{M}$ such that $\rho(H_1 \cap H_2) = \rho(H_1) \wedge \rho(H_2)$. The mapping ρ is called an allocation of probability and $\rho(H)$ is the total probability mass committed to H. Furthermore, for every $H \in \mathcal{E}$, we have $Bel(H) = \mu(\rho(H))$. In this way, $\mathcal{M}$ plays the role of the former bpa in the finite case and ρ corresponds to the mapping u introduced in subsection 2.1.

The definition of Dempster's rule of combination in this framework is not easy and straightforward. Shafer, in an unpublished paper (1978), developed a process of combination corresponding to Dempster's rule. Another approach to the definition of Dempster's rule is given in Shafer, 1976 b.

(ii) *Random Sets.* The framework of random sets provides for a straightforward generalization of the approach in the finite case. With infinite frames Θ, it will be no more possible to take as algebra the power set 2^Θ for the definition of m. Hestir et al. (1991) propose, following Matheron (1975), two particular algebras in order to define a probability distribution for the random set. Dempster's closed random intervals (Dempster, 1968 b) can also be seen in this framework of random sets.

It is straightforward to define Dempster's rule of combination as the intersection of random sets. The drawback of random sets however is that not all belief functions as defined by Shafer (1979, see (i) above) can be generated from random sets ! This will be explained in (iii) below. This is a first sign, that in the infinite case not all approaches are equivalent as in the finite case.

(iii) *Theory of Hints:* Dempster's multivalued mappings (subsection 2.1 (i)) may be generalized to infinite frames not only in the obvious way of random sets but also in a more subtle way. The point is that any possible interpretation of a hint determines a set of implied propositions, namely the supersets of $\Gamma(\omega)$) (if the truth is in $\Gamma(\omega)$, then it is necessarily also in every superset of $\Gamma(\omega)$). This is a family of sets $S(\omega)$ which has the following properties:

(1) $H \in S(\omega)$ and $H \subseteq H'$ imply $H' \in S(\omega)$,
(2) $H_1 \in S(\omega)$ and $H_2 \in S(\omega)$ imply $H_1 \cap H_2 \in S(\omega)$,
(3) $\Theta \in S(\omega)$, $\emptyset \notin S(\omega)$.

A collection of sets satisfying these three conditions is called a *filter.* Note that a filter is only closed under finite intersections, such that in the case of infinite Θ, the intersection $\cap \{H \in S(\omega)\}$ is not necessarily in $S(\omega)$. Thus, a filter is more general than simply the supersets of a set.

Now, a general hint $\mathcal{H}$ is constructed as follows. First, consider a possibly infinite set of possible interpretations Ω, which forms a probability space $(\Omega, \mathcal{A}, P)$. Each possible interpretation ω of Ω determines a filter $S(w)$ of implied propositions: if ω proves to be the correct interpretation, then each proposition in $S(\omega)$ is necessarily true. Defining $\Gamma(\omega) = S(\omega)$ allows to define a hint as a quintuple $\mathcal{H} = (\Omega, \mathcal{A}, P, \Gamma, \Theta)$. Note that this is a structure which is more general than a random set. The set $u(H) = \{\omega \in \Omega : H \in \Gamma(\omega)\}$ contains all the possible interpretations of $\mathcal{H}$ which imply the hypothesis H. If $u(H) \in \mathcal{A}$, then the degree of support $sp(H) = P(u(H))$ is defined. It can be shown, that the family $\mathcal{E}$ of subsets H of Θ, for which $u(H)$ is $\mathcal{A}$-measurable is a multiplicative class and that on this class sp is monotone of order ∞ (Kohlas, 1990). Furthermore, it is also true that for every function which is monotone of order ∞ on a multiplicative class $\mathcal{E}$, there is a hint, the so called canonical hint, which has this function as support function (Kohlas, 1990).

Because it is clear how to define the intersection of two filters, that is the family of subsets implied by two independent interpretations, it is straightforward to define Dempster's rule of combination as a generalization of the finite case (see Kohlas, 1990).

Even for a hypothesis H such that $u(H)$ is not $\mathcal{A}$-measurable, the degree of support can be defined. In fact, it is natural to define the degree of support of H by

$$sp_e(H) = \sup \{P(A) : A \subseteq u(H), A \in \mathcal{A}\} = P_*(u(H)). \tag{2.33}$$

The function sp_e turns out to be still monotone of order ∞ (Kohlas, 1990), a result which generalizes the theorem of Fagin, Halpern (1989) stating that the inner measure of a probability is monotone of order ∞. sp_e is an extension of the support function sp from the class $\mathcal{E}$ to the whole power set 2^{Θ}. Now, there may be several different hints, with the same support function on the class $\mathcal{E}$, but having different extensions sp_e of sp to 2^{Θ}. This remark is the key to a finer study of the structure of hints and support functions in infinite frames Θ (see Kohlas, Monney, 1991 a).

3. Applications of Evidence Theory

Do the structures introduced above correspond to anything interesting in our surrounding world? Does evidence theory have interesting, nontrivial and convincing applications? Misunderstandings of evidence theory led to disappointing results and as a consequence to a rejection of the theory by some authors. Nevertheless, there are today some domains where evidence theory begins to emerge as a natural approach which gives convincing results. Among these fields are statistical inference, diagnostics, risk analysis and decision analysis. These applications will be surveyed briefly in this section. But in addition to these applications a few others have to be mentioned. At the Stanford Research Institute International, evidence theory was studied very early from an application point of view, see Lowrance et al. 1986, Lowrance, 1988. This institute developed also a Lisp based general system (a shell) for evidential reasoning. Shafer himself applied evidence theory to auditing (Shafer, Srivastava, 1990). Other applications concern image processing (Wesley, 1986; Provan, 1990 b; Lohmann, 1991), geometric and temporal reasoning (Kohlas, Monney, 1991 b; Monney, 1991), including project scheduling and financial modelling (Kohlas, 1989). Evidence theory was of course considered as a general approach to uncertainty management in expert systems or knowledge bases. A recent architecture of knowledge bases incorporating evidence theory has been developed by Saffiotti (1991 a and b); he addresses in particular the important problem of using an appropriate language like first order logic as a base for the automatic construction of a propositional or extended mathematical model of evidence.

3.1. Statistical Inference from an Evidential Point of View

Dempster's motivation to introduce multivalued mappings and the related lower and upper probabilities was statistical inference (Dempster, 1966, 1967, 1968, 1990). The goal of this new approach was to generalize Bayesian inference and also to reconcile the latter with Fisher's fiducial probabilities. Dempster was able to derive lower and upper probabilities on the unknown parameters to be studied by statistical inference by specifying a multivalued mapping and without using an a priori distribution on the parameter space. If an a priori distribution is known, then Dempster's method can incorporate it and leads to the classical Bayesian approach, but his method works also, if the a priori information about the parameter is less specific than a distribution, if it is only described by a belief function.

This approach will be sketched here in a way proposed by Kohlas (1992), which shows that the lower and upper probabilities of Dempster can in fact be seen as degrees of support and plausibility of hypotheses about the unknown parameter derived from observations and a process model specifying how observations are generated. Let Π be the set of possible values of an unknown parameter π. Inference about this parameter is made by means of an observation process which yields an observation x in some obervation space X. It is assumed that the actual observation x depends only on the unknown parameter π and a further unknown chance element ω which belongs to a known probability space $(\Omega, \mathcal{A}, P)$. This is expressed by a known function

$$x = f(\pi, \omega) \tag{3.1}$$

Equation (3.1) is called a process model of observation. Now, if the experiment is carried out, then an observation $x \in X$ becomes available. Given such an observation x and the process model (3.1), what can be said about the unknown parameter π?

