ON THE JUSTIFICATION OF DEMPSTER'S RULE OF COMBINATION

Frans Voorbraak
Department of Philosophy, Rijksuniversiteit te Utrecht

Logic Group
Preprint Series
No. 42
December 1988

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

On the justification of Dempster's rule of combination

Frans Voorbraak

Department of Philosophy, University of Utrecht
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

Abstract
In Dempster-Shafer theory it is claimed that the pooling of evidence is reflected by Dempster's rule of combination, provided certain requirements are met. The justification of this claim is problematic, since the existing formulations of the requirements for the use of Dempster's rule are not completely clear. In this paper, randomly coded messages, Shafer's canonical examples for Dempster-Shafer theory, are employed to clarify these requirements and to evaluate Dempster's rule. The range of applicability of Dempster-Shafer theory will turn out to be rather limited. Further, it will be argued that the mentioned requirements do not guarantee the validity of the rule and some possible additional conditions will be described.

Acknowledgment: I would like to thank Gerard Renardel de Lavalette for drawing my attention to the problem of justifying Dempster's rule of combination.

December 1988

Contents

1 Introduction 3

2 Dempster-Shafer theory 4

2.1 Belief functions 4

2.2 Dempster's rule of combination 5

3 Arguments against Dempster's rule 9

3.1 Lemmer's counterexample 9

3.2 An amended example 11

4 Discerning the relevant interaction 13

5 Independent bodies of evidence 16

5.1 Shafer's constructive probability 16

5.2 DS-independence 17

5.3 The interaction of evidence 19

6 Assumptions underlying Dempster's rule 21

6.1 Conditions for the validity of Dempster's rule 21

6.2 Bayesian approximations 23

7 Conclusion 25

Appendix: propositions and proofs 26

References 30

1 Introduction

Dempster's rule of combination is the most important tool of Dempster-Shafer theory (also known as evidence theory or theory of belief functions). This theory, which originates from Arthur Dempster [2] and has been developed to its present form by Glenn Shafer [10,11,12,13], is meant to be a theory of evidence and probable reasoning and has recently received much attention as a promising theory for the handling of uncertain information in expert systems. Mathematically, Dempster's rule of combination is simply a rule for computing a belief function from two other belief functions. However, it is claimed that the rule reflects the pooling of evidence within Dempster-Shafer theory, provided certain requirements are met. The justification of this claim presents a problem which is the subject of this paper.

Unfortunately, Shafer's first formulation of the requirements for the use of Dempster's rule was rather vague and, despite several attempts to clarify the meaning of the requirements, the existing formulations are still not completely clear. This obscurity may have contributed to the fact that in the literature these requirements are often more or less ignored: frequently, the only proviso mentioned is that the bodies of evidence to be combined have to be independent, which may at best be considered to be a very superficial account of the requirements. In fact, the use of the term "independent" may be misleading, since in the context of Dempster-Shafer theory independence cannot be equated with stochastical independence.

One of the main objectives of this paper is to contribute to the clarification of the requirements for the use of Dempster's rule and in particular to the clarification of the Dempster-Shafer theory notion of independence. Further, it will be argued that Shafer's requirements are not sufficient to guarantee the validity of the rule and some possible additional conditions for the use of the rule will be described.

Section 2 consists of a brief introduction to Dempster-Shafer theory. In section 3 we describe and amend John F. Lemmer's argument against the applicability of Dempster's rule under a sample space interpretation of Dempster-Shafer theory: he gives an example in which the application of Dempster's rule yields an undesirable result. This (amended) example is used to give concrete form to the considerations of sections 4 and 5 where the requirements for the use of Dempster's rule are studied in detail. In section 6 the additional assumptions underlying the rule are described.

2 Dempster-Shafer theory

In this section we briefly explain some notions and terminology of Dempster-Shafer theory. For a more detailed exposition and some background information see Shafer [10] or Gordon and Shortliffe [4].

2.1 Belief functions

Let Θ be a set of mutually exclusive and exhaustive hypotheses about some problem domain. (For simplicity, Θ will always be assumed to be finite.) Relevant propositions are represented as subsets of this set Θ which is called the **frame of discernment**, or simply **frame**. Suppose for example $\Theta = \{a,b,c,d\}$, then the proposition $(a \vee c) \wedge \neg b \wedge \neg d$ is represented by the set $\{a,c\}$. In Fig. 1 the hierarchy of propositions is depicted. Notice that set-theoretical inclusion between sets corresponds with logical implication between the represented propositions.

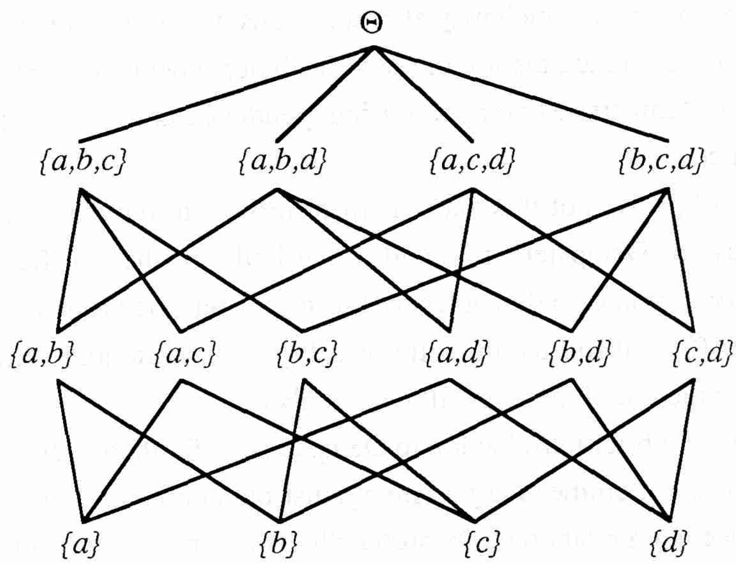

Fig. 1. Hierarchy of propositions for the frame $\{a,b,c,d\}$

Definition 2.1 Let Θ be a frame. A **basic probability assignment** (bpa) on Θ is a function m from 2^{Θ}, the powerset of Θ, to [0,1] such that

$$m(\varnothing) = 0 \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1$$

The quantity $m(A)$, called A's **basic probability number**, corresponds to the measure of belief that is committed exactly to the proposition (represented by the set) A and in general not to the total belief committed to A, since the total belief includes the measures of belief committed to subsets of A. This explains the following:

Definition 2.2 The **belief function** Bel induced by the bpa m on Θ is defined by

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad (A, B \subseteq \Theta)$$

$Bel(A)$ measures the total belief committed to A. Each belief function Bel is induced by a unique bpa m which can be recovered from Bel as follows:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \qquad (A \subseteq \Theta)$$

A proper evaluation of the degree of belief in a proposition A takes into consideration not only the measure of belief committed to A, but also that committed to the negation of A. Information concerning this latter measure is indirectly coded in the following function:

Definition 2.3 The **plausibility function** Pl induced by the bpa m is defined by

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

It is easy to see that $Bel(A) \leq Pl(A)$ and that $Pl(A) = 1 - \mathrm{Bel}(A^c)$, where A^c denotes the set-theoretical complement of A. $[\mathrm{Bel}(A), \mathrm{Pl}(A)]$ is called the **belief interval** of A.

Definition 2.4 Let m be the bpa of Bel.

(1) If $m(A) > 0$, then A is called a **focal element** of Bel.

(2) The union of all focal elements of Bel is called the **core** of Bel.

(3) A belief function is called **vacuous** if Θ is its only focal element.

(4) If all focal elements of Bel are singletons, then Bel is called **Bayesian**.

2.2 Dempster's rule of combination

The following explanation of Dempster's rule of combination is essentially the one given in Glenn Shafer's *A Mathematical Theory of Evidence*.

Let m and m' be the bpa's of the belief functions Bel and Bel' with cores $\{A_1,...,A_p\}$ and $\{B_1,...,B_q\}$ respectively. The probability masses assigned by the bpa's can be visualized as segments of the unit interval, as shown in Fig. 2 below.

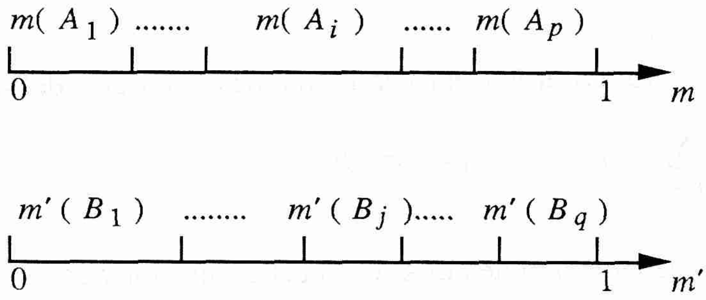

Fig. 2. Probability masses assigned by m and m'.

Fig. 3 shows how the two intervals can be orthogonally combined to obtain a square representing the total probability mass assigned by $Bel \oplus Bel'$, the combination of Bel and Bel', where Bel commits vertical strips to its focal elements and Bel' horizontal ones.

