

Dempster's Rule as Seen by Little Colored Balls

Audun Jøsang¹ and Simon Pope²

¹University of Oslo ²Microsoft

Dempster's rule is traditionally interpreted as an operator for fusing belief functions. While there are different types of belief fusion, there has been considerable confusion regarding the exact type of operation that Dempster's rule performs. Many alternative operators for belief fusion have been proposed, where some are based on the same fundamental principle as Dempster's rule, and others have a totally different basis, such as the cumulative and averaging fusion operators. In this article we analyze Dempster's rule from a statistical and frequentist perspective and compare it with cumulative and averaging belief fusion. We prove, and illustrate by examples on colored balls, that Dempster's rule in fact represents a method for serial combination of stochastic constraints. Consequently, Dempster's rule is not a method for cumulative fusion of belief functions under the assumption that subjective beliefs are an extension of frequentist beliefs. Having identified the true nature of Dempster's rule, appropriate applications of Dempster's rule of combination are described such as the multiplication of orthogonal belief functions, and the combination of preferences dictated by different parties.

Key words: Belief theory, Fusion, Cumulative operator, Averaging operator, Subjective logic.

Received 18 Jan. 2010; Revised 22 Dec. 2010; Accepted 26 Apr. 2011; Published online 28 May 2012.

1. INTRODUCTION

Belief theory has its origin in a model for upper and lower probabilities proposed by Dempster in 1960. Shafer later proposed a model for expressing beliefs [Shafer (1976)]. The main idea behind belief theory is to abandon the additivity principle of probability theory, i.e. that the sum of probabilities on all pairwise exclusive possibilities must add up to one. Instead belief theory gives observers the ability to assign so-called belief mass to any subset of the frame of discernment, i.e. non-exclusive possibilities including the whole frame itself. The advantage of this approach is that uncertainty about the probability values, i.e. the lack of evidence to support any specific probability value, can be explicitly expressed by assigning belief mass to the whole frame or to arbitrary subsets of the frame. Shafer's book [Shafer (1976)] describes many aspects of belief theory, but the two main elements are 1) a flexible way of expressing beliefs, and 2) a method for combining beliefs, commonly known as Dempster's rule.

The traditional interpretation of Dempster's rule is that it fuses separate argument beliefs from independent sources into a single belief. There are well known examples where Dempster's rule produces counterintuitive results when interpreted in this way, especially in case of strong conflict between the two argument beliefs. Motivated by this observation, numerous authors have proposed alternative methods for fusing beliefs [Destercke and Dubois (2010); Jøsang *et al.* (2010); Jøsang (2002); Lefevre *et al.* (2002); Daniel (2000); Murphy (2000); Smets (1990); Dubois and Prade (1988); Yager (1987)]. These rules are not only different, they also model different situations. The fact that different situations require different rules is often ignored, and therefore represents a significant source of confusion in the belief theory community.

The importance of correctly interpreting the situation before selecting the appropriate operator can be illustrated as follows. The right operator for modeling the strength of a chain is the principle of the weakest link, and the right operator for modeling the strength of a relay swimming team is the average strength of each member. Applying the weakest swimmer principle to assess the overall strength of the relay swimming team might represent an approximation, but it is incorrect in general, and would give very unreliable predictions. Similarly, applying the principle of average strength of the links in a chain when assessing the overall strength of the chain might represent an approximation, but it is incorrect in general and could be fatal if life depended on it. The observation of these simple examples tells us that it is crucial to properly understand the situation at hand in order to find the correct model for analyzing it.

Corresponding author: Audun Jøsang, email: josang@mn.uio.no

Because Dempster's rule seems to produce counterintuitive results when fusing highly conflicting beliefs, considerable effort has been invested into finding a satisfactory solution to conflict management in belief fusion. However, without trying to interpret the specific situation at hand it is difficult to judge the merit of the various proposed fusion rules. This confusion can be seen as the tragedy of belief theory for two reasons. Firstly, instead of advancing belief theory, researchers have been trapped in the search for a solution to the same problem for 25 years. Secondly, this controversy has given belief theory a bad taste despite its obvious advantages in ignorance and uncertainty management.

The traditional interpretation has been that Dempster's rule provides a method for fusing evidence in cumulative or averaging situations. In this article we investigate the properties of Dempster's rule from a frequentist perspective, and prove that it represents a method of combining belief constraints. The basis for our analysis is to interpret cumulative, averaging and constraining situations in a statistical frequentist sense. More specifically, a cumulative situation is when the fused belief should be derived from the sum of statistical observations. Furthermore, an averaging situation is when the fused belief should be derived from the average of statistical observations. Finally, a constraining situation is when the fused belief should be derived by considering the argument beliefs as stochastic constraints. The operators for cumulative and averaging belief fusion [Jøsang *et al.* (2010)] that are based on this frequentist perspective are described in order to provide a basis for comparison. By assuming that results from the frequentist analysis are also applicable to subjective and non-frequentist situations it emerges that Dempster's rule is a method for fusing belief constraints, not for cumulative or averaging belief fusion. This is illustrated by several examples where colored balls are used to represent frequentist situations.

When Dempster's rule is interpreted as a method for cumulative belief fusion the well known counterintuitive results resulting from combining conflicting beliefs are to be expected. In case of low conflict between the two argument beliefs, Dempster's rule represents an approximation of cumulative belief fusion, but as the level of conflict increases, the accuracy of the approximation decreases. However, we show that even in case of low conflict between the two argument beliefs, Dempster's rule, when interpreted as a method for cumulative belief fusion, can produce quite counterintuitive results.

2. BELIEF THEORY

In this section several concepts of the Dempster-Shafer theory of evidence [Shafer (1976)] are recalled in order to introduce notations used throughout the article. Let $\Theta = \{\theta_i, i = 1, \dots, n\}$ denote a finite set of exhaustive and exclusive possible values for a state variable v of interest¹, where n denotes the set cardinality. A set of this form is traditionally called a frame of discernment, or frame for short. The powerset of Θ , denoted as 2^Θ , represents the set of all subsets of Θ , including Θ itself. A so-called *basic belief assignment* (bba) m on Θ is defined as a function from 2^Θ to $[0, 1]$ satisfying:

$$\sum_{x \subseteq \Theta} m(x) = 1, \quad m(\emptyset) = 0. \quad (1)$$

Values of a bba are called *belief masses*. A bba m such that $m(\emptyset) = 0$ is said to be normal. This condition was originally imposed by Shafer, but is relaxed by some authors [Smets (1990)] under the open-world assumption stating that the set Θ might not be complete and that v might take its value outside Θ . Each subset $x \subseteq \Theta$ such that $m(x) > 0$ is called a focal element of m . A bba m can be equivalently represented by a belief function $bel: 2^\Theta \rightarrow [0, 1]$, defined as

$$bel(x) \triangleq \sum_{\emptyset \neq y \subseteq x} m(y) \quad \forall x \subseteq \Theta. \quad (2)$$

The quantity $bel(x)$ can be interpreted as a measure of one's belief that hypothesis x is true. Note that functions m and bel are in one-to-one correspondence [Shafer (1976)] and can be seen as two facets of the same piece of information. If all the focal elements are singletons (i.e. one-element subsets of Θ) then we speak about *Bayesian belief functions*. If all the focal elements are nestable (i.e. linearly ordered by inclusion) then we speak about *consonant belief functions* [Shafer (1976)]. A *dogmatic belief function* is defined by Smets as a belief function for which $m(\Theta) = 0$. Let us note, that trivially, every Bayesian belief function is dogmatic.

¹Some authors have proposed to allow non-exhaustive sets [Smets (1990)].

3. FREQUENTIST INTERPRETATION OF DEMPSTER'S RULE

Belief theory normally interprets the conjunctive rule and Dempster's rule in a non-frequentist fashion, meaning that bbas represent subjective beliefs and that each instance of combination is a one-off process that combines the two bbas in a single operation unrelated to previous or future combinations. Below we will describe an alternative interpretation of the combination bbas as a series of stochastic constraints, and describe frequentist mathematical models for the combination of such constraints. It turns out that frequentist constraint combinations in their non-normalized and normalized forms are equivalent to the conjunctive rule and Dempster's rule respectively.

Assume two bbas as m^A and m^B that represent beliefs about the truth values of states in a specific frame Θ . The superscripts A and B are attributes that identify the respective belief source or belief owners. These two bbas can be mathematically merged with the conjunctive rule of combination " $\odot$ " denoted as $m^{A\&B} = [m^A \odot m^B]$, or with its normalized version called Dempster's rule " $\ominus$ " denoted as $m^{A\&B} = [m^A \ominus m^B]$. Belief owner combination with $\&$ thus corresponds to the conjunctive bba combination rule $\odot$, and belief owner combination with $\&$ corresponds to Dempster's bba combination rule $\ominus$. The algebraic expression of the conjunctive rule is defined next, followed by that of Dempster's rule.