Define $\Gamma_x(\omega) = \{\pi \in \Pi : x = f(\pi, \omega)\}$, the set of all possible parameter values if x is the given observation, and ω the random element which generated the observation. Then $\mathcal{H}_x = (\Omega, \mathcal{A}, P, \Gamma_x, \Pi)$ is a hint about the unknown parameter value π (see section 2.1). Hence, if $H \subseteq \Pi$ is any hypothesis about the unknown parameter, then $u_x(H) = \{\omega \in \Omega : \emptyset \neq \Gamma_x(\omega) \subseteq H\}$ represents all random elements ω for which the hypothesis is necessarily true and $v_x(H) = \{\omega \in \Omega : \Gamma_x(\omega) \cap H \neq \emptyset\}$ the random elements for which the hypothesis is possibly true. Thus, $sp_x(H) = P(u_x(H))/P(v_x(\Pi))$ and $pl_x(H) = P(v_x(H))/P(v_x(\Pi))$ are the degrees of support and plausibility of H given the observation x (assuming that $u_x(H), v_x(H)$ are measurable, otherwise the inner or outer measures of P have to be considered, see subsection 2.3). Note that $\Gamma_x(\omega)$ may well be empty. But such an ω cannot possibly have generated the observation x, because there is no parameter π which in conjunction with ω generates x. Thus, the conditioning in sp_x and pl_x is well founded. Application of this approach is possible for so called measurement or also structural models (Frazer, 1968; Kohlas, 1992). But also sampling problems may be formulated and analyzed this way (Dempster, 1967 b, 1968 a and b, 1990; Shafer, 1982 a; Kohlas, 1992).

A further reference for the application of belief functions in statistical inference is given by Wasserman (1990); this paper takes however not the evidential point of view, but uses belief functions as a means to describe convex sets of a priori probability measures for a Bayesian analysis. In the cited paper, other references to similar uses of lower and upper probability in statistical inference may be found.

3.2 Diagnostics and Risk Analysis

Diagnostics is the task to infer explanations for a set of observations. In many cases these observations are symptoms of diseases or faults of a medical or technical system. The explanations furnish then the possible diseases causing the observed facts. Symptoms can be considered as evidence for certain possible diseases and evidence theory is therefore an obvious approach to diagnostics.

An early example of a diagnostic expert system was MYCIN. The authors of this system already noted that symptoms are not infallible evidence but may have exceptions. Also they stated that symptoms are evidences which bear on sets of hypotheses rather than on individual hypotheses. They concluded that probability theory was not an appropriate tool for their purposes and they developed their own ad hoc uncertainty calculus. Later, it became clear that evidence theory would have been well adapted to the structure of MYCIN (Gordon, Shortliffe, 1985). Applications of evidence theory to diagnosis have been discussed by Smets (1979, 1981). GERTIS is a newer expert system which is of a similar type (Yen, 1989), although it mixes evidential theory and Bayesian analysis.

In the mean time, a quite different architecture of diagnostic systems emerged, the model based inference systems. Here a model of the properly functioning system serves as a basis to predict output values from given input values. Differences between these predicted values and the observed ones are then used to derive possible diagnoses (a fundamental paper in this domain of logic based diagnosis is Reiter, 1987). These approaches are essentially symbolic and provide no means to rank the possible diagnoses according to their likelihood or to judge the likelihood that a particular component of a system is faulty. It will be explained how these models can be extended in a very natural way by evidence theory to include likelihood measures. Some of these models may not only be used for diagnostic purposes, but for predictive risk analysis as well.

(i) *Model Based Diagnostics.* A model of a properly functioning system is described by a set of variables and relations between them. Let $M = \{1, \ldots, m\}$ be an index set, X_i with $i \in M$ a set of variables taking values in the frames Θ_i. Let R be a family of subsets of M and for every $J \in R$ let $R_J \subseteq \prod_{i \in J} \Theta_i = \Theta_J$ be a given relation between the variables X_i for $i \in J$. These relations correspond to physical devices which realize those relations as input-output relations (where some variables are considered as inputs and the others as outputs). As an example, the X_i may be integer variables, taking values in $\mathcal{N}$, and the relations are arithmetic relations like additions $X_i + X_j = X_k$ or multiplications $X_i \times X_j = X_k$. Or else the X_i are Boolean variables with values $\{0, 1\}$ and the relations are Boolean relations as they arise for example in digital circuits (see Figure 1 for an example).

If the values of some variables are observed, then such a model permits to test whether the observations are consistent with the assumed relations between the variables. If this is not the case, then some relations must be violated and the corresponding devices must be faulty. For example, in Figure 1, for the input values $A = 1$ and $B = 0$, one would expect the output values $S = 1$ and $C = 0$. However, if for S the value 0 is actually observed, then at least one the devices d_1, d_2, d_3 must be faulty. However this analysis does not permit to rank these diagnoses according to their likelihood.

Now fault models of the components can be introduced. Let $\Omega_J = \{\omega_0, \omega_1, \ldots, \omega_s\}$ denote a set of different mutually exclusive working modes of the component associated to the relation $J \subseteq M$. It is supposed that the probabilities $p_J(\omega_i)$ of these different modes are known. ω_0 is assumed to correspond to a properly functioning component, whereas ω_1 to ω_s represent different faulty modes. To any mode ω_i corresponds a relation $\Gamma_J(\omega_i) \subseteq \Theta_J$; $\Gamma_J(\omega_0) = R_J$ is the above relation of a properly functioning device. An and-gate as used in a digital circuit may for example have two fault modes, one where the output is 0 for every input configuration and one where the output is entirely unpredictable for each input. The later mode corresponds to the vacuous relation $\{0, 1\} \times \{0, 1\} \times \{0, 1\}$, all combinations of input and output are possible. This associates clearly a hint $(\Omega_J, p_J, \Gamma_J, \Theta_J)$ to every component in the system. The observation of certain variables X_i provides further deterministic and precise hints relative to Θ_i.

These hints refer to different but related frames. This situation differs from those considered so far and especially in section 2, where hints refered always to the same frame of discernment. The situation encountered here is however typical for most applications of evidence theory. Dempster's rule for the combination of hints can be extended to such situations as will be explained in section 4. Hence these hints may be combined to obtain degrees of support and plausibility for different diagnoses, that is for hypotheses about the possible faulty configurations of modes which may explain the observed values of variables. The computational methods for this analysis are described in section 4. This shows that a well known model in diagnosis theory can very naturally be extended using evidence theory so as to incorporate information about the likelihood of possible diagnoses.

Such systems may also often be conveniently formulated in the framework of logic with uncertain arguments (see subsection 2.1 (iii)) and related to truth maintenance systems (Provan, 1988). A further application of the model introduced above is to the detection of bad data in measurement models (Kohlas, 1992).

(ii) *Risk Analysis.* Boolean systems as described above arise not only in modeling digital circuits, but also in fault or event trees as used for reliability or risk analysis. These structures can be used for diagnostic purposes, if some events are observed exactly in the way described above. They are however also used for predictive studies of risks. Almond (1991, 1992 a) discusses the statistical analysis of failure rates of components, using methods similar as those described in subsection 3.1. This leads to only partially known component models, in the sense that only belief functions for the unknown parameters can be given. This in turn leads to belief functions for the failure of components and thus for the events in the fault tree. This is an example of how evidence theory may be applied in risk analysis.