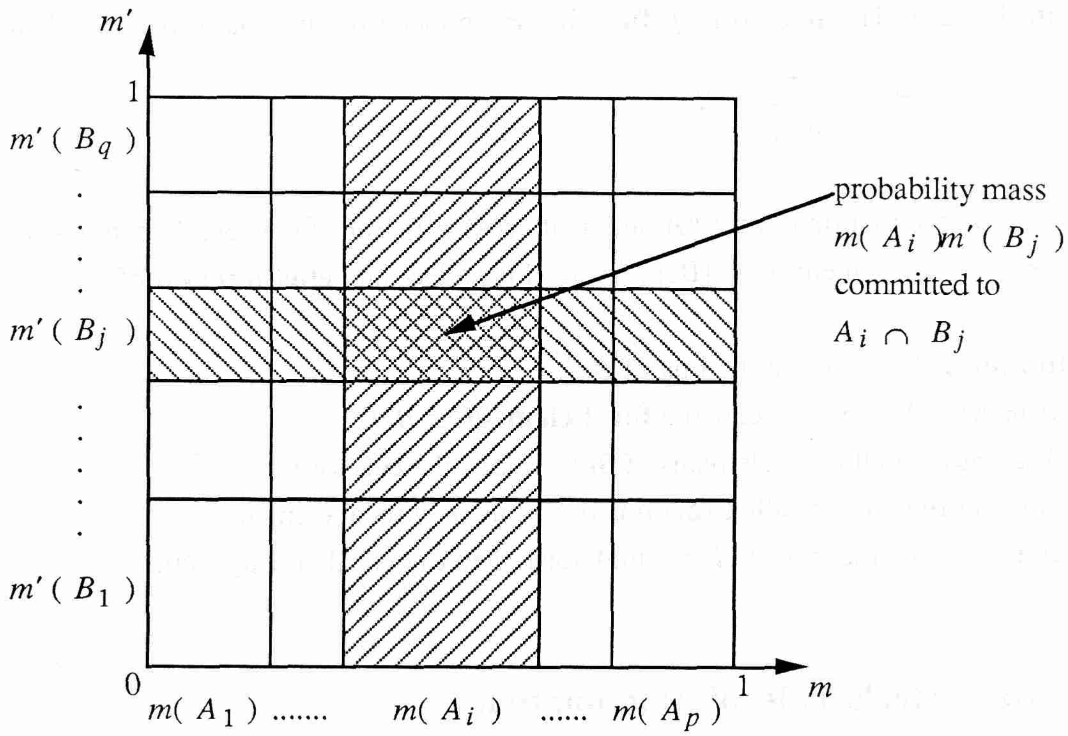

Fig. 3. Probability masses assigned by the combination of Bel and Bel'.

The exact commitment of the intersection of the vertical strip of measure $m(A_i)$ and the horizontal strip of measure $m'(B_j)$ has measure $m(A_i)\cdot m'(B_j)$ and since it is committed both to A_i and B_j, we may say that it is committed $A_i \cap B_j$. A given subset A of Θ may of course have more than one of these rectangles exactly committed to it. Hence to obtain the total probability mass exactly committed to A by the combination m and m' (notation $m \oplus m'$) we have to take the sum of all $m(A_i)\cdot m'(B_j)$ such that $A = A_i \cap B_j$.

However, in this way some probability mass may be committed to $\varnothing$, since there may be a focal element Ai of m and a focal element B_j of m' such that $A_i \cap B_j = \varnothing$. But then $m \oplus m'$ would fail to be a bpa. Therefore all rectangles committed to the empty set are discarded and the measures of the remaining rectangles are rescaled by dividing through the sum of all $m(A_i)\cdot m'(B_j)$ such that $A_i \cap B_j \neq \varnothing$, provided this sum does not equal 0; otherwise we say that $Bel \oplus Bel'$ does not exists or that Bel and Bel' are not combinable. Hence we arrive at the following definition:

Definition 2.5 (Dempster's rule of combination) Let Bel and Bel' be belief function induced by the bpa's m and m' respectively. If $\Sigma\{m(A_i)\cdot m'(B_j) \mid A_i \cap B_j \neq \varnothing\} = 0$, then Bel and Bel' are called **not combinable**. Otherwise, $Bel \oplus Bel'$, the **combination** of Bel and Bel' **by Dempster's rule**, is the belief function induced by $m \oplus m'$, where

$$m \oplus m'(\mathrm{A}) = \frac{\sum_{A_i \cap B_j = A} m(A_i)\cdot m'(B_j)}{\sum_{A_i \cap B_j \neq \varnothing} m(A_i)\cdot m'(B_j)} \qquad (\text{if } A \neq \varnothing;\ m \oplus m'(\varnothing) = 0.)$$

The factor $[\Sigma\{m(A_i)\cdot m'(B_j) \mid A_i \cap B_j \neq \varnothing\}]^{-1}$ is called **renormalizing constant** of Bel and Bel'.

Shafer exhibits some "sensible and intuitive results" of Dempster's rule in the simplest cases, but he does not give an *a priori* justification of the rule:

> Given several belief functions over the same frame of discernement but based on distinct bodies of evidence, Dempster's rule of combination enables us to compute their orthogonal sum, a new belief function based on the combined evidence. Though this essay provides no conclusive *a priori* argument for Dempster's rule, we will see in the following chapters that the rule does seem to reflect the pooling of evidence, provided only that the belief functions to be combined are actually based on entirely distinct bodies of evidence and that the frame of discernment discerns the relevant interaction of these bodies of evidence. (Shafer [10], p. 57)

What it means for a frame of discernment to discern the relevant interaction was adequately explained in chapter 8 of Shafer [10]. (See also section 4 of this paper.) But the meaning of "entirely distinct bodies of evidence" remained obscure.

It seems obvious that one should not use Dempster's rule to combine a body of evidence X with itself, or more generally, with some body of evidence Y which is implied by X (i.e. with some body of evidence Y for which $x \rightarrow y$ holds, where x (y) denotes the proposition which expresses the evidence X (Y)). The reason for this is that if X implies Y, then the evidence Y is already taken into account in the basic probability assignment m_X. Hence one should have $Bel_X \oplus Bel_Y = Bel_X$, whereas in general this does not hold:

Example 2.6 Let Θ be the frame $\{A, \neg A\}$, where A denotes the proposition "patient P has the flue". Suppose that X represents the observation that P has a fever > 39° C, that Y represents the observation that P has a fever > 38.5° C and that the basic probability assignments of X and Y are: $m_X(\mathrm{A}) = 0.6$, $m_X(\Theta) = 0.4$, $m_Y(A) = 0.4$ and $m_Y(\Theta) = 0.6$. Then $Bel_X \oplus Bel_Y(A) = 0.76$, which does not equal $Bel_X(A) = 0.6$.

A natural following case to be considered is that where, although X does not (logically) imply Y, X considerably increases the probability of Y, i.e. $\mathrm{P}(y \mid x) >> \mathrm{P}(y)$. It might seem plausible that also in this case the effect of the body of evidence Y is (for a large part, at least) already taken into account in the basic probability assignment m_X. In that case m_Y should not be allowed to function as an equal partner of m_X in Dempster's rule of combination. But this would suggest that the rule is only justified if the bodies of evidence to be combined are independent in the sense that $\mathrm{P}(x \wedge y) = \mathrm{P}(x) \cdot \mathrm{P}(y)$.

Indeed, the requirement that the belief functions to be combined have to be based on independent evidence is often mentioned in the literature as a proviso on the application of Dempster's rule. However, authors largely ignore Shafer's opinion that the belief function concept of independence, which will be referred to as **DS-independence**, differs from the usual probability theory concept. (Cf. Shafer [11].) The inadequacy of the usual interpretation of independence is illustrated in the following sections by a description of a proposed "counterexample" to Dempster's rule followed by a detailed study of Shafer's requirements for the use of the rule.

3 Arguments against Dempster's rule

Dempster-Shafer theory has been criticized by means of arguments purporting to show that probability theory is the unique correct description of uncertainty (see e.g. Lindley [9]). These arguments, which are based on the work of Savage and De Finetti, undoubtedly have some appeal, but they are not entirely convincing. In spite of them, several authors have argued for a relaxation of the probability axioms. In particular, the additivity axiom is felt to be too restrictive in case one has to deal with uncertainty deriving from (partial) ignorance. (See Kyburg [7] or Fishburn [3].)

In this paper, we will focus on more concrete arguments against Dempster-Shafer theory, which do not find fault with the failure of additivity in the theory, but question the validity of Dempster's rule by claiming that in some situations its application yields undesirable results. (see Kyburg [7], Lemmer [8] or Zadeh [16].) These arguments are at least incomplete, since it is not shown that the requirements for the use of Dempster's rule are satisfied in the particular situations under consideration. Below we desribe (and amend) one such argument which is relatively explicit about the kind of independence that is assumed.

3.1 Lemmer's counterexample

In Lemmer [8] John Lemmer proposes the following sample space interpretation of Dempster-Shafer theory: Imagine balls in an urn which have a single "true" label. The set of these labels (or rather, the set of propositions expressing that a ball has a particular label from the set of these labels) functions as the frame of discernment Θ. Henceforth we will call these "true" labels **frame labels**. Belief functions are formed empirically on the basis of evidence acquired from observation processes which are called **sensors**. These sensors attribute to the balls labels which are subsets of Θ.

The labelling of each sensor is assumed to be **accurate** in the sense that the frame label of a particular ball is consistent with the attributed label. (In other words, a sensor s is accurate if s attributes to each ball with frame label t a label U such that $\{t\}\subseteq U$. Each sensor s gives rise to a bpa m by defining for all $U\subseteq\Theta$ $m(U)$ to be the fraction of balls labelled U by sensor s. Then, due to the assumed accurateness of the sensors, $Bel(U)$ corresponds to the minimum fraction of balls for with a frame label $t\in U$ and $Pl(U)$ corresponds to the maximum fraction of balls which could have a frame label which is an element of U.

Lemmer proceeds by giving the following example which shows that the combination by Dempster's rule of belief functions which derive from accurate labelling processes does not necessarily yield a belief function which assigns accurate probability ranges to each proposition, even if the labelling processes are independent when conditioned by the value of the frame label. It is not clear why he chooses for conditional independence, but he claims that unconditional independence of labelling processes is undesirable in real applications since it would imply a non-zero probability of inconsistent labelling by the two sensors.