Definition 1 (The Conjunctive Rule of Combination):

$$\begin{aligned} m^{A\&B}(x) &= [m^A \odot m^B](x) \\ &= \sum_{y \cap z = x} m^A(y) m^B(z) \quad \forall x \subseteq \Theta. \end{aligned} \quad (3)$$

Merging two bbas with the conjunctive rule defined above produces a sub-additive bba, meaning that the sum of belief masses on focal elements can be less than one, in which case it is assumed that the missing or complement belief mass gets assigned to the emptyset. If desirable, the normality assumption $m(\emptyset) = 0$ can be recovered by dividing each belief mass by a normalization coefficient. The resulting operator is Dempster's rule which is defined next.

Definition 2 (Dempster's Rule of Combination):

$$\begin{aligned} m^{A\&B}(x) &= [m^A \ominus m^B](x) \\ &= \frac{[m^A \odot m^B](x)}{1 - [m^A \odot m^B](\emptyset)} \quad \forall x \subseteq \Theta, x \neq \emptyset. \end{aligned} \quad (4)$$

The use of Dempster's rule is mathematically possible only if m^A and m^B are not totally conflicting, i.e., if at least there exists a focal element y of m^A and a focal element z of m^B satisfying $(y \cap z) \neq \emptyset$. This rule is associative, commutative and non-idempotent, and its use has been theoretically justified on the basis of these properties by several authors [Dubois and Prade (1986); Klawonn and Schwecke (1992); Voorbraak (1991)] according to specific axioms. For an algebraic analysis of Dempster's rule applied to a binary frame of discernment see e.g. [Hájek *et al.* (1992); Hájek and Valdés (1991)].

The normalization in Dempster's rule redistributes conflicting belief masses to non-conflicting ones, and thereby tends to eliminate any conflicting characteristics in the resulting belief mass distribution. The so-called "open world assumption" of the conjunctive rule of combination can be applied to avoid this particular problem by allowing all conflicting belief masses to be allocated to the empty set. Smets [Smets (1990)] justifies this interpretation by arguing that the presence of highly conflicting beliefs indicates that some possible event must have been overlooked (the open world assumption) and therefore is missing in the frame of discernment. The idea is that conflicting belief masses should be allocated to this missing (empty) event. Smets has also proposed the much less controversial, and rather obvious interpretation of the belief mass allocated to the empty set simply as a measure of conflict between separate bbas².

The frequentist mathematical expression of the conjunctive rule is described next. Assume a repetitive

²Note that conflicting beliefs can exist inside a bba as non-consonant beliefs. This internal conflict contributes to conflict between separate beliefs when combined with Dempster's rule.

process U which generates instances of unconstrained subsets Θ that are modified by serially arranged belief constraints A and B to produce output subsets. The serial constraining configuration is determined by two argument bbas m^A and m^B in the following way. A specific subset $y \subseteq \Theta$ which is a focal element of m^A is selected with probability equal to $m^A(y)$ as the constraint at A , and a subset $z \subseteq \Theta$ which is a focal element of m^B is selected with probability equal to $m^B(z)$ as the constraint at B . The unconstrained subset Θ from U is then constrained at A by computing $y = (\Theta \cap y)$, which in turn is constrained at B to produce $x = (y \cap z)$. This is illustrated in Fig.1.

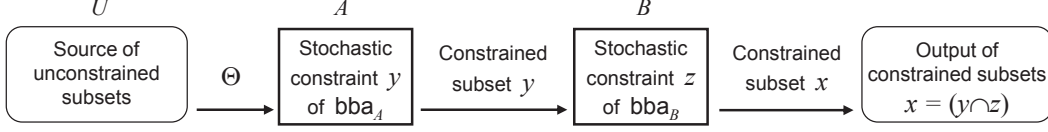


FIGURE 1. Frequentist interpretation of the conjunctive rule

If e.g. for a specific instance i the constraints are such that $(y_i \cap z_i) = x \neq \emptyset$, then the output is x . If for another instance j the constraints are such that $(y_j \cap z_j) = \emptyset$, then the output is $\emptyset$. If for yet another instance k the constraints are $y_k = z_k = \Theta$ so that $(y_k \cap z_k) = \Theta$, then the output is Θ . Relative to the total number of output subsets (including $\emptyset$ and Θ) the proportions of output of subset x can be expressed by a frequentist probability $P^{A\&B}(x)$. Let n denote the number of instances originating from U , and let T denote the binary logic truth function such that $T(\text{TRUE}) = 1$ and $T(\text{FALSE}) = 0$. Then the convergence value of $P^{A\&B}(x)$ when n goes to infinity is expressed by Def.3 below.

Definition 3 (Non-Normalized Serial Stochastic Belief Constraints):

$$P^{A\&B}(x) = \lim_{n \rightarrow \infty} \left(\frac{\sum_{i=1}^n T((y_i \cap z_i) = x)}{n} \right) \quad \forall x \subseteq \Theta. \quad (5)$$

It can be shown that this is the same as the conjunctive rule which leads to the following theorem.

THEOREM 1 (Equivalence with the Conjunctive Fusion Rule):

The conjunctive rule of Def.1 is equivalent to probabilistic non-normalized serial stochastic belief constraints of Def.3, formally expressed as $m^{A\&B}(x) \equiv P^{A\&B}(x)$.

Proof. Transformation of the expression for $P^{A\&B}(x)$ in simple successive steps demonstrates the equivalence.

$$\begin{aligned} 1: \quad P^{A\&B}(x) &= \lim_{n \rightarrow \infty} \left(\frac{\sum_{i=1}^n T((y_i \cap z_i) = x)}{n} \right) & \forall x \subseteq \Theta \\ 2: \quad &= \frac{n(\sum_{y \cap z = x} m_A(y) m_B(z))}{n} & \forall x \subseteq \Theta \\ 3: \quad &= \sum_{y \cap z = x} m^A(y) m^B(z) & \forall x \subseteq \Theta \\ 4: \quad &= [m^A \odot m^B](x) &= m^{A\&B}(x) \quad \forall x \subseteq \Theta \end{aligned} \quad (6)$$

The crucial point is the transformation from step 1 to step 2. The validity of this transformation is evident because the probability of $T((y_i \cap z_i) = x)$ at every instance i is $\sum_{y \cap z = x} m^A(y) m^B(z)$ so that the relative frequency will converge towards this value. $\square$

In order to describe the frequentist interpretation of Dempster's rule, it must be assumed that the constraining process rejects any instance of emptyset $\emptyset$ resulting at B as illustrated in Fig.2.

If e.g. for a specific instance i the constraints are such that $(y_i \cap z_i) = x \neq \emptyset$, then the output is still x . However, if for another instance j the constraints are such that $(y_j \cap z_j) = \emptyset$, then there is no output because $\emptyset$ gets rejected at B . Relative to the total number of output subsets (including Θ but excluding $\emptyset$) the proportions of output of subset x can be expressed by a frequentist probability $P^{A\&B}(x)$. Let n denote the number of instances

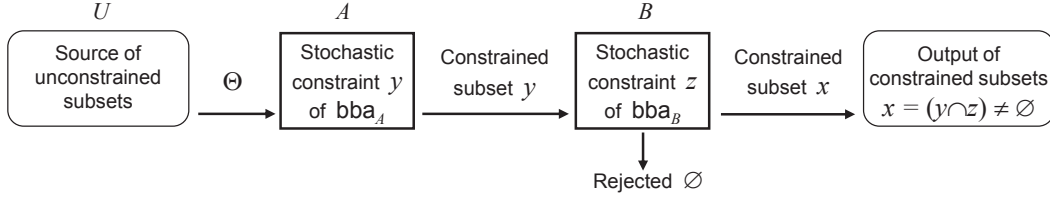


FIGURE 2. Frequentist interpretation of Dempster's rule

originating from U , and let T denote the logic truth function such that $T(\text{TRUE}) = 1$ and $T(\text{FALSE}) = 0$. Then the convergence value of $P^{A\&B}(x)$ when n goes to infinity is expressed by Def.4 below.

Definition 4 (Normalized Serial Stochastic Belief Constraints):

$$P^{A\&B}(x) = \lim_{n \rightarrow \infty} \left(\frac{\sum_{i=1}^n T((y_i \cap z_i) = x)}{n - \sum_{i=1}^n T((y_i \cap z_i) = \emptyset)} \right) \quad \forall x \subseteq \Theta, x \neq \emptyset. \quad (7)$$

It can be seen that this is the same as Dempster's rule which leads to the following theorem.