3.3 Decision Analysis

(i) *Generalized Decision Principles.* Analysis of uncertain perspectives and data is an important activity in decision support. One aspect is the problem of choice. In the standard choice model, a set $A = \{a_1, \ldots, a_m\}$ of possible activities and a set $S = \{s_1, \ldots, s_n\}$ of possible states of the world is given, and it is assumed that for every activity a_i and state s_j a utility index u_{ij} is defined which expresses the preferences of the decision maker. Classical decision theory considers two extreme situations relative to the knowledge of the state of the world. Either there is full ignorance about the state; in this case decision principles like the Hurwicz principle, Laplace's principle of insufficient reason or the principle of minimal regret are proposed to select an activity. Or else a probability distribution $p(s_j)$ over the set S is assumed to be given. Then the activity is selected by maximizing the expected value of the utility of an action a_i

$$E_i(u_{ij}) = \sum_{j=1}^{n} u_{ij} p(s_j). \tag{3.2}$$

This method can be justified by certain axioms of rationality (von Neumann, Morgenstern, 1944; Fishburn, 1970).

In our terminology, the first case corresponds to a vacuous hint about the state of the world and the second to a precise hint about the state. So it is natural to look at intermediate cases, where an arbitrary hint $\mathcal{H} = (\Omega, p, \Gamma, S)$ about the state is given. If a fixed interpretation ω is known to be the correct one, then the true state is known to be in the set $\Gamma(\omega)$. With respect to this set of states we have a decision problem with full ignorance and we may apply any one of the decision principles to evaluate the actions. We may then take the expected value over these evaluations for every interpretation ω in Ω. Each decision principle leads in this way to a *generalized decision principle.*

First of all, note that an action a_i is dominated by an action a_k, if for all ω max $\{u_{ij} : s_j \in \Gamma(\omega)\} \leq$ min $\{u_{kj} : s_j \in \Gamma(\omega)\}$. Dominated actions may be eliminated from further considerations.

Laplace' principle of insufficient reason amounts to assume a uniform distribution over the states in $\Gamma(\omega)$ for every ω, taking the expected value of the utilities relative to these uniform distributions and then taking the expected value over Ω:

$$\sum_{\omega \in \Omega} p(\omega) \sum \{u_{ij}/|\Gamma(\omega)| : s_j \in \Gamma(\omega)\}, \text{ for all } a_i \in A. \tag{3.3}$$

This generalized principle of insufficient reason has been proposed by Dubois, Prade (1982), Williams (1982), Smets (1989), and Smets, Kennes (1990).

Hurwicz's principle uses an index of optimism α between 0 and 1 to weight the pessimistic minimal utility value and the optimistic maximal utility value over the possible states. The generalized Hurwicz' principle considers thus the following index for every action a_i:

$$\sum_{\omega \in \Omega} p(\omega)(\alpha \max \{u_{ij} : s_j \in \Gamma(\omega)\} + (1-\alpha) \min \{u_{ij} : s_j \in \Gamma(\omega)\}). \tag{3.4}$$

This approach has been proposed by Strat (1990). It can also be seen as a linear interpolation between lower and upper bounds of expectation intervals defined by the hint. Strat shows also how to use this principle in the context of generalized decision trees. For $\alpha = 1$ the principle reduces to the generalized maximax principle and for $\alpha = 0$ to the generalized maximin principle.

The regret principle computes for every state s_j and action a_i the regret $r_{ij} = \max \{u_{ij} : i = 1, \ldots, m\} - u_{ij}$. It applies then the generalized minimax principle to the matrix (r_{ij}). This procedure has not been yet proposed in the literature, but it is in the line of the above principles.

(ii) *Expected Utility Theory.* Let us generalize the basic decision model introduced above. Instead of describing the outcome of an action a_i and a state s_j directly by an numerical utility index u_{ij}, suppose that the outcome is described by some more basic, possibly nonnumeric element $\theta_{ij} = \theta(a_i, s_j) \in \Theta$. The hint $\mathcal{H}$ on the set of possible states s is transformed for each action a_i by the function $\theta(a_i, \cdot) : S \rightarrow \Theta$ into a hint $\mathcal{H}_i$ on Θ. The problem of choice of an action a_i reduces then to the choice of a hint on Θ or to a comparison of hints or belief functions on Θ. Now, a hint on Θ can be seen as probability distribution over the subsets of Θ. Let $\mathcal{P}_\Theta$ denote the set of probability distributions over 2^Θ, where Θ is assumed to be finite. Then the following theorem holds:

Theorem 3.1 (see for example Fishburn, 1970, theorem 8.2) Let $\prec$ denote a binary relation on $\mathcal{P}_\Theta$. Then there is a real-valued function u (called a utility function) on 2^Θ that satisfies

$$P \prec Q \text{ iff } \sum_{B \subseteq \Theta} P(B)u(B) < \sum_{B \subseteq \Theta} Q(B)u(B) \tag{3.5}$$

for all P, Q in $\mathcal{P}_\Theta$

if and only if, for all P, Q, R in $\mathcal{P}_\Theta$

(1) $\prec$ on $\mathcal{P}_\Theta$ is a weak order,
(2) (Independence). $P \prec Q$ implies $\alpha P + (1-\alpha)R < \alpha Q + (1-\alpha)R$ for all $0 < \alpha < 1$,
(3) (Contuinity). $P \prec Q, Q \prec R$ implies $\alpha P + (1-\alpha)R < Q$ and $Q < \beta P + (1-\beta)R$ for some $\alpha, \beta \in (0,1)$.

Moreover, u is unique up to a positive linear transformation $au + b$.

Inequality (3.9) states that under the conditions of the theorem the action can be selected by maximizing an expected utility. This has been noted by Jaffray (1989). (3.9) may however not be very practical given the large numbers of subsets of Θ for which utilities $u(B)$ must be determined. Jaffray (1989) adds a further condition to (1) to (3) above in theorem 3.1, which reduces (3.9) essentially to the generalized Hurwicz principle. Jaffray, Wakker (1992) relate the above result to Savage's sure thing principle and show how a weaker version of this principle leads to the above result. Jaffray (1992 a) then considers this expected utility approach also in a dynamic setting.

(iii) *Support for Actions.* Another approach, but very much in the spirit of evidence theory, is to look for arguments for or against the hypothesis that a given action is satisfactory (achieves the specified goals), or that it is better than another one, or even that it is the best among a given set of actions. Hints and belief functions about the consequences of an action may be used for this purpose. According to Strat (1990) such a "constructive" approach to decision theory has been proposed by Shafer (1982) in an unpublished paper. This approach seems not to have been pursued in the literature (see however Schaller, 1991). It would be close in spirit to the outranking methods of multicriteria analysis (see for example Roy, 1985). Again, it should be emphasized that evidential decision analysis as understood here is different from decision analysis with partially known probabilities.

4. Computational Aspects of Evidence Theory

The practical implementation of evidence theory sets several difficult computational problems. How can the available evidence be best encoded for representation on a computer and what are the corresponding memory requirements? Which are the best algorithms to compute degrees of support and plausibility or to combine pieces of evidence by Dempster's rule and what are their complexities? Are there special structures of evidence which facilitate representation and computation? These are the questions which are briefly addressed in this section. References to computer packages which implement evidence theory will also be given.

4.1 Representation of evidence

In practice, evidence is given in pieces represented by different hints $\mathcal{H}_1, \ldots, \mathcal{H}_n$ and the total, combined evidence is obtained by combining these hints by Dempster's rule $\mathcal{H} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_n$. A first implementation decision concerns the question whether the evidence should be represented and stored decomposed in pieces $\mathcal{H}_1, \ldots, \mathcal{H}_n$ or rather in the "compiled" form $\mathcal{H}$. In the first case any query about the support or plausibility of a hypothesis requires the necessary combination calculations, in the latter case such a query corresponds to a table lookup. However the first variant needs in many cases far less memory than the latter.