Example 3.1 Let $\Theta = \{a,b,c\}$ and assume that balls labelled a are light in weight while balls labelled either b or c are heavy and that balls labelled either a or b are red while balls labelled c are blue. Let s be a sensor which can classify weight and labels balls having frame label a with $\{a\}$ and balls having either b or c as frame label with $\{b,c\}$ and s' a sensor which can classify color and attributes the label $\{a,b\}$ to balls with either a or b as frame label and the label $\{c\}$ to balls with frame label c. Let $P(t)$ represent the actual fraction of balls with frame label t. The bpa's m and m' resulting from the sensors s and s' respectively are given by:
$m(\{a\}) = P(a)$, $m(\{b,c\}) = 1 - P(a)$, $m'(\{c\}) = P(c)$ and $m'(\{a,b\}) = 1 - P(c)$.

It is easy to see that although the labelling processes are unconditionally independent only if $P(a) = 0$ or $P(c) = 0$, they are independent when conditioned by the value of the frame label. Dempster's rule yields

$$\sum_{\{a\}\cap U\neq\varnothing} m\oplus m'(U) = \frac{m(\{a\})\cdot m'(\{a,b\})}{1 - m(\{a\})\cdot m'(\{c\})} = \frac{P(a)\cdot(1-P(c))}{1 - P(a)\cdot P(c)}$$

For $Bel\oplus Bel'$ to be accurate, $P(a) \leq Pl\oplus Pl'(\{a\})$ must be valid. Hence the following inequality must hold:

$$P(a) \leq \frac{P(a)\cdot(1-P(c))}{1 - P(a)\cdot P(c)} \Rightarrow P(a) = 0 \vee P(c) = 0.$$

Hence if neither $P(c)$ nor $P(a)$ equals zero, then the belief function $Bel\oplus Bel'$ does not correctly bound the value of $P(a)$.

Lemmer concludes that Dempster's rule is not applicable in situations which can be modelled by sample spaces in which sample points have "true" labels independent of the labelling process. However, we do not believe that his example provides strong evidence for this conclusion, since we doubt the adequacy of the described sample space interpretation of Dempster-Shafer theory. Consider Lemmer's own words on the sample space interpretation of classical and Bayesian probability theory:

> The problem in classical probability theory is to estimate, before drawing the ball, the probability of drawing a ball labeled "a". This probability, under appropriate assumptions about how the ball is drawn, is usually taken to be the fraction of balls in the urn which are labeled "a".....The problem in Bayesian probability theory is to estimate, having drawn the ball and made an observation about label "a", the probability of the ball also being marked with label "b". Under Bayes' rule this probability is taken to be the fraction of balls having label "b" in that portion of the sample defined by balls having label "a". (Lemmer [8], p.119)

Likewise, the problem in Dempster-Shafer theory is to attribute, having drawn the ball and obtained some information about it, probability masses to the different possible propositions concerning the frame label of the ball. In this way Dempster-Shafer theory can be seen as a generalization of Bayesian probability theory, whereas Lemmer's sample space interpretation is a generalizaton of the interpretation of classical probability theory: although he describes an experiment in his model to be the "drawing of a ball and assigning a probability range to each possible proposition about the true label of the ball", the result of the experiment is in fact independent of the chosen ball.

Below we adjust Lemmer's sample space interpretation in a more or less obvious manner so that the evidence that a particular ball is drawn is taken into account and we reformulate Lemmer's argument in terms of this adjusted interpretation.

3.2 An amended example

Consider an urn containing balls with multiple labels. Each ball has among its multiple labels exactly one label from some set of frame labels, say $\Theta = \{a,b\}$. Suppose X denotes the evidence that a particular ball drawn from the urn has label x. Given X, the probability that the ball has label a (b) equals $P(a \mid x)$ ($P(b \mid x)$), i.e. the fraction of the x-labelled balls which are labelled a (b). Let Bel_X be a belief function on the frame of discernment $\{a,b\}$ based on the evidence that the drawn ball has label x and possibly on some evidence concerning the (fraction of the) frame labels of the x-labelled balls.

Definition 3.2 Let X be the evidence that x is the case, Bel_X a belief fuction based on X on a frame Θ and P a probability measure on $\{u \wedge x \mid u \in \Theta\} \cup \{u \wedge \neg x \mid u \in \Theta\}$. Bel_X is called **conditionally accurate** w.r.t. P if for all $U \subseteq \Theta$:

$$Bel_X(U) \leq P(U \mid x) \leq Pl_X(U).$$

The notion of conditional accurateness is of course closely related to Lemmers notion of accurateness: one can think of a Bel_X which is conditionally accurate w.r.t. a P measuring actual fractions of balls as deriving from an accurate sensor whose working is restricted to the x-labelled balls. The following example shows that the conditional accurateness of two belief functions Bel_X and Bel_Y does not imply the conditional accurateness of $Bel_X \oplus Bel_Y$, even in case X and Y are independent in the sense that $P(x \wedge y) = P(x) \cdot P(y)$.

Example 3.3 Consider an urn containing 100 balls which are labelled according to Table 1 below.

Table 1. Contents of urn of example 3.3. Each string of labels is followed by the number of balls that have exactly those labels.

labels	number of balls
axy	4
ax	4
ay	16
a	16
bxy	10
bx	10
by	20
b	20

Let the bpa's m_X and m_Y be given by: $m_X(\{a\}) = 2/7$, $m_X(\{b\}) = 5/7$, $m_Y(\{a\}) = 2/5$ and $m_Y(\{b\}) = 3/5$ and let P measure the actual fractions of balls in the urn. Then Bel_X and Bel_Y are conditionally accurate w.r.t to P since for all $U \subseteq \{a,b\}$:

$$Bel_X(U) \leq P(U \mid x) \leq Pl_X(U) \text{ and } Bel_Y(U) \leq P(U \mid y) \leq Pl_Y(U).$$

Notice that in this case Bel_X and Bel_Y are even Bayesian belief functions and that $P(x \wedge y) = 0.14 = 0.28 \cdot 0.5 = P(x) \cdot P(y)$. We also have conditional independence: $P(x \mid a) \cdot P(y \mid a) = P(x \wedge y \mid a)$ and $P(x \mid b) \cdot P(y \mid b) = P(x \wedge y \mid b)$. Dempster's rule yields $Bel_X \oplus Bel_Y(\{b\}) = 15/19 \approx 0.79$, while $P(b \mid x \wedge y) = 5/7 \approx 0.71$. Hence $Bel_X \oplus Bel_Y$ is not conditionally accurate, since $Bel_X \oplus Bel_Y(\{b\}) > P(b \mid x \wedge y)$.

Before drawing any conclusion from this example, we will have a closer look at the requirements for the use of Dempster's rule.

4 Discerning the relevant interaction

Dempster's rule should only be used if the frame of discernment Θ is fine enough to discern all relevant interaction of the evidence to be combined. The following example shows what can go wrong if a frame does not satisfy this requirement:

Example 4.1 During his investigation of a burglary Sherlock Holmes has come up with the following two pieces of evidence:

E_1: from his inspection of the safe Holmes was able to conclude, with a high degree of certainty, say 0.7, that the thief was left-handed.

E_2: from the state of the door which gives access to the room with the safe he was able to conclude, again with a high degree of certainty, say 0.8, that the theft was an "inside job".

Let $\Theta = \{LI, LO, RI, RO\}$, where LI stands for the proposition that the thief was a left-handed insider, etc. Then the bpa's m_1 and m_2, based on E_1 and E_2 respectively, are given by:

$$m_1(\{LI,LO\}) = 0.7 \qquad m_2(\{LI,RI\}) = 0.8$$
$$m_1(\Theta) = 0.3 \qquad m_2(\Theta) = 0.2$$

Dempster's rule yields:

$$m_1 \oplus m_2(\{LI\}) = 0.56 \qquad m_1 \oplus m_2(\{LI,LO)\}) = 0.14$$
$$m_1 \oplus m_2(\Theta) = 0.06 \qquad m_1 \oplus m_2(\{LI,RI\}) = 0.24$$

However, if one takes as frame of discernment $\Omega = \{LI, \neg LI\}$, then both E_1 and E_2 give rise to a vacuous belief function and thus $Bel_1 \oplus Bel_2$ is also vacuous on Ω, although the combination of E_1 and E_2 does support $\{LI\}$. Ω fails to discern this support because it can discern E_1's support for $\{LI,LO\}$ and E_2's support for $\{LI,RI\}$ only as support for Ω and $\Omega \cap \Omega = \Omega \neq \{LI\} = \{LI,LO\} \cap \{LI,RI\}$.

Before we can give an exact description of what it means for a frame to discern all relevant interaction of two bodies of evidence, we need some technical definitions from chapters 6 and 8 of Shafer [10]:

Definition 4.2 Let Θ and Ω be frames of discernment.

(1) $\omega : 2^\Theta \rightarrow 2^\Omega$ is a **refining** if for all $A \subseteq \Theta$ $\omega(A)$ denotes the same proposition as A.

(2) $\omega : 2^\Theta \rightarrow 2^\Omega$ is an **ultimate refining** if for each refining $\phi: 2^\Theta \rightarrow 2^\Psi$ there exists a refining $\psi : 2^\Psi \rightarrow 2^\Omega$ such that $\omega = \psi\phi$.

If there exists a refining $\omega : 2^\Theta \rightarrow 2^\Omega$, then Ω is called a **refinement** of Θ and Θ is called a **coarsening** of Ω. If there exists an ultimate refining $\omega : 2^\Theta \rightarrow 2^\Omega$, then Ω is called an **ultimate refinement** of Θ.