THEOREM 2 (Equivalence with Dempster's rule):

Dempster's rule of Def.2 is equivalent to probabilistic normalized serial stochastic belief constraints of Def.4, formally expressed as $m^{A\&B}(x) \equiv P^{A\&B}(x)$.

Proof. Transformation of the expression for $P^{A\&B}(x)$ in simple successive steps demonstrates the equivalence.

$$\begin{aligned}
 1: \quad P^{A\&B}(x) &= \lim_{n \rightarrow \infty} \left(\frac{\sum_{i=1}^n T((y_i \cap z_i) = x)}{n - \sum_{i=1}^n T((y_i \cap z_i) = \emptyset)} \right) && \forall x \subseteq \Theta, x \neq \emptyset \\
 2: &= \frac{n(\sum_{y \cap z = x} m^A(y) m^B(z))}{n(1 - \sum_{y \cap z = \emptyset} m^A(y) m^B(z))} && \forall x \subseteq \Theta, x \neq \emptyset \\
 3: &= \frac{\sum_{y \cap z = x} m^A(y) m^B(z)}{1 - \sum_{y \cap z = \emptyset} m^A(y) m^B(z)} && \forall x \subseteq \Theta, x \neq \emptyset \\
 4: &= \frac{[m^A \odot m^B](x)}{1 - [m^A \odot m^B](\emptyset)} = [m^A \odot m^B](x) = m^{A\&B}(x) && \forall x \subseteq \Theta, x \neq \emptyset
 \end{aligned} \quad (8)$$

The crucial point is the transformation from step 1 to step 2. The validity of this transformation is evident because the probability of $T((y_i \cap z_i) = x)$ at every instance i is $\sum_{y \cap z = x} m^A(y) m^B(z)$, and because the probability of $T((y_i \cap z_i) = \emptyset)$ at every instance i is $\sum_{y \cap z = \emptyset} m^A(y) m^B(z)$. $\square$

While Defs.3&4 are based on frequentist situations in the form of long term probabilities of constrained output subsets, the results can be extended to the combination of subjective bbas in the same way that frequentist probability calculus can be extended to subjective and non-frequentist situations. According to [de Finetti (1974b)] a frequentist probability is no more objective than a subjective (non-frequentist) probability, because even if observations are objective, their translation into probabilities is always subjective. [de Finetti (1974a)] provides further justification for this view by explaining that subjective knowledge of a system often will carry more weight when estimating probabilities of future events than purely objective observations of the past. The only case where probability estimates can be purely based on frequentist information is in abstract examples from text books. Frequentist information is thus just another form of evidence used to estimate probabilities. Because a bba is simply a probability distribution over the powerset of a frame, it is natural that de Finetti's view can be extended to belief theory. Based on this argumentation there is not only a mathematical equivalence, but also an interpretational equivalence between the conjunctive rule and non-normalized stochastic belief constraints, as well as between Dempster's rule and normalized stochastic belief constraints.

4. CUMULATIVE AND AVERAGING BELIEF FUSION

In the literature Dempster's rule is often applied to situations that are cumulative or averaging in nature. In order to have a basis for judging the performance of Dempster's rule in such situations, this section briefly describes the cumulative and averaging operators of belief fusion that can be derived from traditional Bayesian analysis. More specifically, the rules are equivalent to the accumulation and averaging respectively of statistical evidence in the multinomial Dirichlet model.

4.1. The Dirichlet Distribution and bba Mapping

Let X be a reduced version of the powerset of the frame of discernment Θ . More precisely, let the reduced powerset X contain only proper subsets of Θ so that it is defined by:

$$X = 2^\Theta \setminus \{\Theta, \emptyset\}, \quad (9)$$

and let k be the cardinality of X . Given that n is the cardinality of Θ , then $k = 2^n - 2$ because $\{\emptyset\}$ and $\{\Theta\}$ are not counted. By considering X as a frame by itself, a belief mass on a proper subset of Θ is at the same time a belief mass on a singleton of X . Similarly, belief mass on Θ is the same as belief mass on X . It can be shown that this bba over X is equivalent to a Dirichlet probability distribution over X [Jøsang *et al.* (2010)]. The Dirichlet distribution $D(\vec{p} | \vec{\alpha})$ over X is a multinomial probability distribution of the probability vector $\vec{p}$ over X based on the evidence parameters $\vec{\alpha}$ about the state elements $x_i \in X$, where $\vec{p}$ satisfies the traditional additivity criterion $\sum_{i=1}^k \vec{p}(x_i) = 1$. The Dirichlet distribution is expressed as:

$$D(\vec{p} | \vec{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^k \vec{\alpha}(x_i)\right)}{\prod_{i=1}^k \Gamma(\vec{\alpha}(x_i))} \prod_{i=1}^k p(x_i)^{(\vec{\alpha}(x_i)-1)} \quad \text{where } \vec{\alpha}(x_i) > 0, \forall x_i \in X. \quad (10)$$

In order to separate between the non-informative prior distribution, and the posterior distribution based on observation evidence, the general evidence parameters $\vec{\alpha}$ must be split into two parts as:

$$\vec{\alpha}(x_i) = \vec{r}(x_i) + W\vec{d}(x_i) \quad \text{where} \quad \begin{cases} \vec{r}(x_i) \geq 0, \forall x_i \\ \vec{d}(x_i) \in [0, 1], \forall x_i \\ \sum_{i=1}^k \vec{d}(x_i) = 1 \end{cases} \quad (11)$$

Here $\vec{r}$ represents the observation evidence, W represents the prior non-informative weight and $\vec{d}$ represents the prior base rates, so that the Dirichlet distribution can be denoted as $D(\vec{p} | \vec{r}, \vec{d})$.

In case of a binary state space $X = \{x, \bar{x}\}$ with default base rates $\vec{d}(x) = \vec{d}(\bar{x}) = 0.5$ it is normally required that the *a priori* non-informative distribution is uniform, which requires that the *a priori* weight $W = 2$. The *a priori* weight W will always be equal to the cardinality of the state space over which a non-informative uniform distribution is assumed. The choice of W determines the influence that observation evidence will have over the prior non-informative distribution. However, as will be shown below the cumulative and averaging operators are not a function of W , which makes the choice of W irrelevant and outside the scope of this discussion.

The parameters of the Dirichlet distribution can be mapped to bbas in a bijective fashion that preserves the probability expectation values [Jøsang and Elouedi (2007)]. This mapping is defined below.

Definition 5 (Mapping between observation evidence and bba):

The cases of an uncertain bba ($m(X) \neq 0$) and a dogmatic bba ($m(X) = 0$) are treated slightly differently.

$$\forall x_i \in X \quad \begin{cases} m(x_i) = \frac{\vec{r}(x_i)}{W + \sum_{i=1}^k \vec{r}(x_i)} \\ m(X) = \frac{W}{W + \sum_{i=1}^k \vec{r}(x_i)} \end{cases} \Leftrightarrow \begin{pmatrix} \text{For } m(X) \neq 0: & \text{For } m(X) = 0: \\ \begin{cases} \vec{r}(x_i) = \frac{Wm(x_i)}{m(X)} \\ 1 = m(X) + \sum_{i=1}^k m(x_i) \end{cases} & \begin{cases} \vec{r}(x_i) = m(x_i) \infty \\ 1 = \sum_{i=1}^k m(x_i) \end{cases} \end{pmatrix} \quad (12)$$

4.2. Cumulative Fusion of Evidence

The cumulative fusion operator is equivalent to *a posteriori* updating of Dirichlet distributions, and is based on the bijective mapping described in the previous section. Assume a process with possible outcomes defined by

the frame Θ , and with the reduced powerset X defined according to Eq.(9) as $X = 2^\Theta \setminus \{\Theta, \emptyset\}$. Let agents A and B make independent observations, e.g. by observing the outcomes of the process over two separate time periods, and assume that they apply the same base rate vector $\vec{d}$ to Θ . The symbol “ $\diamond$ ” denotes the semantic cumulative fusion of two observers A and B into a single imaginary observer denoted as $A \diamond B$.