In any case, the question arises how to represent hints or equivalent forms of evidence. To represent a hint, its elements, i.e. the set of possible interpretations, their probabilities, the corresponding multivalued mapping and the frame of discernment Θ must be represented. This is an extensive representation of evidence. More compact, but slightly less informative, would be a representation of evidence by the corresponding basic probability assignment (bpa) $m(B)$ for all focal sets $B \subseteq \Theta$. In the worst case there are $2^{|\Theta|} - 1$ focal sets (i.e. every non empty subset of Θ may be a focal set). Already for $|\Theta|$ greater that say 30 this will become impracticable. In practice however the basic pieces of evidence $\mathcal{H}_i$ will each have only very few focal sets; then the bpa will be a convenient representation of evidence. However in many cases $\mathcal{H}$ will already have much more focal sets than each $\mathcal{H}_i$ individually. This is an argument against a "compiled" representation of evidence. In view of theorem 2.8 other possible representations of evidence would be in terms of the support-, plausibility- or commonality functions. These representations are however impracticable because they all need in most cases $2^{|\Theta|} - 1$ sets with their corresponding numerical evaluation.

It will be shown below, that there are important special structures which make the representation of the pieces of evidence $\mathcal{H}_i$ more simple. Thus, a decomposed representation of evidence becomes in most practical situations quite feasible and the main problem is then to find efficient algorithms for the combination of the hints.

4.2. Conversion algorithms and Dempster's Rule

Whatever representation in terms of bpa m, support (or belief) function sp (Bel), plausibility function pl or commonality function q is chosen, there is a need for conversion algorithms from one form into another. Each set function m, sp, pl or q can be represented by a vector in the space $\mathcal{R}^s$, where $s = 2^{|\Theta|}$. The transformation of m into Bel (or sp) as defined by (2.29) is called a Möbius transformation, which is an application of $\mathcal{R}^s$ into itself. Kennes, Smets (1991 a and b) and Thoma (1991) independently have devised the fast Möbius transform. It transforms a vector of $\mathcal{R}^s$ in $|\Theta|$ steps as follows: Arrange the element of Θ in an arbitrary order $\theta_1, \ldots, \theta_n$, $n = |\Theta|$ and then compute for $i = 1, \ldots, n$

$$m_i(A) = \begin{cases} m_{i-1}(A) + m_{i-1}(A - \{\theta_i\}), & \text{if } \theta_i \in A, \\ m_{i-1}(A), & \text{otherwise.} \end{cases} \tag{4.1}$$

If $m_0 = m$, then it can be shown that $m_n = sp$ (see for example Thoma, 1991). Similar algorithms may be derived for the conversion between all other pairs of set functions m, sp, pl and q (Kennes, 1992). The practical limit of these algorithms is again given by the size $2^{|\Theta|}$ of the vectors in $\mathcal{R}^s$. If the support is needed only for some subsets of Θ, but not for all of them, then coarsening the frame Θ may help (see Shafer, 1976 a for this notion).

The conversion algorithms need only be applied after evidences have been combined by Dempster's rule. Orponen (1990) studied the computational complexity of this rule. His results showed that Demspter's rule is hopelessly complex in the worst case. Practical cases are fortunately not necessarily worst cases. They exhibit often special structures which may be exploited to reduce complexity and increase the efficiency of algorithms. This is discussed below.

Kennes, Smets (1991 b) propose to use fast Möbius transform to transform the bpas m_1 and m_2 of two evidences into their respective commonality functions q_1 and q_2 and then use theorem 2.5 to get the commonality function of the combined evidence and then to use again a fast Möbius transform to get the other set functions relative to the combined evidence. This is an alternative to the use of (2.22). However the commonality function tends to have much more non null values than the bpa and needs therefore much more storage, such that it is not clear whether this approach is preferable.

4.3 Graphical Models

In practical applications of evidence theory the reasoning concerns a group of distinct variables $X_1, \ldots, X_m$, each one with its own domain (or frame of discernment) Θ_i. Then arbitrary groupings of variables $\{X_j : j \in J \subseteq M$ (with $M = \{1, \ldots, m\}$) can be considered and such a group of variables has the frame $\Theta_J = \prod_{j \in J} \Theta_j$, the product space of the frames of the variables in the group. Denote Θ_M by Θ; this is the overall frame of the model to be considered. The available pieces of evidence very often do not concern all variables together, but only some groups of them. Such pieces of evidence concerning a group J of variables may be described by the hints $\mathcal{H}_i, i = 1, \ldots, r$, relative to the frame Θ_J. More precisely, let J_i denote the group of variables of the hint $\mathcal{H}_i$. As an example, consider model based diagnostics described in section 3.2 (ii) above.

As these different hints refer to different frames, they can not be directly combined by Dempster's rule. A hint $\mathcal{H} = (\Omega, p, \Gamma, \Theta_J)$ relative to the group of variables J can be extended to a group $K \supset J$ simply by replacing the focal sets $\Gamma(\omega)$ by their cylindrical extension $\Gamma(\omega) \uparrow K = \Gamma(\omega) \times \Theta_{K-J}$. The new hint $\mathcal{H} \uparrow K = (\Omega, p, \Gamma \uparrow K, \Theta_K)$ is called the *vacuous extension* of $\mathcal{H}$ from J to K. Then the hints relative to different groups of variables may all be vacuously extended to M and then combined on the common frame Θ,

$$\mathcal{H} = (\mathcal{H}_1 \uparrow M) \oplus \ldots \oplus (\mathcal{H}_r \uparrow M) \tag{4.2}$$

In general, let us define Dempster's rule by $\mathcal{H}_1 \oplus \mathcal{H}_2 = (\mathcal{H}_1 \uparrow M) \oplus (\mathcal{H}_2 \uparrow M)$. A representation of $\mathcal{H}$ by its pieces $\mathcal{H}_i$ is clearly advantageous from a memory point of view. Not only has each $\mathcal{H}_i$ in general much less focal sets than $\mathcal{H}$, but also their focal sets are generally in a much smaller dimension $|J_i|$ than $|M|$.

It seems however at first sight that this last advantage is lost, when the hints are combined, because all of them must be extended to M. This however can be avoided in many cases by the ability to combine hints locally. Trivially, if two hints refer to the same group J of variables, then they may be combined on Θ_J beforehand rather than on Θ. Also if one hint refers to a group J, the other to a group K and $J \subseteq K$, then the two hints can be combined on Θ_K. If all these combinations are executed beforehand, then a family $\mathcal{J} = \{J_1, \ldots, J_r\}$ of incomparable groups of variables remain (i.e. all groups are distinct and none is contained in another one). $\mathcal{J}$ is called a scheme (Thoma, 1991) and it defines a *hypergraph* (with hyperedges J_i).