In fact, Shafer defines a refining to be a function $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ such that

(i) $\omega(A) = \bigcup_{\theta \in A} \omega(\{\theta\})$

(ii) the sets $\omega(\{\theta\})$ with $\theta \in \Theta$ constitute a disjoint partition of Ω, i.e.

(1) $\omega(\{\theta\}) \neq \varnothing$ for all $\theta \in \Theta$

(2) $\omega(\{\theta\}) \cap \omega(\{\theta'\}) = \varnothing$ if $\theta \neq \theta'$

(3) $\bigcup_{\theta \in \Theta} \omega(\{\theta\}) = \Omega$

Although it is clear that if ω is a refining according to definition 4.2, then ω satisfies Shafer's requirements for a refining, the two definitions are not equivalent, since the fact that ω is a refining according to Shafer's definition does not imply that $\omega(A)$ expresses the same proposition as A (For example, if $\Theta = \{\theta, \neg\theta\}$ and ω is defined by $\omega(\varnothing) = \varnothing$, $\omega(\{\theta\}) = \{\neg\theta\}$, $\omega(\{\neg\theta\}) = \{\theta\}$and $\omega(\Theta) = \Theta$, then according to Shafer's definition, ω is a refining $2^{\Theta} \rightarrow 2^{\Theta}$.) We believe that definition 4.2 is to be preferred, since Shafer implicitly assumes that if ω is a refining, then $\omega(A)$ and A denote the same proposition. The following useful properties of a refining $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ follow trivially from definition 4.2: (i) $\omega(A) = \varnothing$ iff $A = \varnothing$

(ii) for all $A,B \subseteq \Theta$ $\omega(A \cap B) = \omega(A) \cap \omega(B)$

Definition 4.3 Let $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ be a refining.
The **inner reduction** of ω is the mapping $\omega^{i} : 2^{\Omega} \rightarrow 2^{\Theta}$ given by

$$\omega^{i}(A) = \{\theta \in \Theta \mid \omega(\{\theta\}) \subseteq A\}$$

The **outer reduction** of ω is the mapping $\omega^{o} : 2^{\Omega} \rightarrow 2^{\Theta}$ given by

$$\omega^{o}(A) = \{\theta \in \Theta \mid \omega(\{\theta\}) \cap A \neq \varnothing\}$$

Definition 4.4 A frame Θ is said to **discern the relevant information** of the bodies of evidence E_1 and E_2 if $\omega^{o}(A \cap B) = \omega^{o}(A) \cap \omega^{o}(B)$ whenever ω^{o} is the outer reduction of a refining $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ and A and B are focal elements of Bel_1 and Bel_2 respectively, where Bel_i is the belief function on Ω which arises from evidence E_i.

In general there exists no method to decide conclusively whether a particular frame discerns all relevant interaction of the evidence to be combined, but according to Shafer one can often acquire confidence that the frame is fine enough from a general appraisal of the evidence. E.g. he claims that one does not need a detailed analysis to become confident that the details which will receive support from E_1 after refining the frame Θ in example 4.1 are independent from those which will receive support from E_2.

In case of the (highly) idealized example 3.3 we are able to be more definite, since here the frame $\Theta = \{a,b\}$ has an ultimate refinement. The following proposition shows that the existence of an ultimate refinement considerably simplifies the matter.

Proposition 4.5 If $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ is an ultimate refining and $\omega^{\rho}(A \cap B) = \omega^{\rho}(A) \cap \omega^{\rho}(B)$ whenever A and B are focal elements of Bel_1 based on evidence E_1 and Bel_2 based on evidence E_2 respectively, then Θ discerns the relevant interaction of E_1 and E_2.

(Proofs of the propositions can be found in the appendix.)

Consider example 3.3.
Write Ω for $\{axy,ax,ay,a,bxy,bx,by,b\}$ and let the refining $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ be given by:

$$\omega(\{a\}) = \{axy,ax,ay,a\}$$

$$\omega(\{b\}) = \{bxy,bx,by,b\}.$$

It is easy to see that ω is an ultimate refining and it is a matter of routine to check that $\omega^{\rho}(A \cap B) = \omega^{\rho}(A) \cap \omega^{\rho}(B)$, for all A and $B \subseteq 2^{\Omega}$ which are focal elements of Bel_X and Bel_Y respectively. For example, $\omega^{\rho}(\{axy,ax\} \cap \{axy,ay\}) = \omega^{\rho}(\{axy\}) = \{a\} = \{a\} \cap \{a\} = \omega^{\rho}(\{axy,ax\}) \cap \omega^{\rho}(\{axy,ay\})$. We may conclude that in example 3.3 the requirement that the frame has to discern all relevant interaction of the evidence to be combined is met. In the following section we investigate the other (less clear) requirement for the use of Dempster's rule.

5 Independent bodies of evidence

The first serious attempt to explain what it means for two bodies of evidence to be "entirely distinct" or independent appeared in Shafer [11]. This explanation, which was further developed in subsequent papers (Shafer [12,13]), was placed in the context of what Shafer called the "constructive interpretation of probability". Hence before describing the Dempster-Shafer theory notion of independence, we first give an outline of this constructive interpretation.

5.1 Shafer's constructive probability

According to the constructive interpretation of probability, probability judgments are made by comparing the particular situation at hand to abstract canonical examples in which the uncertainties are governed by known chances. These canonical examples matching the particular situation may e.g. be games which can be played repeatedly and for which the long-run frequencies of the possible outcomes are known. In this case one arrives at "classical" probability judgments. Shafer claims that sometimes, in particular in those situations where one does not have sufficient information to make a comparison to examples like these games, other kinds of examples may be judged to be appropriate and that the use of these other canonical examples may give rise to judgments formulated in terms of Dempster-Shafer theory. We will describe the most general kind of these examples: randomly coded messages.

Let Θ be a frame of discernment. Suppose someone sends us an infallible encoded message X, where the code is randomly taken from the list $c_1,c_2,\ldots,c_n$ and the chance that code c_i is used is p_i. (Both the list and the associated chances are supposed to be known by us.) Further suppose that, for all i, decoding the encoded message using c_i yields "the truth is in A_i", where $A_i \subseteq \Theta$. (Notation: $c_i(X) = A_i$.) Since in addition to the message both the list and the chances of the codes are supposed to be known by us, evidence which is judged to be like the receipt of such a message may be represented by a pair (X,c), where X denotes an encoded message and c is a probability space $(\{c_1,\ldots,c_n\},P_c)$. The evidence (represented by) (X,c) may then be expressed in terms of Dempster-Shafer theory by defining the basic probability assignment $m_{(X,c)}$ as follows:

$$m_{(X,c)}(A) = \sum\{P_c(c_i) \mid c_i(X) = A\}.$$

Indeed, $m_{(X,c)}(A)$ is, in a certain sense, the chance the message was "the truth is in

A" and $Bel_{(X,c)}(A)$, with $Bel_{(X,c)}$ the belief function induced by $m_{(X,c)}$, is the chance that "the truth is in A" is implied by the message. Hence if the coded message is our only evidence, then we will want to call $Bel_{(X,c)}(A)$ our degree of belief that the truth is in A. By choosing the right probability space c, any belief function can be obtained by the above described method. E.g. the evidence X of example 3.3 may be represented by $Bel_{(X,c)}$, where X denotes the message that the ball has label x, c is the probability space $(\{c_1,c_2\},P_c)$, $P_c(c_1) = 2/7$, $P_c(c_2) = 5/7$, $c_1(X) = \{a\}$ and $c_2(X) = \{b\}$.

If (X,c) and (Y,d), with $c = (\{c_1,...,c_n\},P_c)$ and $d = (\{d_1,...,d_m\}, P_d)$, represent bodies of evidence, then the combination of those bodies of evidence may be represented by $((X,Y),(c,d))$, where (X,Y) denotes the conjunction of the encoded messages X and Y and (c,d) is a probability space $(\{(c_i,d_j) \mid 1 \le i \le n,\ 1 \le j \le m\}, P_{((c,d)) \mid (X,Y))})$. (Notice that the probability measure of (c,d) depends on X and Y. The reason for choosing the notation $P_{((c,d)) \mid (X,Y))}$ will be given below.)

5.2 DS-independence

Although still not completely clear, the following reformulation of the requirement that the bodies of evidence to be combined with Dempster's rule must be entirely distinct is somewhat more informative than the original one: "The uncertainties in the arguments being combined,..., must be independent when the arguments are viewed abstractly - i.e., before the interactions of their conclusions are taken into account". ([Shafer [11], p. 49.) The kind of independence informally characterized by the above formulation will be called **DS-independence**.

In Shafer [13] (X,c) and (Y,d) are treated as DS-independent if their combination would be represented by $((X,Y),(c,d))$ with (c,d) the product probability space of c and d, in case X and Y were considered to be messages concerning totally unrelated questions. (Let $c = (\{c_1,...,c_n\},P_c)$ and $d = (\{d_1,...,d_m\}, P_d)$ be probability spaces, then $(\{(c_i,d_j) \mid 1 \le i \le n,\ 1 \le j \le m\}, P)$ is the **product probability space** of c and d if $P(c_i,d_j) = P_c(c_i) \cdot P_d(d_j)$.)