Definition 6 (Cumulative Fusion Operator):

Let m^A and m^B be bbas respectively held by agents A and B over the same frame Θ . Let k be the cardinality of the reduced powerset $X = 2^\Theta \setminus \{\Theta, \emptyset\}$. The elements of X , which are the proper subsets of Θ , are denoted by x_i so that $X = \{x_i; i = 1, \dots, k\}$. Belief mass on a proper subset $\{x_i\} \subset \Theta$ is the same as belief mass on the corresponding element $x_i \in X$, and belief mass on $\{\Theta\}$ is the same as belief mass on X . Let $m^{A \diamond B}$ be the bba such that:

Case I: For $m^A(\Theta) \neq 0 \vee m^B(\Theta) \neq 0$:

$$\begin{cases} m^{A \diamond B}(x_i) &= \frac{m^A(x_i)m^B(\Theta) + m^B(x_i)m^A(\Theta)}{m^A(\Theta) + m^B(\Theta) - m^A(\Theta)m^B(\Theta)} \\ m^{A \diamond B}(\Theta) &= \frac{m^A(\Theta)m^B(\Theta)}{m^A(\Theta) + m^B(\Theta) - m^A(\Theta)m^B(\Theta)} \end{cases} \quad (13)$$

Case II: For $m^A(\Theta) = 0 \wedge m^B(\Theta) = 0$:

$$\begin{cases} m^{A \diamond B}(x_i) = \gamma_A m^A(x_i) + \gamma_B m^B(x_i) \\ m^{A \diamond B}(\Theta) = 0 \end{cases} \quad \text{where} \quad \begin{cases} \gamma_A = \lim_{\substack{m^A(\Theta) \rightarrow 0 \\ m^B(\Theta) \rightarrow 0}} \frac{m^B(\Theta)}{m^A(\Theta) + m^B(\Theta)} \\ \gamma_B = \lim_{\substack{m^A(\Theta) \rightarrow 0 \\ m^B(\Theta) \rightarrow 0}} \frac{m^A(\Theta)}{m^A(\Theta) + m^B(\Theta)} \end{cases} \quad (14)$$

Then $m^{A \diamond B}$ is the cumulatively fused bba of m^A and m^B , representing the summation of their respective Dirichlet evidence parameters. $m^{A \diamond B}$ can be interpreted as the bba of an imaginary agent $[A \diamond B]$, as if that agent represented both A and B . By using the symbol ‘ $\oplus$ ’ to designate this belief operator, we define $m^{A \diamond B} \equiv m^A \oplus m^B$.

The above defined fusion operator is considered cumulative in the sense of the following theorem.

THEOREM 3 (Equivalence with Statistical Cumulation):

The cumulative fusion operator of Def.6 is equivalent to cumulation of statistical observation evidence under the mapping of Def.5.

Proof. Let the two observers’ respective observations be expressed as $\vec{r}_A$ and $\vec{r}_B$. The Dirichlet distributions resulting from these separate bodies of evidence can be expressed as $D(\vec{p} \mid \vec{r}_A, \vec{d})$ and $D(\vec{p} \mid \vec{r}_B, \vec{d})$. The fusion of these two bodies of evidence simply consists of vector addition of $\vec{r}_A$ and $\vec{r}_B$.

In terms of bbas, the expression for cumulative fusion of Def.6 is derived by mapping the argument bbas to Dirichlet distributions according to Eq.(12), then adding their $\vec{r}$ parameters, and finally mapping the result back to a bba [Jøsang (2007); Jøsang *et al.* (2010)]. $\square$

In Case II, γ_A and γ_B are relative weights satisfying $\gamma_A + \gamma_B = 1$. Their limit values according to Eq.(14) are meaningful when considering the frequentist interpretation of $m(\Theta)$ through the mapping of Eq.(12) in case $\sum_{t=1}^k \vec{r}(x_t) \rightarrow \infty$. Their default values can be set to $\gamma_A = \gamma_B = 0.5$. It can be verified that the cumulative fusion operator is commutative, associative and non-idempotent. The cumulative fusion operator represents a generalization of the consensus operator for independent opinions used in subjective logic [Jøsang (1997, 2002)].

4.3. Averaging Fusion of Evidence

The averaging operator is equivalent to averaging the evidence of Dirichlet distributions, and is based on the bijective mapping between the belief and evidence notations described in Sec.4.1. Assume a process with possible outcomes defined by the frame Θ where X is a reduced powerset defined according to Eq.(9) as $X = 2^\Theta \setminus \{\Theta, \emptyset\}$. Let two sensors A and B make dependent observations, e.g. by observe the same outcomes of the process over the same time period, and assume that they apply the same base rate vector $\vec{d}$ to Θ . The symbol “ $\oslash$ ” denotes the semantic averaging fusion of two observers A and B into a single imaginary observer denoted as $A \oslash B$.

Definition 7 (Averaging Fusion Operator):

Let m^A and m^B be bbas respectively held by agents A and B over the same frame Θ . Let k be the cardinality of the reduced powerset $X = 2^\Theta \setminus \{\Theta, \emptyset\}$. The elements of X , which are the proper subsets of Θ , are denoted by x_i so that $X = \{x_i; i = 1, \dots, k\}$. Belief mass on a proper subset $\{x_i\} \subset \Theta$ is the same as belief mass on the corresponding element $x_i \in X$, and belief mass on $\{\Theta\}$ is the same as belief mass on X . Let $m^{A \oslash B}$ be the bba such that:

Case I: For $m^A(\Theta) \neq 0 \vee m^B(\Theta) \neq 0$:

$$\begin{cases} m^{A \oslash B}(x_i) &= \frac{m^A(x_i)m^B(\Theta) + m^B(x_i)m^A(\Theta)}{m^A(\Theta) + m^B(\Theta)} \\ m^{A \oslash B}(\Theta) &= \frac{2m^A(\Theta)m^B(\Theta)}{m^A(\Theta) + m^B(\Theta)} \end{cases}$$

Case II: For $m^A(\Theta) = 0 \wedge m^B(\Theta) = 0$:

$$\begin{cases} m^{A \oslash B}(x_i) = \gamma_A m^A(x_i) + \gamma_B m^B(x_i) \\ m^{A \oslash B}(\Theta) = 0 \end{cases} \quad \text{where} \quad \begin{cases} \gamma_A = \lim_{\substack{m^A(\Theta) \rightarrow 0 \\ m^B(\Theta) \rightarrow 0}} \frac{m^B(\Theta)}{m^A(\Theta) + m^B(\Theta)} \\ \gamma_B = \lim_{\substack{m^A(\Theta) \rightarrow 0 \\ m^B(\Theta) \rightarrow 0}} \frac{m^A(\Theta)}{m^A(\Theta) + m^B(\Theta)} \end{cases} \quad (15)$$

Then $m^{A \oslash B}$ is the averaged bba of m^A and m^B , representing the average of their respective Dirichlet evidence parameters. By using the symbol “ $\oplus$ ” to designate this belief operator, we define $m^{A \oslash B} \equiv m^A \oplus m^B$.

The above defined fusion operator is considered averaging in the sense of the following theorem.

THEOREM 4 (Equivalence with Statistical Averaging):

The averaging fusion operator of Def.7 is equivalent to averaging of statistical observation evidence under the mapping of Def.5.

Proof. Let the two sensors’ respective observations be expressed as $\vec{r}_A$ and $\vec{r}_B$. The Dirichlet distributions resulting from these separate bodies of evidence can be expressed as $D(\vec{p} | \vec{r}_A, \vec{d})$ and $D(\vec{p} | \vec{r}_B, \vec{d})$. The averaging fusion of these two bodies of evidence simply consists of the average vector value of $\vec{r}_A$ and $\vec{r}_B$.

In terms of bbas, the expression for averaging fusion of Def.7 is derived by mapping the argument bbas to Dirichlet distributions according to Eq.(12), then taking the average of their $\vec{r}$ parameters, and finally mapping the result back to a bba [Jøsang (2007); Jøsang *et al.* (2010)]. $\square$

In Case II, γ_A and γ_B are relative weights satisfying $\gamma_A + \gamma_B = 1$. Their limit values according to Eq.(15) are meaningful when considering the frequentist interpretation of $m(\Theta)$ through the mapping of Eq.(12) in case $\sum_{t=1}^k \vec{r}(x_t) \rightarrow \infty$. Their default values can be set to $\gamma_A = \gamma_B = 0.5$. It can be verified that the averaging fusion operator is commutative and idempotent, but not associative. The averaging fusion operator represents a generalization of the consensus operator for dependent opinions used in subjective logic [Jøsang (1997, 2002)].

So far we have analyzed Dempster’s rule from a frequentist perspective and proven that it is equivalent to the fusion of stochastic belief constraints. We have also described the cumulative and averaging fusion operators based on observing evidence from a frequentist perspective.

The next sections contain examples that describe various situations that can be interpreted as either cumulative, averaging or constraining. The three operators DR (Dempster's Rule), CO (Cumulative Operator) and AO (Averaging Operator) are applied to the examples in order to compare their performance in the various situations.