Now, even beyond the trivial cases above, *local combinations* are possible, due to a basic theorem given below. If $\mathcal{H} = (\Omega, p, \Gamma, \Theta_J)$ is a hint relative to a group of variables J, and K is a subset of J, then $\mathcal{H} \downarrow K = (\Omega, p, \Gamma \downarrow K, \Theta_K)$ is the restriction or projection of $\mathcal{H}$ from J to K, if $\Gamma(\omega) \downarrow K$ is the projection of $\Gamma(\omega)$ from Θ_J to Θ_K. Contrary to the vacuous extension, where information is neither added nor lost (i.e. $(\mathcal{H} \uparrow K) \downarrow J = \mathcal{H}$ if $J \subseteq K$), projecting hints causes in general a loss of information, that is $(\mathcal{H} \downarrow K) \uparrow J \neq \mathcal{H}$, if $K \subseteq J$. Vacuous extension, followed by projection permit to transport a hint relative to a group of variables J to any other group K, $\mathcal{H}|K = (\mathcal{H} \uparrow J \cup K) \downarrow K$. Then the following theorems of local computations hold:

Theorem 4.1 (see for example Shafer et al. 1987). If $\mathcal{H}_1$ and $\mathcal{H}_2$ are hints relative to the groups of variables H and K respectively and $H \cap K \subseteq J$, then

$$(\mathcal{H}_1 \oplus \mathcal{H}_2)|J = (\mathcal{H}_1|J) \oplus (\mathcal{H}_2|J). \tag{4.3}$$

Theorem 4.2 (Shafer et al. 1987). If $\mathcal{H}$ is a hint relative to a group of variables H, and K, J denote groups of variables such that $H \cap K \subseteq J$, then

$$\mathcal{H}|K = (\mathcal{H}|J)|K. \tag{4.4}$$

This last theorem shows that transporting a hint from a group of variables H to another group K can be done simpler by projecting the hint first to $H \cap K$ and then extending it to K, rather than first extending it to $H \cup K$ and then projecting it to K. Theorem 4.1 shows that in some cases hints may be combined locally on certain smaller frames Θ_J relative to a group of variables J rather than on the overall frame Θ.

If $\mathcal{J}$ is a hypergraph, then it may be possible to arrange its hyperedges J_i into a family $\mathcal{E}$ of pairs $\{J_i, J_k)\}$such that

(1) $(\mathcal{J}, \mathcal{E})$ is a tree,
(2) if J_i and J_k are distinct vertices of the tree $(\mathcal{J}, \mathcal{E})$, then $J_i \cap J_k$ is contained in every vertex on the unique path from J_i to J_k in the tree $(\mathcal{J}, \mathcal{E})$.

Such a tree is called a *Markov tree* (also called a join tree in data base theory, see Maier, 1983). If such a tree exists, then $\mathcal{J}$ is called a hypertree or an acyclic hypergraph. If one now considers any vertex J of a Markov tree $T = (\mathcal{J}, \mathcal{E})$, then it has a set $N(J)$ of neighboring vertices and if node J is eliminated then T decomposes into a set of subtrees $T(J_k)$, each one associated with a vertex $J_k \in N(J)$. It can be shown that every subtree of a Markov tree is still a Markov tree. If $T = (\mathcal{J}, \mathcal{E})$ is a Markov tree and $\mathcal{J}$ is a scheme, then there is a hint $\mathcal{H}_J$ associated with every vertex J of T. Let $\mathcal{H}(T)$ denote $\oplus \{\mathcal{H}_J : J \in \mathcal{H}\}$. Property (2) of a Markov tree permits the application of both theorems 4.1 and 4.2 to give the following result:

Theorem 4.3 (Shafer et al. 1987). If T is a Markov tree and J is a vertex of T, then

$$\mathcal{H}(T)|J = \mathcal{H}_J \oplus [\oplus \{(\mathcal{H}(T(J_k))|J_k)|J : J_k \in N(J)\}]. \tag{4.5}$$

This is a recursive formula for combining hints on Markov trees, which can be used to compute projections of the combined hint to groups of variables in the hypergraph $\mathcal{J}$, using local combinations only.

Theorem 4.3 is a basic result for the computational aspects of evidence theory. In fact it is a special case of a much more general abstract framework (Shenoy, Shafer, 1990; Shafer, 1991) which covers such diverse models as nonserial dynamic programming (Bertele, Brioschi, 1972), probabilistic (Bayes or Markov) networks (Lauritzen, Spiegelhalter, 1988), constraint propagation, influence diagrams (Oliver, Smith, 1990) and others.

Sometimes the hypergraph has an evident Markov tree structure. Such is the case for example for models of dynamical processes (Kohlas, 1991 c). However, it is not always possible to construct a Markov tree for a given hypergraph $\mathcal{J}$. In such a case, the hypergraph must be enlarged by augmenting the hyperedges until a Markov tree can be obtained; this is called constructing a hypertree cover of the hypergraph. Although it is not difficult to find a hypertree cover of a hypergraph, it is difficult to find a good one, that is one with hyperedges of small cardinality. There is a growing literature on this subject; see for example Rose, 1970, Bertele and Brioschi, 1972, Tarjan and Yannakakis, 1984, Kong, 1986, Arnborg et al., 1987, Mellouli, 1987, Zhang, 1988 and Almond, Kong, 1991.

Meanwhile, several software packages based on this framework have been developed. Examples are DELIEF (Zarley et al. 1988), MacEvidence (Hsia, Shenoy, 1989), PULCINELLA (Safiotti, Umkehrer, 1991 and 1992), BELIEF (Almond, 1992 b, c) and TRESBEL (Xu, 1992).

4.4. Logical models

Propositional knowledge based on assumptions (see subsection (iii) in 2.1) can be considered in the framework of the graphical models introduced above. Their special structure permits however the application special methods. One of them propagates degrees of support in inference nets, applying the method of factorization well known from reliability theory, where necessary (Cardona et. al., 1991, 1992). A software package implementing these ideas is available (Cardona, 1993).

A variant of this method would be to Monte Carlo sample assumptions and the use automatic proof procedures (e.g. resolution) to see whether the assumptions are contradictory or prove the hypothesis. This would permit to estimate the degree of support of the hypothesis.

Finally, an alternative is to proceed in two steps: In the first step, the contradictory configurations of assumptions and the configurations of assumptions supporting a given hypothesis are determined. For this purpose techniques from Assumption Bases Truth Maintenance (ATMS) can be used (see DeKleer, 1986, Reiter, DeKleer, 1987, Siegel, 1987 and especially Inoue, 1992). In the second stage the degree of support of the hypothesis is calculated as the conditional probability of the supports of the hypothesis using (2.23). This can be done by applying methods from combinatorial reliability theory. The mathematical foundations of this approach are given in Kohlas, Monney (1993).

5. Conclusion and Outlook

Evidence theory is still young. There remain many open questions. These relate to its theoretical foundations, its practical use in modeling and its computational complexity.

On the level of the theory, there is surely a large field of mathematical research open for studying evidence, hints, support or belief functions, plausibility functions or whatever with respect to non finite spaces. In particular, the study of evidence relative to topological spaces should be rewarding. Also evidence theory appears as a theory of random sets (or rather random filters). These concepts represent clearly a generalization of the notion of random variables. Now, classical probability theory can largely be considered as a theory of random variables. In what exact sense then can evidence theory be considered as a generalization of probability theory. Note that Dempster's rule as such does not appear in classical probability theory. What is then role of Dempster's rule in probability theory? There are also relations between evidence theory and possibility theory or fuzzy logic (Klir, Folger, 1988; Dubois, Prade, 1985; Shafer, 1987). How do these theories relate ?. And a similar question can be raised concerning the relations between evidence theory and logic. Beside the relation shown above in the domain of propositional systems with probable assumptions, there are relations to default logic, modal logic and other non-standard logics (see for example Deutsch-McLeish, 1991).

It seems that evidential information is essentially *relational*. On the other hand in knowledge engineering the importance of *conditional* information (if something holds, then ...) is stressed. This is well respected for example in Bayesian networks where the knowledge is essentially encoded in conditional probability distributions. Does this aspect really restrict the applicability of evidence theory? The evidential approach to statistical analysis shows that in fact evidence theory needs somehow more detailed information about the observation process than Bayesian statistics. On the other hand it can be dispensed of a priori distributions. This question needs more study.