Under this interpretation X and Y of example 3.3 would clearly be DS-independent since the probability that an x-labelled ball B_1 has label a (b) does not depend on whether an y-labelled ball B_2 ($\neq B_1$) has label a or b. But also X and Y of example 2.6 would be DS-independent, since the probability of P_1's fever > 39° C being caused by the flue does not depend on whether the fever > 38.5° C of P_2 is being caused by the flue or not. Therefore we prefer the following related, but more cautious, interpretation of DS-independence:

Definition 5.1 (X,c) and (Y,d) with $c = (\{c_1,...,c_n\},P_c)$ and $d = (\{d_1,...,d_m\}, P_d)$ are called **DS-independent** if for al codes c_i and d_j $P_{(c,d)}(c_i \mid d_j) = P_c(c_i)$ and $P_{(c,d)}(d_j \mid c_i) = P_d(d_j)$, where $P_{(c,d)}$ is the (a priori) probability measure on $(C,D) = \{(c_i,d_j) \mid 1 \le i \le n, 1 \le j \le m\}$ which does not take (the interaction of) X and Y into account and $P_{(c,d)}(c_i)$ is an abbreviation for $P_{(c,d)}(\{(c_i,d_j) \mid 1 \le j \le m\})$.

Example 5.2 Suppose that two (rather untrustworthy) persons X and Y both answer "yes" to the question "is it slippery outside?" and that X answers truthfully 80% of the time, but that the other 20% of the time he is careless and answers "yes" or "no" without taking into account what is actually the case, while Y answers truthfully 70% of the time and carelessly 30% of the time. Let $\Theta = \{\text{slippery, not slippery}\}$. Since a carelessly given "yes" does not provide support for either "slippery" or "not slippery", the answers give rise to the following bpa's: $m_X(\{\text{slippery}\}) = 0.8$, $m_X(\Theta) = 0.2$, $m_Y(\{\text{slippery}\}) = 0.7$ and $m_Y(\Theta) = 0.3$.

According to def. 5.1 above, the answers of X and Y constitute DS-independent bodies of evidence if, viewed abstractly, i.e. before considering the (interaction of the) given answers, the probability of Y being careless is (probabilistically) independent of X either being careless or truthful (i.e. $P(Y \text{ careless}) = P(Y \text{ careless} \mid X \text{ truthful}) = P(Y \text{ careless} \mid X \text{ careless})$). (Cf. [Shafer [12], p.132.) Notice that this does not imply that the answers are independent in the sense that $P(X$ answers "yes" and Y answers "yes"$) = P(X$ answers "yes"$)\cdot P(Y$ answers "yes"$)$. Further notice that if viewed abstractly $P(Y \text{ careless} \mid X \text{ truthful}) = P(Y \text{ careless})$, then this equality does not necessarily remain true after considering the given answers. E.g. if X and Y contradict each other, then they cannot both be truthful and hence $P(Y \text{ careless} \mid X \text{ truthful}) = 1 > 0.2 = P(Y \text{ careless})$.

In definition 5.1 it is presupposed that one can speak of the probability $P_{(c,d)}((c_i ,d_j))$ that a pair of codes (c_i ,d_j) is being used without taking the received messages X and Y into account. In fact, we will assume that the probability measure P_c of c and P_d of d do not depend on the messages X and Y. (Notice that this does not necessarily mean that the probability spaces c and d are independent of X and Y, since the answers might affect the list of possible codes, but it does mean that if the messages X and X' give rise to the same set of possible codes, then the probability of a code being used does not depend on whether X or X' is the message.)

Under this assumption, which is also (implicitly) made in Shafer [13], it makes sense to speak of the a priori probability of a pair of codes. (See section 5.3 for some remarks on the relation between the a priori probability measure $P_{(c,d)}$ and $P_{((c,d)) \mid (X,Y))}$.)

However, this assumption puts rather strong constraints on the kind of evidence that can be compared with a randomly coded message. For instance, it becomes far from obvious whether the coded messages are adequate canonical examples for the bodies of evidence X and Y of example 3.3: let c_a (c_b) be the code translating message X into "the drawn ball has label a (b)" and d_a (d_b) the code translating message Y into "the drawn ball has label a (b)", then the probability of the codes do not seem to be independent of X and Y. In any case, X and Y could not be considered to be DS-independent bodies of evidence, since $P_{(c,d)}(c_a,d_b) = P_{(c,d)}(c_b,d_a) = 0$. (Similar remarks hold for example 2.6.)

Hence we arrive at the same conclusion as Lemmer: Dempster's rule is not applicable in situations like example 3.3. However, we do not base this conclusion on the intuition that $Bel(A)$ ($Pl(A)$) should be interpreted as something like a lower (upper) probability, but on the fact that X and Y cannot be considered to be DS-independent, in spite of all kinds of (irrelevant) independence properties.

5.3 The interaction of evidence

In order to clarify the relation between $P_{((c,d)) \mid (X,Y))}$ and the a priori probability measure $P_{(c,d)}$ we first digress on the meaning of conditional probabilities and on the notion of a partially specified probability measure.

In general, evidence can be evaluated on different levels of abstraction. Therefore, a probability measure P on a frame Θ also has the meaning of a collection of constraints on probability measures on refinements of Θ: If $\omega : 2^{\Theta} \rightarrow 2^{\Omega}$ is a refining and P and P' are probability measures induced by the evidence E on Θ and Ω respectively, then P' is called a **refinement** of P and for all $A \subseteq \Theta$ $P'(\omega(A)) = P(A)$. If Ω is a proper refinement of Θ, i.e. $\Omega \neq \Theta$, then P only *partially* specifes a probability measure on Ω. E.g. $P_{(c,d)}$ partially specifies a probability measure on any coded messages frame, where Ω is a **coded messages frame** if it discerns codes *and* messages, i.e. if there exists a refining ω: $2^{(C,D)} \rightarrow 2^{\Omega}$ such that $(c_i,d_j) \wedge (X,Y) \in \omega(\{(c_i,d_j)\})$.

More generally, any consistent set of probabilities may be interpreted as a partially specified probability measure. (A set S of probabilities is called **consistent** if there exists (i) a frame Θ which is a superset of every set for which a probability is given by S and (ii) a probability measure P on Θ such that P agrees with S. In Van der Gaag [14] these notions of consistency and of a partially specified probability measure are introduced and employed in the context of Boolean algebras.) In particular, this set of probabilities may contain a conditional probability without containing the relevant absolute probabilities.

Traditionally, $P(A \mid B)$ is defined only if (i) A and B are (elements of the σ-field of) subsets of the sample space on which P is defined and (ii) $P(B) > 0$. ($P(A \mid B)$ may then be calculated by means of the well-known formula $P(A \mid B) = P(A \cap B)/P(B)$.) However, one can argue that conditional probabilities are not always obtained from a priori probabilities and that often absolute probabilities are in fact probabilities conditional upon some data for which the probabilities are unknown.

For instance, the probability measure $P_{((c,d)) \mid (X,Y))}$ measures the probabilities of pairs of codes (c_i,d_j) conditional upon the messages X and Y. Intuitively, these probabilities are the same as the conditional probabilities of (c_i,d_j) given (X,Y) measured by a probability measure which does not take (X,Y) into account. Hence $P_{((c,d)) \mid (X,Y))}$ puts some additional constraints on the partially specified probability measure $P_{(c,d)}$ on any coded messages frame: $P_{((c,d)) \mid (X,Y))}((c_i,d_j)) = P_{(c,d)}((c_i,d_j) \mid (X,Y))$. In the following it will be shown that these are indeed *additional* constraints, although in Dempster's rule it is assumed that $P_{((c,d)) \mid (X,Y))}$ is uniquely determined by $P_{(c,d)}$. In other words, the particular way (the interaction of) the evidence is taken into account in Dempster's rule corresponds to an additional assumption next to DS-independence.

6 Assumptions underlying Dempster's rule

If one applies Dempster's rule of combination, then one (implicitly) assumes more than that the bodies of evidence to be combined are DS-independent and that the frame discerns their interaction. In 6.1 this will be shown in the context of the coded messages, while in 6.2 it is argued more informally that the rule is somewhat biassed: probability masses assigned to sets of a relatively large cardinality are in some sense given more weight than probability masses assigned to sets of a lower cardinality.

6.1 Conditions for the validity of Dempster's rule

In this section we give conditions for the validity of Dempster's rule in the context of the randomly coded messages as canonical examples for Dempster-Shafer theory.

Definition 6.1 $Bel_{(X,c)} \oplus Bel_{(Y,d)}$ is called **valid in the context of the coded messages** if for all $A \subseteq \Theta$

$$m_{(X,c)} \oplus m_{(Y,d)}(A) = m_{((X,Y),(c,d))}(A) \quad (= \sum \{P_{((c,d)) \mid (X,Y))}((c_i,d_j)) \mid c_i(X) \cap d_j(Y) = A\}).$$

Definition 6.2 $((X,Y),(c,d))$ is called **simple** if for all A_i,A_j in the core of $Bel_{(X,c)}$ and all B_k,B_l in the core of $Bel_{(Y,d)}$: $A_i \cap B_k = A_j \cap B_l \rightarrow A_i \cap B_k = \varnothing \vee (A_i = A_j \wedge B_k = B_l)$.

Proposition 6.3 Let $((X,Y),((c,d))$ be simple. $Bel_{(X,c)} \oplus Bel_{(Y,d)}$ is valid in the context of the coded messages iff

(1) $P_{((c,d)) \mid (X,Y))}((c_i,d_j)) = 0$ whenever $c_i(X) \cap d_j(Y) = \varnothing$

(2) for all i,j,k,l such that $c_i(X) \cap d_k(Y) \neq \varnothing \neq c_j(X) \cap d_l(Y)$

$$P_{((c,d)) \mid (X,Y))}((c_i,d_k)) : P_{((c,d)) \mid (X,Y))}((c_j,d_l)) = P_c(c_i) \cdot P_d(d_k) : P_c(c_j) \cdot P_d(d_l)$$

The proof is straightforward (see appendix) and shows that even if $((X,Y),(c,d))$ is not simple, then conditions (1) and (2) are sufficient and condition (1) is necessary for the validity of Dempster's rule. And although, in general, (2) is not strictly necessary for the validity of $Bel_{(X,c)} \oplus Bel_{(Y,d)}$ (see example 6.4), conditions (1) and (2) do seem to represent the assumptions underlying Dempster's rule in the context of the coded messages, viz. giving probability 0 to "impossible" pairs of codes and uniformly rescaling the a priori probability of the possible pairs.