5. ZADEH'S EXAMPLE

We will begin the study of examples with the well known case that Zadeh [Zadeh (1984)] described with the purpose of criticizing Dempster's rule for the counterintuitive results it produces. In the subsequent example we show that by introducing a small portion of uncertainty, the counterintuitive characteristics seem to disappear.

5.1. The Original Example: Dogmatic Conflicting Beliefs

Suppose that we have a murder case with three suspects; Peter Black, Paul Grey and Mary White, and two witnesses W_A and W_B who give highly conflicting testimonies. The example assumes that the most reasonable conclusion about the likely murderer can be obtained by fusing the beliefs expressed by the two witnesses using Dempster's rule. Table 1 gives the witnesses' belief masses in Zadeh's example and the resulting belief masses after applying Dempster's rule. Some values have been rounded off in this and in the subsequent examples.

TABLE 1. Zadeh's example with Dempster's Rule (DR), the Cumulative Operator (CO) and the Averaging Operator (AO)

		Input evidence of:		Results of applying:		
		Witness A	Witness B	DR	CO	AO
$m(\text{Peter Black})$	=	0.99	0.00	0.00	0.495	0.495
$m(\text{Paul Grey})$	=	0.01	0.01	1.00	0.010	0.010
$m(\text{Mary White})$	=	0.00	0.99	0.00	0.495	0.495

In case of DR, the surprising result is that Paul Grey is identified as the guilty although none of the witnesses thinks he is the murderer. Dempster's rule thus does a poor job in fusing the two beliefs in this case. The results of CO and AO are equal in this case because the input belief contains no uncertainty, but they produce what intuition would predict. In our opinion the most appropriate fusion rule in this case is AR, because it is about weighing the evidence of the two witnesses against each other.

There are other types of situations that can be correctly modeled with Dempster's rule and where it produces perfectly normal and intuitive results when fed with exactly the same belief arguments as those used in Zadeh's example. More specifically these are situations of serial stochastic belief constraints, as described in Def.4.

Assume a special type of balls that can take three different colors; black, white and gray. The balls can be unconstrained, in which case they have all the three colors simultaneously. The balls can also be constrained, in which case they only have a subset of the colors, i.e. two or one single colour.

An initial process generates unconstrained balls that are being fed into two serially connected stochastic constraints. Constraint A will receive the unconstrained balls first. For each received ball, Constraint A will decide with given probabilities whether the ball shall be constrained to be a black ball, or a gray ball. Each ball is then sent to Constraint B. For each received ball, Constraint B will decide with given probabilities whether the ball shall be constrained to be a white ball, or a gray ball. When the incoming ball happens to be black and Constraint B decides that it shall be white, the ball is rejected because a black ball can not be constrained to be white. Likewise, when the incoming ball happens to be gray, and Constraint B decides that it shall be white, the ball is also rejected. Only when the incoming ball is gray and Constraint B decides that it shall be gray is the ball accepted. Constraints A and B correspond to witnesses A and B in the original Zadeh's example. This situation is illustrated in Fig.3 below.

The result of this constraint process is as would be expected, i.e. only gray balls can be accepted by the serial conjunction of the two constraints, so that 100% of the accepted balls are gray.

The degree of conflict in this example is very high, and is equal to the rejection rate of Constraint B in Fig.3 which is 0.9999.

Dempster's rule provides an exact model for this situation, and of any other situation of the same type,

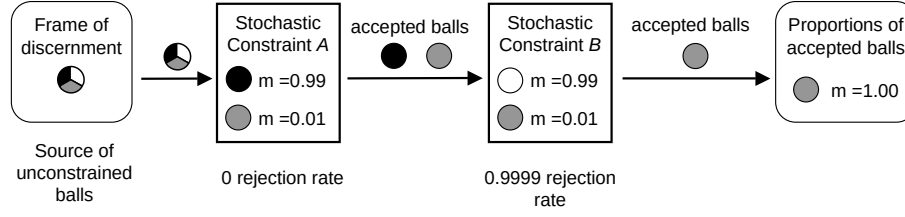


FIGURE 3. Zadeh's example as seen by colored balls

where the two constraints are defined as beliefs. This indicates that Dempster's rule should be interpreted as an operator for serial conjunction of constraints.

5.2. The Modified Example: Introducing uncertainty

In order to illustrate that Dempster's rule, when interpreted as a method for cumulative belief fusion, can produce relatively intuitive results by reducing the level of conflict, the following example modifies Zadeh's example by allocating a small amount of belief mass to the set $\Theta = \{\text{Peter, Paul, Mary}\}$. This has the effect of slightly reducing the level of conflict. Table 2 gives the modified bbas and the results of applying Dempster's rule.

TABLE 2. Results of DR, CO and AO after introducing uncertainty in Zadeh's example

		Input evidence of:		Results of applying:		
		Witness A	Witness B	DR	CO	AO
$m(\text{Peter Black})$	=	0.98	0.00	0.490	0.49246	0.49
$m(\text{Paul Grey})$	=	0.01	0.01	0.015	0.01005	0.01
$m(\text{Mary White})$	=	0.00	0.98	0.490	0.49246	0.49
$m(\Theta)$	=	0.01	0.01	0.005	0.00503	0.01

In this situation it seems that DR rule gives a more intuitive result than in the previous example, i.e. the results give equally strong beliefs in the guilt of Peter Black and Mary White, and only very little belief in Paul Grey being the murderer. However, by recognizing that fusion of witness evidence is most closely modeled as a case of averaging evidence, AO should in principle be best suited for, and the results of applying AO from Table 2 also seems to give the most intuitive result.

When translating this situation into a series of stochastic constraints, we see that Dempster's rule again provides an exact model. This is illustrated in Fig.4.

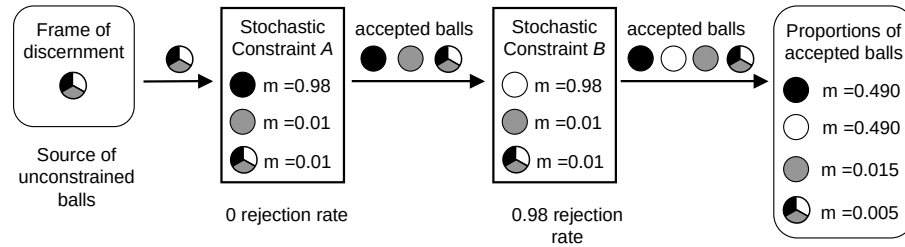


FIGURE 4. The modified Zadeh's example as seen by colored balls

Whereas the results in this example are inconclusive regarding the suitability of applying Dempster's rule in the modified Zadeh's example, there is no doubt that it produces correct results when the argument beliefs are interpreted as stochastic constraints.

The conclusion to be drawn from this is that Dempster's rule seems to provide an approximation of cu-

mulative belief fusion in case the level of conflict is moderate or low. An interesting question is how good this approximation is.

6. OTHER EXAMPLES

6.1. Example: Fusing Equal Dogmatic Sensor Outputs

Assume that a manufacturing process can produce black and white balls, and that two sensors observe the likelihood with which black and white balls are produced. The frame of discernment has two elements: $\Theta = \{\text{Black}, \text{White}\}$ and the powerset three elements: $2^\Theta = \{\text{Black}, \text{White}, \{\text{Black} \cup \text{White}\}\}$. For the purpose of independence it can be assumed that the sensors observe the manufacturing process at different time periods. In Sec.7 we show that there in fact are two different types of independence to consider, where observation independence can be called *external independence*.

Assume that Sensor A and Sensor B produce equal measures about the colors of the balls in the form of the dogmatic beliefs in Table 3 below:

TABLE 3. Results of combining equal dogmatic beliefs from two sensors

		Input evidence of:		Results of applying:		
		Sensor A	Sensor B	DR	CO	AO
$m(\text{Black})$	=	0.99	0.99	0.999898	0.99	0.99
$m(\text{White})$	=	0.01	0.01	0.000102	0.01	0.01
$m(\text{Black} \cup \text{White})$	=	0.00	0.00	0.000000	0.00	0.00

The beliefs expressed by each sensor are thus identical and dogmatic. Despite being equal, the two beliefs would in fact be in slight conflict from the point of view of Dempster's rule because they are non-consonant and have internal conflict.

Let this situation be interpreted as an accumulation of observations. The cumulative and averaging operators produce the average belief, identical to Sensor A's and Sensor B's beliefs, which is exactly what one would expect. When applying Dempster's rule however, the results are very different and rather counterintuitive. In this situation it is therefore correct to apply the cumulative operator, and it would be wrong to apply Dempster's rule.