From the point of view of computation, there is clearly a need for efficient approximations and heuristics for evidential calculations. In order to study a question, to judge a hypothesis, probably not all available evidence must be considered. There may be some far-fetched evidence whose relevance to the question is only marginal. So *focusing* the reasoning on the relevant evidence will in many cases be an efficient strategy to limit the computational effort to reasonable size. Furthermore evidence may be simplified if only coarse judgments are required. Notions of coarsening of evidence have been introduced by Dubois, Prade, 1986; Yager, 1985 and generalized by Kohlas, Monney, 1991 a, see also Vakili, 1993.

Evidence theory adds new aspects to the management of uncertainty, aspects which are not covered by related methods in the field, like probability theory, fuzzy logic or various kinds of nonmonotonic logics. Therefore, in conclusion, it can be said that evidence theory is an interesting and promising new method for managing uncertainty and is a valuable addition to the toolbox of methods in this field.

Acknowledgement

The authors thank Philippe Smets for his helpful comments.

Bibliography

ARNBORG, S.; CORNEIL, D.G.; PROSKUROWSKI, A. (1987): Complexity of Finding Embeddings in a k-Tree, *SIAM J. Algebraic and Discrete Methods,* **8**, 277-284.

ALMOND, R. (1991): Belief Function Models For Simple Series and Parallel Systems. Tech. Report 207, Department of Statistics, Univ. of Washington

ALMOND (1992 a): Models for Incomplete Failure Data. Statistical Sciences, Inc., Seattle, WA.

ALMOND, R. (1992 b): Using the Belief Package. Harvard Univ., Dept. of Statatistics.

ALMOND, R. (1992 c): Graphical-Belief: Project Overview. Statistical Sciences, Inc., Seattle, WA.

ALMOND, R.; KONG, A. (1991): Optimality Issues in Constructing Markov Tree froem Graphical Models. Res. Rep. A-3, Harvard Univ., Dept. of Stat.

BARLOW, R.E.; PROSCHAN, F. (1975): *Statistical Theory of Reliability and Life testing.* New York.

BERTELE,U.; BRIOSCHI, F. (1972): *Nonserial Dynamic Programming.* Academic Press.

BONISSONE, P.P; HENRION, M.; KANAL, L.N.; LEMMER, J.F. (eds.) (1991): *Uncertainty in Artificial Intelligence 6.* North Holland, Amsterdam.

BOUCHON-MEUNIER, B.; YAGER, R.R.; ZADEH, L.A. (eds.) (1991): *Uncertainty in Knowledge Bases.* Springer Lecture Notes in Computer Science, No. 521.

CARDONA, L. (1993): CHRIS: Die Behandlung unzuverlässiger Inferenznetze anhand der Theorie der Hinweise. Diss. Univ. Fribourg, Wirtschafts- und Sozialwiss. Fak.

CARDONA, L.; KOHLAS, J.; MONNEY, P.A. (1991): The Reliability of Reasoning with Unreliable Rules and Propositions. In KRUSE, SIEGEL (1991), 125-129.

CARDONA, L.; KOHLAS, J.; MONNEY, P.A. (1992) Inference with Unreliable Rules, Proc. IPMU'92, 21-24. To appear in the volume of the Lecture Notes in Computer Science series devoted to this conference.

CHOQUET, G. (1953): Theory of Capacities. *Ann. Inst. Fourier* (Grenoble), **5**, 131-295.

D'AMBROSIO, B.; SMETS, P.; BONISSONE, P.P. (Eds.) (1991): *Uncertainty in AI 91.* Morgan and Kaufman, San Mateo, Cal.

DE KLEER, J. (1986): An Assumption-Based TMS. *Artificial Intelligence*, **28**, 127-162.

DEMPSTER, A. (1966): New Methods for Reasoning Towards Posterior Distributions Based on Sample Data. *Annals Math. Stat.*, **37**, 355-374.

DEMPSTER, A. (1967 a): Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals Math. Stat.*, **38**, 325-339.

DEMPSTER, A. (1967 b): Upper and Lower Probability Inferences Based on a Sample from a Finite Univariate Population. *Biometrika*, **54**, 515-528.

DEMPSTER, A. (1968 a): A Generalization of Bayesian Inference. *J. Royal Stat. Soc., Series B*, **30**, 205-247.

DEMPSTER, A. (1968 b): Upper and Lower Probabilities Generated by a Random Closed Interval. *Annals Math. Stat.*, **39**, 957-966.

DEMPSTER, A. (1990): Bayes, Fisher, and Belief Functions. In: GEISSER, S.; HODGES, J.S.; PRESS, S.J.; ZELLNER, A. (eds.): *Bayesian and Likelihood Methods in Statistics and Econometrics.*

DEUTSCH-MCLEISH, M. (1991): A Model for Non-Monotonic Reasoning Using Dempster's Rule. In BONISSONE et al. (1991), 481-494.

DUBOIS, D., PRADE, H. (1982): On Several Representations of an Uncertain Body of Evidence. In GUPTA, M.M.; SANCHEZ, E. (Eds.): *Fuzzy Information and Decision Processes.* North Holland, Amsterdam, 167-181.

DUBOIS, D.; PRADE, H. (1986): A Set-Theoretic View of Belief Functions. *Int. Journal of General Systems*, **12**, 193-226.

DUBOIS, D.; PRADE, H. (1985): *Théorie des possibilités. Applications à la représentation des connaissances en informatique.* Masson, Paris.

DUBOIS, D.; WELLMAN, M.P.; D'AMBROSIO, B.; SMETS, P. (Eds.)(1992): *Uncertainty in AI 92.* Morgan and Kaufman, San Mateo, Cal.

FAGIN, R.; HALPERN, J.Y. (1989): Uncertainty, Belief, and Probability. IJCAI-89, 1161-1167.

FAGIN, R.; HALPERN, J.Y. (1991): A New Approach to Updating Beliefs. In BONISSONE. et al. (eds.) (1991), 347-374.

FISHBURN, P.C. (1970): *Utility Theory for Decision Making.* Wiley, New York.

FRAZER, D.A.S. (1968): *The Structure of Inference.* Wiley, New York.

FEDRIZZI, M.; KACPRZYK, J.; YAGER, R.R. (Eds.) (1993): *Advances in the Dempster-Shafer Theory of Evidence.* Wiley.

GOODMAN,I.R.; NGUYEN, H.T. (1985): *Uncertainty Models for Knowledge-Based Systems.* North-Holland, Amsterdam.

GOODMAN, I.R.; GUPTA, M.M.; NGUYEN, H.T.; ROGERS, G.S. (1991): *Conditional Logic in Expert Systems.* North-Holland, Amsterdam.

GORDON, J.; SHORTLIFFE, E.H. (1985): A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space. *Artificial Intelligence*, **26**, 323-357.

HAJEK, P.; HAVRANEK, T., JIROUSEK, R. (1992): *Uncertain Information Processing in Expert Systems.* CRC Press, London.

HALMOS, P. (1969): *Lectures on Boolean Algebras.* Van Nostrand-Reinhold, London.

HALPERN, J.Y.; FAGIN, R. (1992): Two Views of Belief: Belief as Generalized Probability and Belief as Evidence. *Artificial Intelligence*, **54**, 275-317.

HENRION, M.; SHACHTER, R.D.; KANAL, L.N.; LEMMER, J.F. (eds.) (1990). *Uncertainty in Artificial Intelligence 5.* North Holland, Amsterdam.