Example 6.4 Let $\Theta = \{\alpha,\beta,\gamma\}$ and let (X,c) be given by $c = (\{c_1,c_2\},P_c)$, $P_c(c_1) = P_c(c_2) = 0.5$, $c_1(X) = \{\alpha\}$, $c_2(X) = \Theta$ and (Y,d) by $d = (\{d_1,d_2,d_3\},P_d)$, $P_d(d_1) = P_d(d_2) = P_d(d_3) = 1/3$, $d_1(Y) = \{\alpha,\beta\}$, $d_2(Y) = \{\alpha,\gamma\}$ and $d_3(Y) = \Theta$.
Finally, let $P_{((c,d)) \mid (X,Y))}((c_1,d_1)) = 1/3$, $P_{((c,d)) \mid (X,Y))}((c_1,d_2)) = 0$, $P_{((c,d)) \mid (X,Y))}((c_1,d_3)) = P_{((c,d)) \mid (X,Y))}((c_2,d_1)) = P_{((c,d)) \mid (X,Y))}((c_2,d_2)) = P_{((c,d)) \mid (X,Y))}((c_2,d_3)) = 1/6$.
Then for all $A \subseteq \Theta$ $\quad m_{(X,c)} \oplus m_{(Y,d)}(A) = m_{((X,Y),(c,d))}(A)$, but $P_{((c,d)) \mid (X,Y))}((c_1,d_1)) : P_{((c,d)) \mid (X,Y))}((c_1,d_3)) = 2:1 \neq 1:1 = P_c(c_1) \cdot P_d(d_1) : P_c(c_1) \cdot P_d(d_3)$.

The conditions of proposition 6.3 imply the necessary and sufficient conditions for the validity of Dempster's rule in a concrete example mentioned by Henry E. Kyburg Jr. [6, p. 254]. However, the conditions are not implied by DS-independence. Below we characterize the assumption which is needed in addition to DS-independence.

Definition 6.5 Let P be a (partially specified) probability measure and let $P(E) > 0$. $A_1,\ldots,A_n$ are **equally confirmed by** $\boldsymbol{E}$ if

$$P(A_i)P(A_j \mid E) = P(A_i \mid E)P(A_j) \quad (1 \leq i,j \leq n).$$

Proposition 6.6 Let P be a (partially specified) probability measure and let $P(E) > 0$ and $P(A_i) > 0$ $(1 \leq i \leq n)$. The following are equivalent:

(i) $A_1,\ldots,A_n$ are equally confirmed by E.
(ii) For some λ $\; P(A_i \mid E) = \lambda P(A_i)$ $(1 \leq i \leq n)$.
(iii) $P(E \mid A_i) = P(E \mid A_j)$ $(1 \leq i,j \leq n)$.

Proposition 6.7 Let (X,c) and (Y,d) with $c = (\{c_1,\ldots,c_n\},P_c)$ and $d = (\{d_1,\ldots,d_m\}, P_d)$ be DS-independent. Condition (2) of prop. 6.3 holds iff all (c_i,d_j) with $c_i(X) \cap d_j(Y) \neq \emptyset$ are equally confirmed by (X,Y).

Example 5.2 shows that DS-independence does not imply equal confirmation, since one can imagine the answers of X and Y to be DS-independent without P(X answers "yes" and Y answers "yes" $\mid X$ truthful and Y truthful) being equal to P(X answers "yes" and Y answers "yes" $\mid X$ careless and Y truthful) and to P(X answers "yes" and Y answers "yes" $\mid X$ careless and Y careless). Hence Dempster's rule implicitly employs the often unrealistic assumption that the given answers equally confirm all pairs of codes which are possible given the answers. (The observation that in Dempster's rule the given answers are taken into account in a rather questionable way is also made in [1].)

6.2 Bayesian approximation

In Voorbraak [15] Bayesian approximations of belief functions were introduced for the pupose of achieving possible computational savings in some applications of Dempster-Shafer theory. In this paper they are used to show that Dempster's rule is biassed in some sense.

Definition 6.8 Let *Bel* be a belief function induced by the bpa m. The **Bayesian approximation** $[Bel]$ of *Bel* is induced by the bpa $[m]$ defined by:

$$[m](A) = \frac{\sum_{A \subseteq B} m(B)}{\sum_{C \subseteq \Theta} m(C) \cdot |C|} \quad \text{if } A \text{ is a singleton; otherwise, } [m](A) = 0.$$

The factor $[\Sigma\{m(C) \cdot |C| \mid C \subseteq \Theta\}]^{-1}$ will be called the **Bayesian constant** of *Bel.*

Notice that in general the Bayesian approximation of *Bel* differs from the Bayesian belief function obtained from *Bel* by distributing uniformly all probability mass assigned by m to a subset of Θ over its elements. Probability mass assigned to a subset is not *divided* among its elements, but in some sense assigned *completely* to *all* its elements.

Example 6.9 Let $\Theta = \{a,b,c\}$, $m(\{a\}) = 0.4$ and $m(\{b,c\}) = 0.6$. Then $[m](\{a\}) = 0.4/(0.4 \cdot 1 + 0.6 \cdot 2) = 0.25$ and $[m](\{b\}) = [m](\{c\}) = 0.6/(0.4 \cdot 1 + 0.6 \cdot 2) = 0.375$.

Proposition 6.10

(i) $Bel = [Bel]$ iff *Bel* is Bayesian iff the Bayesian constant is 1.

(ii) If *Bel* and *Bel′* are combinable, then $[Bel] \oplus [Bel'] = [Bel \oplus Bel']$

Corollary 6.11

If $Bel \oplus Bel'$ is Bayesian, then $Bel \oplus Bel' = Bel \oplus [Bel'] = [Bel] \oplus [Bel']$.

Hence Dempster's rule agrees with the assignment of the total probability mass committed to a subset of the frame to all the elements of this subset. Example 6.12 shows that this might give counterintuitive results in case some elements profit more from this assignment than other elements.

Example 6.12 Let $\Theta = \{a,b,c\}$ and $m(\{a\}) = m(\{b,c\}) = m'(\{a,b\}) = m'(\{c\}) = 0.5$.

Then we have $[m](\{a\}) = [m](\{b\}) = [m](\{c\}) = [m'](\{a\}) = [m'](\{b\}) = [m'](\{c\}) = m\oplus m'(\{a\}) = m\oplus m'(\{b\}) = m\oplus m'(\{c\}) = 1/3$. This result is counterintuitive, since intuitively b seems to "share" twice a probability mass of 0.5, while both a and c only have to share once 0.5 with b and are once assigned 0.5 individually.

This counterintuitive result is of course related to the fact that Dempster's rule implicitly assumes that all possible pair of focal elements are equallly confirmed by the combined evidence, while intuitively in example 6.10 $(\{b,c\},\{a,b\})$ is less confirmed than $(\{a\},\{a,b\})$ and $(\{b,c\},\{c\})$.

7 Conclusion

We may conclude that the usual independence notions are inadequate to express the requirements for the use of Dempster's rule of combination. A detailed study of these requirements, and the notion of DS-independence in particular, has shown that the range of applicabiliy of the Dempster-Shafer theory is rather limited, since particular bodies of evidence have to satisfy very strong constraints in order to be called DS-independent. Further it has been shown that applying Dempster's rule involves some additional and often unrealistic assumptions.

Most of these conclusions rely heavily on the assumption that a judgments in terms of Dempster-Shafer theory arises out of a judgment that in a particular situation the evidence may be compared to an abstract canonical example like the receipt of a randomly coded message. This still leaves room for a justification of Dempster-Shafer theory independent of the coded messages. However, we do not believe in the possibility of such u justification, since, as is argued in the last section, Dempster's rule seems to be inherently biassed towards a counterintuitive distribution of probability mass $m(A)$ among the elements of A.

Appendix: propositions and proofs

Proposition 4.5 If $\omega: 2^{\Theta} \rightarrow 2^{\Omega}$ is an ultimate refining and $\omega^{\rho}(A \cap B) = \omega^{\rho}(A) \cap \omega^{\rho}(B)$ whenever A and B are focal elements of Bel_1 based on evidence E_1 and Bel_2 based on evidence E_2 respectively, then Θ discerns the relevant interaction of E_1 and E_2.