6.2. Example: Conjunction of Equal Dogmatic Constraints

In this example we will use the same bba values as in Table 3, but describe a situations where it would be correct to apply Dempster's rule.

Assume that a manufacturing process producing balls that can be configured to be either black or white. Initially, the balls can be either, but once they have been configured, they can not be reconfigured. Two configuration stations are implemented as stochastic constraints and called Constraint A and Constraint B. They are connected in series and take the unconstrained balls from the manufacturing process. The configuration occurs stochastically, where Constraint A and Constraint B configure the products according to the beliefs of Table 3. This situation, which is illustrated in Fig.5, represents conjunctive combination of equal dogmatic constraints.

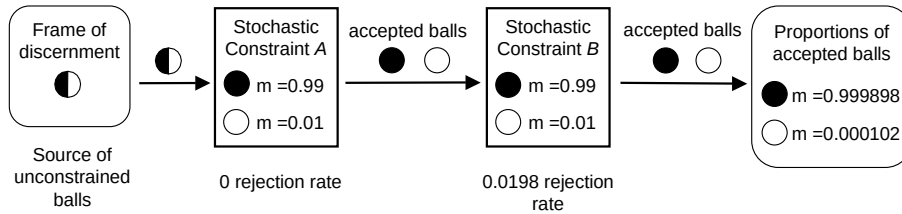


FIGURE 5. Equal constraints over a binary state space combined with Dempster's rule

Dempster's rule correctly models this situation. The level of conflict, expressed by the rejection rate of

Constraint B , is relatively low, and the relative proportion of black balls after two constraints is ≈ 100 times higher than after a single constraint.

6.3. Example: Fusing Non-Conflicting Equal Sensor Outputs

The following example is similar to the previous, but instead of the two sensors assigning belief mass $m = 0.01$ to “White”, this belief mass is instead assigned to “Black $\cup$ White”, and thus introduces uncertainty. The two sensors thus output equal and non-conflicting beliefs given in the Table 4 below:

TABLE 4. Results of combining equal non-conflicting beliefs from two sensors

		Input evidence of:		Results of applying:		
		Sensor A	Sensor B	DR	CO	AO
$m(\text{Black})$	=	0.99	0.99	0.9999	0.995	0.99
$m(\text{White})$	=	0.00	0.00	0.0000	0.000	0.00
$m(\text{Black} \cup \text{White})$	=	0.01	0.01	0.0001	0.005	0.01

Applying Dempster’s rule and the cumulative operator to these beliefs results in a reduction in uncertainty, and a convergence towards the largest belief mass of the sensor outputs. In case of Dempster’s rule, the uncertainty is reduced to the product of the input uncertainties computed as:

$$m^{A \& B}(\text{Black} \cup \text{White}) = m^A(\text{Black} \cup \text{White})m^B(\text{Black} \cup \text{White}) = 0.0001 \quad (16)$$

i.e. by a factor of $1/100$ which represents a very fast convergence. This rate is too fast to be the result of cumulative belief fusion. When fusing two equal beliefs, one should expect the uncertainty to be reduced by $1/2$, and not by a factor of $1/100$.

In case of the cumulative operator, the input beliefs are equivalent to each sensor having observed 198 black balls ($r(\text{Black}) = 198$), and no white balls ($r(\text{White}) = 0$), with the uncertainty computed as $m^{A \& B}(\text{Black} \cup \text{White}) = 2/(r(\text{Black}) + r(\text{White}) + 2) = 0.01$. The output beliefs is equivalent to the observation of $2r(\text{Black}) = 396$ black balls, with the uncertainty computed as $m(\text{Black} \cup \text{White}) = 2/(2r(\text{Black}) + 2) = 0.005025$. The output uncertainty of the cumulative operator is thus halved, which is what one would expect in the case of cumulative fusion.

The difference in convergence i.e. the rate of uncertainty reduction, is significant, and clearly illustrates that Dempster’s rule in fact is not a very good approximation of cumulative belief fusion, even in the case of non-conflicting beliefs.

6.4. Example: Conjunction of Non-Conflicting Equal Constraints

This example is similar to the one in Sec.6.2, but this time the constraints are non-conflicting due to the introduction of uncertainty according to Table 4. The lack of conflict means that there are no rejections. The serial combination of constraints results in a significant increase in the proportion of black balls in the last stage, which also was to be expected. This situation, which is illustrated in Fig.6, is perfectly modeled by Dempster’s rule.

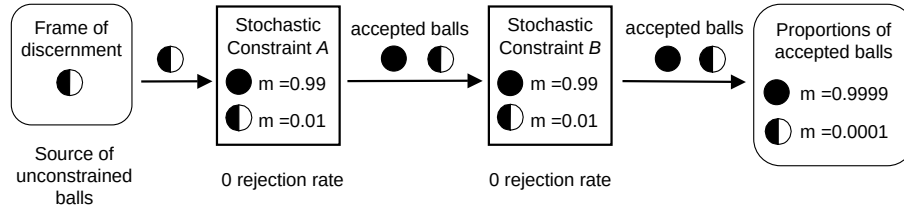


FIGURE 6. Non-conflicting constraints over a binary state space combined with Dempster’s rule

The rejection rate is zero because of the lack of conflict. However, the convergence rate is almost the same as in the example of Sec.6.2 before the uncertainty was introduced. This example thus shows that introducing

uncertainty does not have the effect of making the result more intuitive, as it did in the case of the modified Zadeh's example of Sec.5.2.

7. APPLICATIONS OF DEMPSTER'S RULE OF COMBINATION

The conjunctive rule and Dempster's rule of combination represent an exact model of serial combination of stochastic belief constraints, in the non-normalized and the normalized form respectively. The examples described in the previous sections to illustrate that fact are rather artificial, and do not represent any practical situation if immediate usefulness. This section briefly describes practical situations where the application of Dempster's rule and the conjunctive rule would be correct and useful.

7.1. Multiplication of Orthogonal Belief Functions

Several authors have applied Dempster's rule for a different purpose than belief fusion, for example for combining orthogonal hints [Haenni (2003); Kohlas and Monney (1995)] as described in the Theory of Hints [Haenni (2003)]. A hint can be interpreted as a constraint as described in Defs.1 & 2, so that the Theory of Hints is compatible with interpreting the conjunctive rule and Dempster's rule as a method for serial constraints. In that sense it would be useful to identify the exact situations where the conjunctive rule and Dempster's rule of combination can form the basis of a "Theory of Hints" and how this theory can be interpreted. The examples of hints described in the literature all represent cases of orthogonal bbas defined next.

Definition 8 (Orthogonal bbas):

Let X and Y be two separate frames of cardinality k and l respectively, where $x_i \subseteq X, i = 1, \dots, (2^k - 1)$, and where $y_j \subseteq Y, j = 1, \dots, (2^l - 1)$, i.e. where x_i and y_j denote subsets of X and Y respectively. Subsets of the Cartesian product $X \times Y$ can be denoted as tuples (x_i, y_j) . Let $m_{X\perp}$ denote a bba on $X \times Y$ with focal elements (x_i, Y) and let $m_{Y\perp}$ denote a bba on $X \times Y$ with focal elements (X, y_j) . Then $m_{X\perp}$ and $m_{Y\perp}$ are orthogonal.

It can be proven that a pair of orthogonal bbas is constrained in such a way that the bbas can never be conflicting when combined with the conjunctive rule of combination.

THEOREM 5 (Non-Conflict of Orthogonal bbas): Let X and Y be two separate frames and let $m_{X\perp}$ and $m_{Y\perp}$ be a pair of orthogonal bbas on the product frame $X \times Y$, then $[m_{X\perp} \odot m_{Y\perp}](\emptyset) = 0$.

Proof. Since $m_{X\perp}$ and $m_{Y\perp}$ are a pair of orthogonal bbas, the respective focal elements are represented as tuples (x_i, Y) and (X, y_j) . These focal elements are formed in such a way that their intersection is never empty, i.e. $(x_i, Y) \cap (X, y_j) \neq \emptyset$. For this reason, the conjunctive rule of combination from Def.1 will not assign belief mass to $\emptyset$, which means that a pair of orthogonal bbas are always non-conflicting. $\square$

A bijective mapping can be defined between a pair of orthogonal bbas and a pair of separate bbas on their separate factor frames. Let $m_{X\perp}$ and $m_{Y\perp}$ be a pair of orthogonal bbas on the product frame $X \times Y$, and let m_X be a bba on frame X and let m_Y be a bba on frame Y . We can define the following bijective mapping:

$$\text{Bijective mapping: } \begin{cases} m_X(x_i) = m_{X\perp}(x_i, Y) \\ m_Y(y_j) = m_{Y\perp}(X, y_j) \end{cases} \quad (17)$$

With the mapping of Eq.(17) it can be shown that the application of the conjunctive rule of combination to a pair of orthogonal bbas is equivalent to belief multiplication of bbas on the corresponding factor frames. Multiplication (product) of bbas is defined first.