HESTIR, K.; NGUYEN, H.T.; ROGERS, G.S. (1991): A Random Set Formalism for Evidential Reasoning. In GOODMAN, I.R. et al. (eds.) (1991), 309-344.

HSIA, Y.; SHENOY, P.P. (1989): Mac Evidence: A Visual Evidential Language for Knowledge Based Systems. Working Paper No. 211, Univ. of Kansas, School of Business.

INOUE, K. (1992): Linear Resolution for Consequence Finding. *Artificial Intelligence*, **56**, 301-353.

JAFFRAY, J.Y. (1989): Linear Utility Theory for Belief Functions. *Operations Research Letter*, **8**, 107-112.

JAFFRAY, J.Y. (1992 a): Dynamic Decision Making with Belief Functions. *Cahiers de Recherche en Economie, Mathématiques et Applications*, Univ. de Paris I, No. 92-15

JAFFRAY, J.Y. (1992 b): Bayesian Updating and Belief Functions. *IEEE Trans. SMC,* **22**, 1144-1152.

JAFFRAY, J.Y., WAKKER, P. (1992): Decision Making with Belief Functions: Compatibility and Incompatibility with the Sure-Thing Principle. Unpubl. Paper, Univ. de Paris VI, LID Paris, France, University of Nijmegen, NICI Nijmegen, The Netherlands.

KENNES, R. (1992): Computational Aspects of the Möbius Transform of a Graph. *IEEE Trans. SMC,* **22**, 201-223.

KENNES, R.; SMETS, P. (1991 a): Fast Algorithms for Dempster-Shafer Theory. In: BOUCHON et. al. (1991), 14-30.

KENNES,R.; SMETS,P. (1991 b): Computational Aspects of the Möbius Transform. In. BONISSONE et al (1991), 401-416.

KLAWONN, F.; SMETS, P. (1992): The Dynamic of Belief in the Transferable Belief Model and Specification-Generalization Matrices. In DUBOIS et al. (1992): pp. 130-137.

KLIR G.J. , FOLGER T.A. (1988) : *Fuzzy sets, Uncertainty, and Information.* Prentice-Hall.

KOHLAS, J. (1987). *Zuverlässigkeit und Verfügbarkeit.* Teubner, Stuttgart.

KOHLAS, J. (1989): Modeling Uncertainty with Belief Functions in Numerical Models. *European Journal of Operational Research*, **40**, 377-388.

KOHLAS, J. (1990): A Mathematical Theory of Hints. Inst. for Automation and Op. Res., University of Fribourg, Switzerland, Tech. Rep. No. 173 (to appear in *Int. J. General Systems* under the title: Support and Plausibility Functions Induced by Filter-Valued Mappings).

KOHLAS, J. (1991 a): The Reliability of Reasoning with Unreliable Arguments. *Annals of Operations Reseaarch*, **32**, 76-113.

KOHLAS, J. (1991 b): Describing Uncertainty in Dynamical Systems by Uncertain Restrictions. Article in: DI MASI G.B. et al. (Eds.) (1991): Modeling, Estimation and Control of Systems with Uncertainty. Birkhäuser.

KOHLAS, J. (1992): Evidential Reasoning About Parametric Models. Inst. for Automation and Op. Res., University of Fribourg, Switzerland, Tech. Rep. No. 194.

KOHLAS, J.; MONNEY, P.A. (1991 a): Representation of Evidence by Hints. Inst. for Automation and Op. Res., University of Fribourg, Switzerland, Tech. Rep. No. 186 (to be published in FEDRIZZI M. et al. (Eds.) (1993): *Advances in the Dempster-Shafer Theory of Evidence.* Wiley).

KOHLAS, J.; MONNEY, P.A. (1991 b): Propagating Belief Functions Through Constraint Systems, *Int. J. Approximate Reasoning*, **5**, 433-461. (a shorter version is published in BOUCHON et al., 1991, 50-57)

KOHLAS, J.; MONNEY, P.A. (1993): Probabilistic Assumption-Based Reasoning. Inst. for Automation and Op. Res., University of Fribourg, Switzerland, Tech. Rep. No. 208.

KONG, A. (1986): Multivariate Belief Functions and Graphical Models. Doct. diss. Harvard Univ., Dept. of Stat.

KRUSE, R.; SCHWECKE E., HEINSOHN J. (1991): *Uncertainty and Vagueness in Knowledge-Based Systems.* Springer.

KRUSE, R.; SIEGEL, P. (eds.) (1991): Symbolic and Quantitative Approaches to Uncertainty. Springer Lecture Notes in Computer Science No. 548.

LASKEY, K.B.; LEHNER, P.E. (1988): Belief Maintenance: An Integrated Approach to Uncertainty Management. Proc. Amer. Ass. for AI, 210-214 (reprinted in SHAFER, G.; PEARL, J. (eds.). (1990), 694-698)

LAURITZEN, S.L.; SPIEGELHALTER, D.J. (1988): Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems, *J. Royal Stat. Soc.*, Series B, **50**, 157-224.

LOHMANN, G. (1991): An Evidential Reasoning Approach to the Classification of Satellite Images. In KRUSE, R.; SIEGEL, P. (1991): p. 227-231.

LOWRANCE, J.D. (1988) : Automated Argument Construction. *J. Stat. Planning Inference*, **20**, 369-387.

LOWRANCE, J.D.; GARVEY, T.D.; STRAT, T.M. (1986): A Framework for Evidential Reasoning Systems, AAAI-86, 869-903.

MAIER, D. (1983): *The Theory of Relational Data Bases.* Computer Science Press.

MATHERON, G. (1975): *Random Sets and Integral Geometry.* Wiley, New York.

MELLOULI, K. (1987): On the Propagation of Beliefs in Networks Using the Dempster-Shafer Theory of Evidence. Doct. Diss., Univ. of Kansas, School of Business.

MONNEY P.A. (1991 a): *Le raisonnement temporel et géométrique sous incertitude.* Doctoral Thesis, Math. Institute, Univ. Fribourg, Switzerland.

MONNEY P.A. (1991 b): Planar Geometric Reasoning With the Theory of Hints. In BIERI H., NOLTEMEIER H. (Eds.) (1991): *Computational Geometry. Methods, Algorithms and Applications.* Lecture Notes in Computer Science, vol. 553, p. 141-159.

NGUYEN, H.T. (1978): On Random Sets and Belief Functions. *J. Math. Analysis and Applications*, **65**, 531-542.

NGUYEN, A.T.; SMETS, P. (1991): On Dynamics of Cautious Belief and Conditional Objects. Working Paper, TR/IRIDIA/91.13. To appear in *Int. J. of Approximate Reasoning.*

OLIVER R.M.; SMITH J.Q. (Eds.) (1990): *Influence Diagrams, Belief Nets and Decision Analysis.* Wiley

ORPONEN, P. (1990): Dempster's Rule of Combination is #P-complete. *Artificial Intelligence*, **44**, 245-253.

PEARL, J. (1988): *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, Cal.

PEARL, J. (1990): Reasoning with Belief Functions: An Analysis of Compatibility. *Int. J. Approximate Reasoning*, **4**, 363-389.

PROVAN, G.M. (1988): Solving Diagnostic Problems Using Extended Truth Maintenance Systems. Proc. European Conf on Art. Int., 547-552.

PROVAN, G.M. (1990 a): A Logic-Based Analysis of Dempster-Shafer Theory. *Int. J. Approximate Reasoning*, **4**, 451-495.

PROVAN, G.M. (1990 b): The Application of Dempster-Shafer Theory to a Logic-Based Visual Recognition System. In HENRION M. et al. (1990), 389-405.