Proof

Suppose $\phi: 2^{\Theta} \rightarrow 2^{\Phi}$ is a refining, then there is a refining $\psi: 2^{\Phi} \rightarrow 2^{\Omega}$ such that $\omega = \psi\phi$.

$$\begin{aligned}
\theta \in \phi^{\rho}(A) \wedge \theta \in \phi^{\rho}(B) &\Rightarrow \phi(\{\theta\}) \cap A \neq \emptyset \wedge \phi(\{\theta\}) \cap B \neq \emptyset \\
&\Rightarrow \psi(\phi(\{\theta\}) \cap A) \neq \emptyset \wedge \psi(\phi(\{\theta\}) \cap B) \neq \emptyset \\
&\Rightarrow \psi(\phi(\{\theta\})) \cap \psi(A)) \neq \emptyset \wedge \psi(\phi(\{\theta\})) \cap \psi(B) \neq \emptyset \\
&\Rightarrow \theta \in \omega^{\rho}(\psi(A)) \wedge \theta \in \omega^{\rho}(\psi(B)) \\
&\Rightarrow \theta \in \omega^{\rho}(\psi(A) \cap \psi(B)) \\
&\Rightarrow \psi(\phi(\{\theta\})) \wedge \psi(A) \wedge \psi(B) \neq \emptyset \\
&\Rightarrow \psi(\phi(\{\theta\}) \cap A \cap B) \neq \emptyset \\
&\Rightarrow \phi(\{\theta\}) \cap A \cap B \neq \emptyset \\
&\Rightarrow \theta \in \phi^{\rho}(A \cap B)
\end{aligned}$$

Hence $\phi^{\rho}(A) \cap \phi^{\rho}(B) \subseteq \phi^{\rho}(A \cap B)$. The inclusion $\phi^{\rho}(A \cap B) \subseteq \phi^{\rho}(A) \cap \phi^{\rho}(B)$ is trivial. ∎

Proposition 6.3 Let $((X,Y),((c,d))$ be simple. $Bel_{(X,c)} \oplus Bel_{(Y,d)}$ is valid in the context of the coded messages iff

(1) $P_{((c,d)) \mid (X,Y))}((c_i,d_j)) = 0$ whenever $c_i(X) \cap d_j(Y) = \emptyset$

(2) for all i,j,k,l such that $c_i(X) \cap d_k(Y) \neq \emptyset \neq c_j(X) \cap d_l(Y)$

$$P_{((c,d)) \mid (X,Y))}((c_i,d_k)) : P_{((c,d)) \mid (X,Y))}((c_j,d_l)) = P_c(c_i) \cdot P_d(d_k) : P_c(c_j) \cdot P_d(d_l)$$

Proof

Let $I(A) = \{(i,j) \mid c_i(X) \cap d_j(Y) = A\}$.

only if: Assume $Bel_{(X,c)} \oplus Bel_{(Y,d)}$ is valid in the context of the coded messages.

$m_{(X,c)} \oplus m_{(Y,d)}(\emptyset) = 0$ implies $\sum_{(i,j) \in I(\emptyset)} P_{((c,d)) \mid (X,Y))}(c_i,d_j) = 0$, which implies (1).

To prove condition (2), suppose $c_i(X) \cap d_k(Y) \neq \emptyset \neq c_j(X) \cap d_l(Y)$.

On the one hand, we have $m_{(X,c)} \oplus m_{(Y,d)}(c_i(X) \cap d_k(Y)) = K \cdot m_{(X,c)}(c_i(X)) \cdot m_{(Y,d)} d_k(Y)) = K \cdot P_c(c_i) \cdot P_d(d_k)$, where K is the renormalizing constant. (Here we need the assumption that $((X,Y),((c,d))$ is simple.)

On the other hand, $m_{(X,c)} \oplus \mathrm{m}_{(Y,d)}(c_i(X) \cap d_k(Y)) = P_{((c,d)) \mid (X,Y))}(c_i,d_k)$.

Hence $P_{((X,Y),(c,d))}(c_i,d_k) = K{\cdot}P_c(c_i){\cdot}P_d(d_k)$. Similarly one obtains $P_{((c,d))\,|\,(X,Y))}(c_j,d_l) = K{\cdot}P_c(c_j){\cdot}P_d(d_l)$ and condition (2) follows.

if: Asumme conditions (1) and (2).
(1) implies $\sum_{(i,j)\in I(\varnothing)} P_{((c,d))\,|\,(X,Y))}(c_i,d_j) = 0$, which implies $m_{(X,c)} \oplus m_{(Y,d)}(\varnothing) = 0$.
Let $A \neq \varnothing$. Then $m_{(X,c)} \oplus m_{(Y,d)}(A) = K{\cdot}\sum_{(i,j)\in I(A)} P_c(c_i){\cdot}P_d(d_j)$.
Since $\sum_{A\neq\varnothing} m_{(X,c)} \oplus m_{(Y,d)}(A) = 1$, we have $K{\cdot}\sum_{(i,j)\notin I(\varnothing)} P_c(c_i){\cdot}P_d(d_j) = 1$.
We also have $\sum_{(i,j)\notin I(\varnothing)} P_{((c,d))\,|\,(X,Y))}(c_i,d_j) = 1$.
By condition (2), we may conclude $P_{((c,d))\,|\,(X,Y))}(c_i,d_j) = K{\cdot}P_c(c_i){\cdot}P_d(d_j)$.
Hence $m_{(X,c)} \oplus m_{(Y,d)}(A) = \sum_{(i,j)\,\in\, I(A)} P_{((c,d))\,|\,(X,Y))}(c_i,d_j) = m_{((X,Y),(c,d))}(A)$. ∎

Proposition 6.6 Let P be a (partially specified) probability measure and let $P(E) > 0$ and $P(A_i) > 0$ $(1 \le i \le n)$. The following are equivalent:
(i) $A_1,\dots,A_n$ are equally confirmed by E.
(ii) For some λ $P(A_i \mid E) = \lambda P(A_i)$ $(1 \le i \le n)$.
(iii) $P(E \mid A_i) = P(E \mid A_j)$ $(1 \le i,j \le n)$.

Proof
(ii) ⇒ (i): Let for all i $P(A_i \mid E) = \lambda P(A_i)$. Then for all i and j $P(A_i)P(A_j \mid E) = P(A_i)\lambda P(A_j) = \lambda P(A_i)P(A_j) = P(A_i \mid E)P(A_j)$.
(i) ⇒ (iii): Suppose (*) for all i and j $P(A_i)P(A_j \mid E) = P(A_i \mid E)P(A_j)$.
$P(A_i)P(E \mid A_i) = P(A_i \mid E)P(E)$ (Bayes)
$\Rightarrow P(A_j)P(A_i)P(E \mid A_i) = P(A_j)P(A_i \mid E)P(E)$
$\Rightarrow P(A_j)P(A_i)P(E \mid A_i) = P(A_i)P(A_j \mid E)P(E)$ (*)
$\Rightarrow P(A_j)P(E \mid A_i) = P(A_j \mid E)P(E)$
$\Rightarrow P(A_j)P(E \mid A_i) = P(A_j)P(E \mid A_j)$ (Bayes)
$\Rightarrow P(E \mid A_i) = P(E \mid A_j)$.
(iii) ⇒ (ii): Define $\lambda := P(A_i \mid E)/P(A_i)$ for some i.
$P(E \mid A_i) = P(E \mid A_j) \Rightarrow \lambda P(E) = (P(A_j \mid E)P(E))/P(A_j) \Rightarrow P(A_j \mid E) = \lambda P(A_j)$. ∎

Proposition 6.7 Let (X,c) and (Y,d) with $c = (\{c_1,\dots,c_n\},P_c)$ and $d = (\{d_1,\dots,d_m\}, P_d)$ be DS-independent. Condition (2) of prop. 6.3 holds iff all (c_i,d_j) with $c_i(X) \cap d_j(Y) \neq \varnothing$ are equally confirmed by (X,Y).

Proof

if: $P_{((c,d))\,|\,(X,Y))}((c_i,d_k)) : P_{((c,d))\,|\,(X,Y))}((c_j,d_l))$
$= P_{(c,d)}((c_i,d_k) \mid (X,Y)) : P_{(c,d)}((c_j,d_l) \mid (X,Y))$
$= \lambda \cdot P_{(c,d)}((c_i,d_k)) : \lambda \cdot P_{(c,d)}((c_j,d_l))$ (by assumption of equal confirmation)
$= \lambda \cdot P_c(c_i) \cdot P_d(d_k) : \lambda \cdot P_c(c_j) \cdot P_d(d_l)$ (by assumption of DS-independence)
$= P_c(c_i) \cdot P_d(d_k) : P_c(c_j) \cdot P_d(d_l)$ ($\lambda \neq 0$, since otherwise for all i,j $P_{((X,Y),(c,d))}((c_i,d_j)) = 0$)

only if: Choose i and j such that $P_c(c_i) \cdot P_d(d_j) \neq 0$ and define

$\lambda := P_{((c,d))\,|\,(X,Y))}((c_i,d_j))/(P_c(c_i) \cdot P_d(d_j))$.

(2) implies that for all i,j with $c_i(X) \cap d_j(Y) \neq \varnothing$ $P_{((c,d))\,|\,(X,Y))}((c_i,d_j)) = \lambda \cdot P_c(c_i) \cdot P_d(d_j)$.
Hence, by assumption of DS-independence, for all i,j with $c_i(X) \cap d_j(Y) \neq \varnothing$:

$P_{(c,d)}((c_i,d_j) \mid (X,Y)) = \lambda \cdot P_{(c,d)}((c_i,d_j))$). ∎

Proposition 6.10

(i) $Bel = [Bel]$ iff Bel is Bayesian iff the Bayesian constant is 1.

(ii) If Bel and Bel' are combinable, then $[Bel] \oplus [Bel'] = [Bel \oplus Bel']$

Proof

(i) Bel is $[Bel] \Rightarrow$ ($|A| > 0$ implies $m(A) = [m](A) = 0$) $\Rightarrow$ Bel is Bayesian.
Bel is Bayesian $\Rightarrow$ ($|C| > 1 \rightarrow m(C) = 0$) $\Rightarrow \sum_{C \neq \varnothing} m(C)|C| = \sum_{C \neq \varnothing} m(C) = 1$.
$\sum_{C \neq \varnothing} m(C)|C| = 1 \Rightarrow$ ($|C| > 1 \rightarrow m(C) = 0$) and $[m](A) = \sum_{A \subseteq B} m(B) \Rightarrow Bel = [Bel]$.