Definition 9 (Multiplication of bbas):

Let $X \times Y$ be the Cartesian product of frames X and Y , where product elements are denoted as the ordered tuples $(x_i, y_j) \in X \times Y$. Let m_X and m_Y be bbas on X and Y respectively. The product of m_X and m_Y , denoted as $[m_X \cdot m_Y]$, is defined as:

$$[m_X \cdot m_Y](x_i, y_j) = m_X(x_i) \cdot m_Y(y_j) \quad \forall x_i \subseteq X, y_j \subseteq Y. \quad (18)$$

The following are properties of belief multiplication.

- The cardinality of the product frame of discernment is the product of the cardinalities of the factor frames.
- Multiplication is commutative.
- Multiplication is associative.
- The number of focal elements in the product frame is the product of the number of focal elements in the factor frames.

It can be noted that, given m_X and m_Y , the number of possible focal elements in the product frame $X \times Y$ is less than the number of elements in the powerset of $X \times Y$. More specifically

$$\begin{aligned} \text{Number of possible focal elements in product frame} &= (2^{|X|} - 1) \cdot (2^{|Y|} - 1) \\ \text{Number of elements in powerset} &= 2^{(|X| \cdot |Y|)} - 1 \end{aligned} \quad (19)$$

For example, let X and Y both be binary, i.e. $|X| = |Y| = 2$, then the number of possible focal elements in the product frame is 9, and the number of elements in the powerset of the product frame is 15. This proves that there will always be non-focal elements in the powerset of product frame, as already pointed out in [Jøsang and McAnally (2004)]. The equivalence between the conjunctive combination of orthogonal bbas and the multiplication of bbas on factor frames can be stated as a theorem.

THEOREM 6 (Multiplicative equivalence of conjunctive combination of orthogonal bbas):

Let $m_{X\perp}$ and $m_{Y\perp}$ be a pair of orthogonal bbas on the product frame $X \times Y$, and let m_X be a bba on frame X and let m_Y be a bba on frame Y . Assuming the bijective mapping of Eq.(17), then $[m_X \cdot m_Y] = [m_{X\perp} \odot m_{Y\perp}]$.

Proof. Let (x_i, y_j) denote the elements of $X \times Y$. According to Eq.(3), the conjunctive combination of $m_{X\perp}$ and $m_{Y\perp}$ is expressed as

$$[m_{X\perp} \odot m_{Y\perp}](x_i, y_j) = m_{X\perp}(x_i, Y) \cdot m_{Y\perp}(X, y_j). \quad (20)$$

According to the bijective mapping of Eq.(17) the multiplication of belief masses $m_X(x_i)$ and $m_Y(y_j)$ is expressed as:

$$\begin{aligned} [m_X \cdot m_Y](x_i, y_j) &= m_X(x_i) \cdot m_Y(y_j) \\ &= m_{X\perp}(x_i, Y) \cdot m_{Y\perp}(X, y_j) \\ &= [m_{X\perp} \odot m_{Y\perp}](x_i, y_j). \end{aligned} \quad (21)$$

This is valid for all (x_i, y_j) so that $[m_X \cdot m_Y] = [m_{X\perp} \odot m_{Y\perp}]$ in general. $\square$

There are many practical situations where multiplication of bbas could be applied. In the case of two separate frames of discernment such as e.g. $\Theta_{\text{Gender}} = \{\text{male}, \text{female}\}$ and $\Theta_{\text{Species}} = \{\text{cat}, \text{dog}\}$ it is possible to construct the product frame $\Theta_{\text{Gender} \times \text{Species}} = \{\text{male cat}, \text{female cat}, \text{male dog}, \text{female dog}\}$, and corresponding orthogonal bbas can easily be expressed.

There are also situations where a frame of discernment does not have any apparent product structure, such as e.g. the frame of discernment consisting of the possible outcomes of throwing a dice, i.e. $\Theta_{\text{Dice}} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. Depending on the nature of the applicable beliefs, it might be possible to construct an *overlay product frame* on top of the original frame. The elements of the original frame can then be subsets of the elements of the overlay product frame. In order to preserve consistency, each element of the original frame can only be assigned to one product element in the overlay product frame, and each product element must get assigned to it at least one element from the original frame.

Defining an overlay product frame requires that the bbas are related to separate *aspects* of the elements of the original frame. Two possible aspects of the dice frame can e.g. be *parity* (i.e. whether a number is odd or even) and *primality* (i.e. whether a number is non-prime or prime). It is now possible to consider the separate frames $\Theta_{\text{Parity}} = \{\text{odd}, \text{even}\}$ and $\Theta_{\text{Primality}} = \{\text{non-prime}, \text{prime}\}$, and create the product frame:

$$\Theta_{\text{Parity} \times \text{Primality}} = \{(\text{odd}, \text{non-prime}), (\text{odd}, \text{prime}), (\text{even}, \text{non-prime}), (\text{even}, \text{prime})\}.$$

This product frame can be used as an overlay on top of the dice frame, by making the elements of the dice frame subsets of the elements of the overlay product frame. This overlay product frame, which has 4 product elements, is illustrated in Table 5.

In order to illustrate the application of the conjunctive rule in this example, consider the following situation. An informant has special knowledge about a loaded dice, and an observer is trying to predict the outcome of throwing the dice based on hints from the informant.

TABLE 5. Example overlay product frame on the dice frame

	Primality	
	non-prime (gray)	prime (cross pattern)
Parity		
odd (black)	{1}	{3}, {5}
even (white)	{4}, {6}	{2}

First the informant provides hint A which says that the dice will always produce an even number. The observer translates this into the bbas defined by

$$m^A(\{2\}, \{4\}, \{6\}) = m_{\text{Parity}}^A(\text{even}) = 1. \quad (22)$$

Then the informant provides hint B which says that the dice will always produce a prime number. The observer translates this into the bbas defined by

$$m^B(\{2\}, \{3\}, \{5\}) = m_{\text{Primality}}^B(\text{prime}) = 1. \quad (23)$$

The observer notices that the two bbas are orthogonal in the overlay product frame $\Theta_{\text{Parity} \times \text{Primality}}$. He therefore applies the conjunctive rule of combination to reach the conclusion

$$[m^A \odot m^B](\{2\}) = m_{\text{Parity}}^A(\text{even}) \cdot m_{\text{Primality}}^B(\text{prime}) = 1, \quad (24)$$

i.e. that the dice can only produce 2's, which is the only number which is both even and prime. This example is illustrated with colored balls in Fig.7, where the colour codes are given in Table 5.

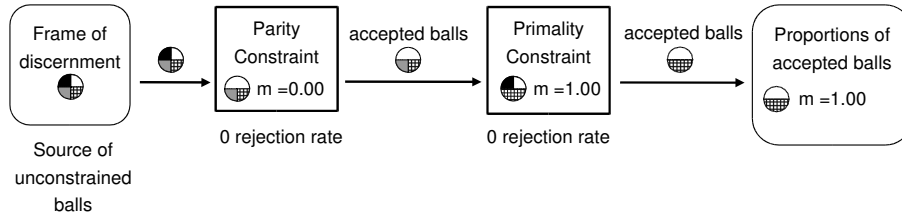


FIGURE 7. Example of applying the conjunctive rule on orthogonal bbas.

The conjunctive rule (as well as Dempster's rule) produces the correct result in this situation. As already pointed out e.g. in [Haenni (2003)], cases of orthogonal bbas on product frames represent practical situations where Dempster's rule can be applied. Equivalently, the bbas can be defined on the original separate frames, and multiplied according to multiplication operator of Def.9.

The principles of external and internal independence are important to understand in order to apply Dempster's, or the conjunctive rule of combination, correctly. Assume for example two sensors/observers that produce beliefs relating to a given frame of discernment. *Internal independence* means that the two sensors/observers measure different aspects of the elements of the frame, so that the bbas produced by the two sensors/observers are orthogonal. *External independence* means that observations are independent because they are made over two different time periods (either by a single sensor/observer, or by different sensors/observers), or alternatively because the observed bbas relate to different aspects of the frame elements. External independence must always be assumed in order for Dempster's rule and the cumulative operator to be meaningful. However, the averaging operator can be applied in case of external dependence.

The application of Dempster's rule to internally independent bbas does not cause any conflict, so that normalization is never required. In this case, Dempster's rule is equivalent to the conjunctive rule of combination as well as to multiplication of bbas on separate frames. In other words, conjunctive belief fusion is equivalent to belief product in this case.