REITER, R. (1987): A Theory of Diagnosis from First Principles. *Artificial Intelligence*, **32**, 57-95.

REITER. R., DE KLEER, J. (1987): Foundations of Assumption-Based Truth Maintenance Systems. Proc. Amer. Assoc. A.I., 183-188.

ROSE, D.J. (1970): Triangulated Graphs and the Elimination Process. *J. Math. Anal. and Appl.*, **32**, 597-609.

ROY, B. (1985): *Méthodologie Multicritère d'Aide à la Décision.* Editions Economica, Paris.

RUSPINI, E.H. (1987): The Logical Foundations of Evidential Reasoning. Tech. Note 408, AI Center, SRI International, Menlo Park, Calif.

SAFFIOTTI, A. (1991 a): Using Dempster-Shafer Theory in Knowledge Representation. In. BONISSONE et al. (1991), 417-431.

SAFFIOTTI, A. (1991 b): A Hybrid Belief System for Doubtful Agents. In: BOUCHON et. al. (1991), 393-402.

SAFFIOTTI, A.; UMKEHRER, E. (1991): PULCINELLA: A General Tool for Propagating Uncertainty in Valuation Networks. Proc. 7-th Conf on Uncertainty in AI, Los Angeles Ca.

SAFFIOTTI, A.; UMKEHRER, E. (1992): Automatic Generation of Valuation networks From General Clauses. Proc. IPMU'92, 341-345.

SCHALLER, J.P. (1991): *Multiple Criteria Decision Aid Under Incomplete Information.* Doctoral Thesis, Economic Sciences Faculty, Univ. of Fribourg, Switzerland.

SHAFER, G. (1976 a): *A Mathematical Theory of Evidence.* Princeton University Press.

SHAFER, G. (1976 b): A Theory of Statistical Evidence. In: HARPER, HOOKER (eds.): Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science. Vol. II, Reidel, Dortrecht, 365-463.

SHAFER, G. (1978): Dempster's Rule of Combination. Unpublished Manuscript.

SHAFER, G. (1979): Allocations of Probability. *Annals of Prob.*, **7**, 827-839.

SHAFER, G. (1982 a): Belief Functions and Parametric Models. *J. Royal Statist. Soc. B*, **44**, 322-352.

SHAFER, G. (1982 b): Constructive Decision Theory. Univ. of Kansas Dept. of Math. Working Paper.

SHAFER, G. (1987): Belief Functions and Possibility Measures. In BEZDEK (ed.): Analysis of Fuzzy Information, vol I: Mathematics and Logic, CRC Press, Boca Raton, 51-84.

SHAFER, G. (1990): Perspectives on the Theory and Practice of Belief Functions. *Int. J. Approximate Reasoning*, **4**, 323-362.

SHAFER, G. (1991): An Axiomatic Study of Computations in Hypertrees. Working Paper 232, Business School, Univ. of Kansas.

SHAFER, G. (1992): Rejoinders to Comments on "Perspectives on the Theory and Practice of Belief Functions". *Int. J. Approximate Reasoning*, **6**, 445-480.

SHAFER, G., LOGAN,R. (1987): Implementing Dempster's Rule for Hierarchical Evidence. *Artificial Intelligence*, **33**, 271-289.

SHAFER, G.; PEARL, J. (eds.) (1990): *Readings in Uncertain Reasoning.* Morgan-Kaufman Publ. San Mateo, Cal.

SHAFER, G.; SHENOY, P.P.; MELLOULI, K. (1987): Propagating Belief Functions in Qualitative Markov Trees. *Int. J. Approximate Reasoning*, **1**, 349-400.

SHAFER, G.; SRIVASTAVA, R.P. (1990): The Bayesian and Belief Function Formalisms: A General Perspective for Auditing, Auditing. *J. Pract. Theory,* **9** (Suppl.).

SHENOY, P.P.; SHAFER, G. (1990): Axioms for Probability and Belief-Function Propagation. In SHAFER, PEARL, 1990, 575-610.

SIEGEL, P. (1987): Représentation et utilisation de la connaissance en calcul propositionnel. Ph. D. thesis, Univ. de Marseilles II.

SMETS, P. (1979): Modèle Quantitatif du Diagnostic Médical. *Bulletin de l'Académie de Médecine de Belgique*, **134**, 330-343.

SMETS, P. (1981): Medical Diagnosis: Fuzzy Sets and Degrees of Belief. *Fuzzy Sets and Systems*, **5**, 259-266.

SMETS, P. (1988): Belief Functions. In SMETS, P.; MAMDANI, A.; DUBOIS, D.; PRADE, H.: *Non-Standard Logics for Automated Reasoning.* Academic Press, London, 253-286.

SMETS, P. (1989): Constructing the Pignistic Probability Function in a Context of Uncertainty. Proc. of the Fifth Workshop on Uncertainty in AI, Windsor, Canada, 319-326. Also in HENRION et al. (1990), 29-39.

SMETS, P. (1990): The Combination of Evidence in the Transferable Belief Model. *IEEE Trans. Pattern Anal. and Machine Int.*, **12**, 447-458.

SMETS, P. (1992): The Nature of the Unnormalized Beliefs Encountered in the Transferable Belief Model. In Dubois et al. (1992), p. 292-297.

SMETS,P.; HSIA, Y.T..(1991): Default Reasoning and the Transferable Belief Model. In BONISSONE et al., 1991, 495-504.

SMETS, P.; KENNES, R. (1990): The Transferable Belief Model. IRIDIA Technical Report TR/IRIDIA/90-14.2, Bruxelles.

STRAT, T.M. (1990): Decision Analysis Using Belief Functions. *Int. J. Approximate Reasoning*, **4**, 391-418.

TARJAN, R.E.; YANNAKAKIS, M. (1984): Simple Linear Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs. *SIAM J. Computing*, **13**, 566-579.

THOMA, H.M. (1991): Belief Function Computations. In GOODMAN, I.R. et al. (eds.) (1991), 269-308.

VAKILI (1993): Approximation of Hints. Inst. for Automation and Op. Res., University of Fribourg, Switzerland, Tech. Rep. No. 209.

VON NEUMANN, J.; MORGENSTERN, O. (1944): *Theory of Games and Economic Behavior.* Princeton Univ. Press, Princeton, N.J.

WASSERMAN, L.A. (1990 a): Prior Envelopes Based on Belief Functions. *Annals of Stat.*, **18**, 454-464.

WASSERMAN, L.A. (1990 b): Belief Functions and Statistical Inference. *Can. J. Stat.*, **18**, 183-196.

WILLIAMS, P.M. (1982): Discussion of Shafer (1982 a), *J. Royal Statist. Soc. B*, **44**, 342.

WESLEY, L.P. (1986): Evidential Knowledge-Based Computer Vision. *Optical Engineering*, **25**, 363-379.

XU, H. (1992): An Efficient Tool for Reasoning with Belief Functions. Proc. IPMU'92 , 65-68.

YAGER, R.R. (1985): The Entailment Principle for Dempster-Shafer Granule. Tech. Report MII-512, Iona College, New Rochelle, N.Y.

YEN, J. (1989): GERTIS: A Dempster-Shafer Approach to Diagnosing Hierarchical Hypotheses. *Communications ACM*, **32**, 573-585.

ZARLEY, D.; HSIA, Y.T., SHAFER, G. (1988): Evidential Reasoning Using DELIEF, Proc. Nat. Conf, A.I., 1988; Reprinted in SHAFER, PEARL, 1990, 619-623.

ZHANG, L. (1988): Studies on Finding Hypertree Covers for Hypergraphs. Working Paper No. 198, Univ. of Kansas, School of Business.