(ii) It is clear that if A is not a singleton, then $[m] \oplus [m'](A) = 0 = [m \oplus m'](A)$. Let c (c') denote the Bayesian constant of Bel (Bel') and let k be the renormalizing constant of Bel and Bel'. Then we have:

$$[m] \oplus [m'](\{a\}) = \frac{\sum_{\{a\} = B \cap C} [m](B) \cdot [m'](C)}{\sum_{B \cap C \neq \varnothing} [m](B) \cdot [m'](C)} = \frac{[m](\{a\}) \cdot [m'](\{a\})}{\sum_{b \in \Theta} [m](\{b\}) \cdot [m'](\{b\})}$$

$$= \frac{c \cdot (\sum_{a \in C} m(C)) \cdot c' \cdot (\sum_{a \in D} m'(D))}{\sum_{b \in \Theta} (c \cdot (\sum_{b \in E} m(E)) \cdot c' \cdot (\sum_{b \in F} m'(F)))} = \frac{(\sum_{a \in C} m(C)) \cdot (\sum_{a \in D} m'(D))}{\sum_{b \in \Theta} ((\sum_{b \in E} m(E)) \cdot (\sum_{b \in F} m'(F)))}$$

$$= \frac{\sum_{a \in C \cap D} m(C) \cdot m'(D)}{\sum_{E \cap F \neq \varnothing} m(E) \cdot m'(F) \cdot |E \cap F|} = \frac{\sum_{a \in B} (\sum_{B = C \cap D} m(C) \cdot m'(D))}{\sum_{C \subseteq \Theta} (\sum_{E \cap F = C} m(E) \cdot m'(F) \cdot |E \cap F|)}$$

$$= \frac{k \cdot \sum_{a \in B} (\sum_{B = C \cap D} m(C) \cdot m'(D)}{k \cdot \sum_{C \subseteq \Theta} (\sum_{E \cap F = C} m(\mathrm{E}) \cdot m'(F) \cdot |E \cap F|)} = \frac{\sum_{a \in B} (k \cdot \sum_{B = C \cap D} m(C) \cdot m'(D))}{\sum_{C \subseteq \Theta} (|C| \cdot k \cdot \sum_{E \cap F = C} m(E) \cdot \mathrm{m}'(F))}$$

$$= \frac{\sum_{a \in B} m \oplus m'(B)}{\sum_{C \subseteq \Theta} m \oplus m'(C) \cdot |C|} = [m \oplus m'](\{a\}). \quad ∎$$

References

1. Cohen, M.S., Watson, S.R. and Barrett, E., Alternative theories of inference in expert systems for image analysis, *Technical Report 85-1* (Decision Science Consortium, Falls Church VA, 1985).
2. Dempster, A.P., Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics* **38** (1967) 325-339.
3. Fishburn, P.C., The axioms of subjective probability, *Statistical Science* **1** (1986) 335-358.
4. Gordon, J. and Shortliffe, E.H. (1984) The Dempster-Shafer theory of evidence, in: B.G. Buchanan and E.H. Shortliffe (eds), *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project* (Addison-Wesley, Reading MA, 1984) 272-292.
5. Kanal, L.N. and Lemmer, J.F. (eds), *Uncertainty in Artificial Intelligence* (North-Holland, Amsterdam, 1986) 117-125.
6. Kyburg Jr., H.E., Bayesian and non-Bayesian evidential updating, *Artificial Intelligence* **31** (1987) 271-293.
7. Kyburg Jr., H.E., Representing knowledge and evidence for decision, in: B. Bouchon and R.R. Yager (eds.), *Uncertainty in Knowledge-Based Systems*, (Springer Verlag, Berlin, 1987) 30-40.
8. Lemmer, J.F., Confidence factors, empiricism and the Dempster-Shafer theory of evidence, in: [4] 117-125.
9. Lindley, D.V., The probability approach to the treatment of uncertainty in artificial intelligence and expert systems, *Statistical Science* **2** (1987) 17-24.
10. Shafer, G., *A Mathematical Theory of Evidence* (Princeton University Press, Princeton NJ, 1976).
11. Shafer, G., Constructive probability, *Synthese* **48** (1981) 1-60.
12. Shafer, G., Probability Judgment in Artificial Intelligence, in: [4] 127-135.
13. Shafer, G., Belief functions and possibility measures, in: J.C. Bezdek (ed.), *Analysis of Fuzzy Information. Vol. I* (CRC Press,1987).
14. van der Gaag, L.C., Partially specified probability measures in expert systems, *Report CS-R8842*, (Centre for Mathematics and Computer Science, Amsterdam, 1988)
15. Voorbraak, F., A computationally efficient approximation of Dempster-Shafer theory, *International Journal of Man-Machine Studies* (to appear).
16. Zadeh, L.A., A simple view on the Dempster-Shafer theory of evidence and its implication for the rule of combination, *The AI Magazine* **7** (1986) 85-90.

Logic Group Preprint Series
Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

nr. 1 C.P.J. Koymans, J.L.M. Vrancken, *Extending Process Algebra with the empty process*, September 1985.
nr. 2 J.A. Bergstra, *A process creation mechanism in Process Algebra,* September 1985.
nr. 3 J.A. Bergstra, *Put and get, primitives for synchronous unreliable message passing,* October 1985.
nr. 4 A.Visser, *Evaluation, provably deductive equivalence in Heyting's arithmetic of substitution instances of propositional formulas*, November 1985.
nr. 5 G.R. Renardel de Lavalette, *Interpolation in a fragment of intuitionistic propositional logic,* January 1986.
nr. 6 C.P.J. Koymans, J.C. Mulder, *A modular approach to protocol verification using Process Algebra* , April 1986.
nr. 7 D. van Dalen, F.J. de Vries, *Intuitionistic free abelian groups,* April 1986.
nr. 8 F. Voorbraak, *A simplification of the completeness proofs for Guaspari and Solovay's R*, May 1986.
nr. 9 H.B.M. Jonkers, C.P.J. Koymans & G.R. Renardel de Lavalette, *A semantic framework for the COLD-family of languages,* May 1986.
nr. 10 G.R. Renardel de Lavalette, *Strictheidsanalyse,* May 1986.
nr. 11 A. Visser, *Kunnen wij elke machine verslaan? Beschouwingen rondom Lucas' argument*, July 1986.
nr. 12 E.C.W. Krabbe, *Naess's dichotomy of tenability and relevance,* June 1986.
nr. 13 Hans van Ditmarsch, *Abstractie in wiskunde, expertsystemen en argumentatie,* Augustus 1986
nr. 14 A.Visser, *Peano's Smart Children, a provability logical study of systems with built-in consistency* , October 1986.
nr. 15 G.R.Renardel de Lavalette, *Interpolation in natural fragments of intuitionistic propositional logic,* October 1986.
nr. 16 J.A. Bergstra, *Module Algebra for relational specifications,* November 1986.
nr. 17 F.P.J.M. Voorbraak, *Tensed Intuitionistic Logic,* January 1987.
nr. 18 J.A. Bergstra, J. Tiuryn, *Process Algebra semantics for queues,* January 1987.
nr. 19 F.J. de Vries, *A functional program for the fast Fourier transform,* March 1987.
nr. 20 A. Visser, *A course in bimodal provability logic*, May 1987.
nr. 21 F.P.J.M. Voorbraak, *The logic of actual obligation, an alternative approach to deontic logic,* May 1987.
nr. 22 E.C.W. Krabbe, *Creative reasoning in formal discussion,* June 1987.
nr. 23 F.J. de Vries, *A functional program for Gaussian elimination*, September 1987.
nr. 24 G.R. Renardel de Lavalette, *Interpolation in fragments of intuitionistic propositional logic,* October 1987.(revised version of no. 15)
nr. 25 F.J. de Vries, *Applications of constructive logic to sheaf constructions in toposes*, October 1987.
nr. 26 F.P.J.M. Voorbraak, *Redeneren met onzekerheid in expertsystemen*, November 1987.
nr. 27 P.H. Rodenburg, D.J. Hoekzema, *Specification of the fast Fourier transform algorithm as a term rewriting system,* December 1987.

nr. 28 D. van Dalen, *The war of the frogs and the mice, or the crisis of the Mathematische Annalen,* December 1987.
nr. 29 A. Visser, *Preliminary Notes on Interpretability Logic,* January 1988.
nr. 30 D.J. Hoekzema, P.H. Rodenburg, *Gauß elimination as a term rewriting system*, January 1988.
nr. 31 C. Smorynski, *Hilbert's Programme,* January 1988.
nr. 32 G.R. Renardel de Lavalette, *Modularisation, Parameterisation, Interpolation,* January 1988.
nr. 33 G.R. Renardel de Lavalette, *Strictness analysis for POLYREC, a language with polymorphic and recursive types*, March 1988.
nr. 34 A. Visser, *A Descending Hierarchy of Reflection Principles,* April 1988.
nr. 35 F.P.J.M. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory*, April 1988.
nr. 36 C. Smorynski, *Arithmetic Analogues of McAloon's Unique Rosser Sentences,* April 1988.
nr. 37 P.H. Rodenburg, F.J. van der Linden, *Manufacturing a cartesian closed category with exactly two objects*, May 1988.
nr. 38 P.H. Rodenburg, J. L.M.Vrancken, *Parallel object-oriented term rewriting : The Booleans*, July 1988.
nr. 39 D. de Jongh, L. Hendriks, G.R. Renardel de Lavalette, *Computations in fragments of intuitionistic propositional logic*, July 1988.
nr. 40 A. Visser, *Interpretability Logic,* September 1988.
nr. 41 M. Doorman, *The existence property in the presence of function symbols,* October 1988.
nr. 42 F. Voorbraak, *On the justification of Dempster's rule of combination*, December 1988.