The application of Dempster's rule to internally dependent bbas may or may not cause conflict. In any case, conjunctive belief fusion in the form of Dempster's rule now represents a method for combining stochastic constraints, which must not be confused with cumulative belief fusion.

Since product frames are formed by the Cartesian product of two separate frames, the cardinality of the product frame is the product of the cardinalities of each separate frame. We do not consider trivial frames

consisting of only one element, so the cardinality of a product frame must therefore be at least 4, and can never be prime. It can be noted that the frame in Zadeh's example is ternary, so that it would not be possible to define an overlay product frame without creating a situation in which at least one of the product elements is empty. It is therefore impossible to define an overlay product frame with orthogonal bbas in Zadeh's example. As a consequence, the bbas in Zadeh's example must always be considered as internally dependent.

7.2. Combination of Preferences

The application of serial constraints as described in Def.4 of the Dempster's rule can be interpreted as the combination of preferences. The following example uses the same bba as in Zadeh's example described in Sec.5. Assume two friends, *A* and *B*, who want to see a film together at the cinema one evening, and that there are only three films showing, called *Black*, *Grey* and *White*. The friends express their preferences in the form of the bbas of Table 1. By applying Dempster's rule they conclude that the only film they are both interested in seeing is *Grey*. This is clearly a constraining situation so that Dempster's rule provides the appropriate way of modeling this situation. Had they expressed their preference in the form of the bbas of the modified Zadeh's example of Table 2, the conclusion would have been that the friends should pick film *Black* or *White* because one of them prefers it, and the other finds it acceptable, whereas none of the friends prefer *Grey*, they only find it acceptable. In fact, all the other examples produce perfectly intuitive results when interpreted as the combination of preferences. Even in the case when their focal sets are non-overlapping, the result is intuitive, namely that they can not agree on which film to see, so that they will not go to the cinema. This indicates that Dempster's rule, or alternatively the conjunctive rule, is the appropriate rule for combining this type of preferences. In subjective logic the *belief constraint operator* [Jøsang (2011); Jøsang and Hankin (2012)], which is based on Dempster's rule, is used for modeling situations of preference combination.

8. CONCLUSION

We have proven and illustrated with examples of colored balls that Dempster's rule represents an exact model for serial combination of stochastic constraints. For this reason, Dempster's rule is not a correct model for cumulative or averaging belief fusion. This conclusion is based on the assumption that results from the frequentist analysis of Dempster's rule as well as of the cumulative and averaging fusion operators are also applicable to subjective and non-frequentist situations, in line with the views of authors such as de Finetti.

An explanation for why Dempster's rule has been confused with belief fusion in the belief theory community could be that it produces an approximation of cumulative/averaging belief fusion in case of low conflict. Several authors have therefore tried to argue that it is wrong to specify highly conflicting belief arguments [Halpern and Fagin (1992); Haenni (2005)]. However, such approximative operators are only warranted when they produce sufficiently good results and when they offer significant advantages over alternative methods. We have shown that even in the case of non-conflicting beliefs, Dempster's rule can be a relatively bad approximation, and provides no advantage over the cumulative and averaging fusion operators. Dempster's rule should therefore never be used as a method for cumulative or averaging belief fusion.

We have defined the concepts of product frames and orthogonal bbas. When applied to orthogonal bbas Dempster's rule is equivalent to the conjunctive rule of combination as well as to multiplication of bbas. There are many practical situations where orthogonal bbas can be defined, and because normalization is not needed, the conjunctive rule can be applied instead of Dempster's rule. We have also shown that the conjunctive rule is appropriate for fusing preferences from different parties when the preferences are expressed as a form of constraints.

The arguments articulated here should represent sufficient evidence for the importance of paying attention to the correct interpretation of the situation to be modeled in order to select the appropriate fusion rule. We hope that this article provides a clear description of the situations where it is appropriate to apply Dempster's rule, and where it is not. With this, we hope that controversy around belief fusion of the past 25 years can come to an end, and that the research community can focus on new challenges and opportunities for the advancement of belief theory.

REFERENCES

- Daniel, M. (2000). Associativity in combination of belief functions. In *Proceedings of 5th Workshop on Uncertainty Processing. - Praha, Edicni oddeleni VSE 2000*, pages 41–54. Springer.
- de Finetti, B. (1974a). The true subjective probability problem. In C.-A. Staël von Holstein, editor, *The concept of probability in psychological experiments*, pages 15–23, Dordrecht, Holland. D.Reidel Publishing Company.
- de Finetti, B. (1974b). The value of studying subjective evaluations of probability. In C.-A. Staël von Holstein, editor, *The concept of probability in psychological experiments*, pages 1–14, Dordrecht, Holland. D.Reidel Publishing Company.
- Destercke, S. and Dubois, D. (2010). Idempotent merging of belief functions: Extending the minimum rule of possibility theory. In T. Denœx, editor, *Workshop on the Theory on Belief Functions (WTBF 2010)*.
- Dubois, D. and Prade, H. (1986). On the unicity of Dempster's rule of combination. *International Journal of Intelligent Systems*, **1**, 133–142.
- Dubois, D. and Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.*, **4**, 244–264.
- Haenni, R. (2003). Ignoring Ignorance is Ignorant. Working paper of the philosophy, probability, and modeling research group, University of Konstanz.
- Haenni, R. (2005). Shedding New Light on Zadeh's Criticism of Dempster's Rule of Combination. In *Proceedings of Information Fusion*.
- Hájek, P. and Valdés, J. J. (1991). Generalized algebraic foundations of uncertainty processing in rule-based expert systems (dempsteroids). *Computers and Artificial Intelligence*, **10**, 29–42.
- Hájek, P., Havránek, T., and Jiroušek, R. (1992). *Uncertain Information Processing in Expert Systems*. CRC Press, Boca Raton, Florida.
- Halpern, J. Y. and Fagin, R. (1992). Two views of belief: Belief as generalised probability and belief as evidence. *Artificial Intelligence*, **54**, 275–317.
- Jøsang, A. (1997). Artificial reasoning with subjective logic. In A. Nayak and M. Pagnucco, editors, *Proceedings of the 2nd Australian Workshop on Commonsense Reasoning*, Perth. Australian Computer Society.
- Jøsang, A. (2002). The Consensus Operator for Combining Beliefs. *Artificial Intelligence*, **142**(1–2), 157–170.
- Jøsang, A. (2007). Probabilistic Logic Under Uncertainty. In *The Proceedings of Computing: The Australian Theory Symposium (CATS2007), CRPIT Volume 65*, Ballarat, Australia.
- Jøsang, A. (2011). Multi-Agent Preference Combination using Subjective Logic. In *International Workshop on Preferences and Soft Constraints (Soft'11)*, Perugia, Italy.
- Jøsang, A. and Elouedi, Z. (2007). Interpreting Belief Functions as Dirichlet Distributions. In *The Proceedings of the 9th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007)*, Hammamet, Tunisia.
- Jøsang, A. and Hankin, R. (2012). Interpretation and Fusion of Hyper Opinions in Subjective Logic. In *Proceedings of the 15th International Conference on Information Fusion (FUSION 2012)*.
- Jøsang, A. and McAnally, D. (2004). Multiplication and Comultiplication of Beliefs. *International Journal of Approximate Reasoning*, **38**(1), 19–51.
- Jøsang, A., Diaz, J., and Rifqi, M. (2010). Cumulative and Averaging Fusion of Beliefs. *Information Fusion*, **11**(2), 192–200. doi:10.1016/j.inffus.2009.05.005.
- Klawonn, F. and Schwecke, E. (1992). On the axiomatic justification of Dempster's rule of combination. *International Journal of Intelligent Systems*, **7**, 469–478.
- Kohlas, J. and Monney, P. (1995). *A Mathematical Theory of Hints. An Approach to Dempster-Shafer Theory of Evidence*, volume 425 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag.
- Lefevre, E., Colot, O., and Vannoorenberghe, P. (2002). Belief functions combination and conflict management. *Information Fusion*, **3**(2), 149–162.
- Murphy, C. K. (2000). Combining belief functions when evidence conflicts. *Decision Support Systems*, **29**, 1–9.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.
- Smets, P. (1990). The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(5), 447–458.
- Voorbraak, F. (1991). On the justification of Dempster's rule of combinations. *Artificial Intelligence*, **48**, 171–197.
- Yager, R. (1987). On the Dempster-Shafer framework and new combination rules. *Information Sciences*, **41**, 93–137.
- Zadeh, L. (1984). Review of Shafer's A Mathematical Theory of Evidence. *AI Magazine*, **5**, 81–83.