6th International Symposium on Imprecise Probability: Theories and Applications, Durham, United Kingdom, 2009

# On the Behavior of the Robust Bayesian Combination Operator and the Significance of Discounting

**Alexander Karlsson**
Informatics Research Centre
University of Skövde, Sweden
alexander.karlsson@his.se

**Ronnie Johansson**
Informatics Research Centre
University of Skövde, Sweden
ronnie.johansson@his.se

**Sten F. Andler**
Informatics Research Centre
University of Skövde, Sweden
sten.f.andler@his.se

## Abstract

We study the *combination problem* for credal sets via the *robust Bayesian combination operator*. We extend Walley's notion of *degree of imprecision* and introduce a measure for *degree of conflict* between two credal sets. Several examples are presented in order to explore the behavior of the robust Bayesian combination operator in terms of imprecision and conflict. We further propose a *discounting operator* that suppresses a source given an interval of *reliability weights*, and highlight the importance of using such weights whenever additional information about the reliability of a source is available.

**Keywords.** Imprecise probabilities, robust Bayesian combination, credal set, discounting, information fusion

## 1 Introduction

We define the *combination problem* as the problem of combining *evidences* regarding some reality of interest (cf., [9]). The problem has gained much attention in several different research fields, in particular *information fusion* (see, e.g., [2]) and *artificial intelligence* (see, e.g., [18]). We have here taken a "set-point-wise Bayesian", or *credal* [11, 5], approach to the combination problem via *the robust Bayesian combination operator*. One important advantage with such an approach is that it is easily adoptable for practitioners and researchers that already are familiar with (standard) Bayesian theory. It should be emphasized that the combination problem is different from the *aggregation problem* where the main goal is to find a common agreement among sources. If an *aggregation operator* [19, Section 1.1] is applied to identical operands, typically the result will also be the same, since it represents a "perfect agreement". If we consider the same scenario, using a *combination operator* instead, the result usually represents stronger evidence in comparison to any of the operands, since both sources agree on some hypotheses, i.e., the result is different from the operands. Several researchers have addressed the aggregation problem (see, e.g., [12, 13, 20]), however, the combination problem is an overlooked area in the case of general credal sets. Combination of evidences in the form of so-called *mass functions* (which can be transformed into a particular type of credal set [2]), have been thoroughly studied within *evidence theory* [16], mainly via some variant of *Dempster's rule*. However, it has been shown that Dempster's rule can yield disparate results in comparison to the robust Bayesian combination operator, in fact, the results can even be disjoint [2].

Our main concern in this paper is to characterize the behavior, interpretation, and implications of utilizing the robust Bayesian combination operator for the combination problem. Furthermore, we introduce a discounting operator which can be used whenever an interval of reliability weights are known for the sources involved in the combination.

The paper is organized as follows: in Section 2, we elaborate on *credal set theory*[1] and derive the robust Bayesian combination operator. In Section 3, we elaborate on *imprecision* and *conflict* with respect to credal sets. In Section 4, we present three examples and utilize imprecision and conflict in order to investigate the results. In Section 5, we introduce the discounting operator and revisit two of the mentioned examples. Lastly, in Section 6, we present a summary, our conclusions, and ideas for future work.

## 2 Preliminaries

We here present some background on credal set theory and derive the robust Bayesian combination operator via its precise counterpart, the Bayesian combination operator.

[1] Also known as *theory of credal sets*. We choose the term "credal set theory" since it is coherent with "Bayesian theory".

### 2.1 Credal Set Theory

*Credal set theory* [4, 5, 6, 11] is a generalization of Bayesian theory where one acknowledges that there might be more than one reasonable probability distribution for representing belief. As a consequence one is allowed to adopt a *closed convex set* of such distributions, commonly referred to as a *credal set*, as the fundamental representation of belief. In order to update such belief, one applies Bayes' theorem *point-wise* to a credal set of *priors* and a *convex set of likelihood functions*. As a last step one utilizes a convex hull operation. Note that in the special case of singleton sets, the theory reduces to standard Bayesian theory.

Let us denote a credal set by $\mathcal{P}_X$, containing probability distributions of the form $p(X)$, $\mathcal{P}_{X|y}$ for distributions in the conditional form $p(X|y)$, and $\mathcal{P}_{X,Y}$ for joint probability distributions $p(X,Y)$. Let $ext\,(\mathcal{P}_X)$ denote the set of *extreme points* (also known as *vertices*) of $\mathcal{P}_X$, i.e., distributions that cannot be expressed as a *convex combination*[2] of any other distributions in the set. We only consider credal sets that have a finite set of extreme points (also known as *polytopes*). Each credal set $\mathcal{P}_X$ can be described as the set of convex combinations of points in $ext\,(\mathcal{P}_X)$, in other words, it suffices to maintain a credal sets' extreme points in order to represent it. In a number of places throughout this paper we will use the credal set that contains all probability distribution for some random variable. Let us therefore formally define this credal set:

**Definition 1.** *Let $\mathcal{P}^*_X$ denote the set of all probability distributions for a random variable $X$ with state space $\Omega_X$, i.e., $\mathcal{P}^*_X \triangleq \{p : 0 \leq p(x_i) \leq 1,\ 1 \leq i \leq |\Omega_X|,\ \sum_{i=1}^{|\Omega_X|} p(x_i) = 1\}$*

One controversy in credal set theory is how one should define *independence* between variables (for an overview see [3]). We here adopt the most commonly used such definition, referred to as *strong independence* [6]:

**Definition 2.** *$X$ and $Y$ are strongly independent iff each $p_i \in ext\,(\mathcal{P}_{X,Y})$ can be expressed as $p_i = p_j p_k$, where $p_j \in \mathcal{P}_X$ and $p_k \in \mathcal{P}_Y$. $X$ and $Y$ are strongly conditionally independent given $Z$ iff $p_i \in ext\left(\mathcal{P}_{X,Y|z}\right)$ can be expressed as $p_i = p_j p_k$, $\forall z \in \Omega_Z$, where $p_j \in \mathcal{P}_{X|z}$ and $p_k \in \mathcal{P}_{Y|z}$.*

[2] A convex combination of points $\{p_i : 1 \leq i \leq n\}$ is defined as $\sum_{i=1}^{n} \lambda_i p_i$, where $\sum_{i=1}^{n} \lambda_i = 1$, $\lambda_i \geq 0$

### 2.2 The Robust Bayesian Combination Operator

Let us first derive, via Bayes' theorem, the Bayesian combination operator, which we then generalize to operate on credal sets. The derivation is inspired by Arnborg [1, 2]. The derivation has previously been utilized in order to define *distinctness of evidences* in variants of evidence theory [17, Sect. 3.1]. Assume that two *sources* have made observations $y_1$ and $y_2$, respectively, related to a random variable $X$. If one wants to formulate one's belief regarding $X$, based on the observations made by the sources, one utilizes Bayes' theorem:

$$p(X|y_1, y_2) = \frac{p(y_1, y_2|X)p(X)}{\displaystyle\sum_{x\in\Omega_X} p(y_1, y_2|x)p(x)} \tag{1}$$

We see that the *posterior belief* $p(X|y_1, y_2)$ is affected by the observations through the *joint likelihood* $p(y_1, y_2|X)$. Hence, it is reasonable to interpret such likelihood as being *evidence* regarding $X$ [9]. Now, if one's posterior belief $p(X|y_1, y_2)$ should be a representation of the available evidence solely, i.e., the posterior belief should be equal to the normalized joint likelihood function, then we need to set our prior belief $p(X)$ to the *uniform distribution* over $\Omega_X$. If we also can assume that the sources have made conditionally independent observations given $X$, i.e.,:

$$p(y_1, y_2|X) = p(y_1|X)p(y_2|X) \tag{2}$$

and that both sources have adopted the uniform distribution as their prior belief $p(X)$, i.e., their belief is completely determined by likelihoods, then we get:

$$\begin{aligned} p(X|y_1, y_2) &= \frac{p(y_1|X)p(y_2|X)p(X)}{\displaystyle\sum_{x\in\Omega_X} p(y_1|x)p(y_2|x)p(x)} && (3)\\ &= \frac{\dfrac{p(X|y_1)p(y_1)}{p(X)}\dfrac{p(X|y_2)p(y_2)}{p(X)}}{\displaystyle\sum_{x\in\Omega_X} \dfrac{p(x|y_1)p(y_1)}{p(x)}\dfrac{p(x|y_2)p(y_2)}{p(x)}} && (4)\\ &= \frac{p(X|y_1)p(X|y_2)}{\displaystyle\sum_{x\in\Omega_X} p(x|y_1)p(x|y_2)} && (5) \end{aligned}$$

We know that:

$$\begin{aligned} p(X|y_i) &= \frac{p(y_i|X)p(X)}{\displaystyle\sum_{x\in\Omega_X} p(y_i|x)p(x)} \\ &= \frac{p(y_i|X)}{\displaystyle\sum_{x\in\Omega_X} p(y_i|x)}, \end{aligned} \tag{6}$$

$i \in \{1, 2\}$, since the sources have adopted the uniform distribution as prior belief. Hence, Eq. 5 constitutes an operator that takes two probability functions, interpreted as evidences, i.e., normalized likelihoods, as operands, and returns a new such function, representing the combined evidence, i.e., normalized joint likelihood. We are now ready to define the *Bayesian combination operator* [1, 2]:

**Definition 3.** *The Bayesian Combination (BC) Operator*[3] *is defined as:*

$$p_1(X) \otimes_{\mathcal{B}} p_2(X) \triangleq \frac{p_1(X)p_2(X)}{\sum_{x \in \Omega_X} p_1(x)p_2(x)},$$

*where $p_1(X)$ and $p_2(X)$ are interpreted as conditionally independent evidences, i.e., normalized likelihoods that are conditionally independent given $X$ (see Eq. 2). The operator is undefined when $\sum_{x \in \Omega_X} p_1(x)p_2(x) = 0$.*

Let us first comment on the case when $\sum_{x \in \Omega_X} p_1(x)p_2(x) = 0$. The case implies that likelihoods are such that at least one of them is zero for every $x \in \Omega_X$, which is exceptional in any properly modeled system. The exact way of dealing with such an exceptional case is application dependent. One technique for resolving the case is to utilize discounting with reliability weights strictly smaller than one (see further Sect. 5).

Note that if the operands strongly agree on some $x \in \Omega_X$ as being the most probable, then the operator will reinforce such probability in the resulting posterior function. As mentioned in the introduction, such behavior is clearly different from what one would expect from an aggregation operator. The reason for why such behavior is reasonable is due to the assumption of conditionally independence between evidences given $X$, as described by Eq. 2. Let us demonstrate this behavior of the BC operator with a simple example:

**Example 1.** *Assume that two sources reports the following probability distributions as a representation of conditionally independent evidences regarding the random variable $X$ with state space $\Omega_X$:*

$$p_1(x_1) = 0.7,\ p_1(x_2) = 0.2,\ p_1(x_3) = 0.1$$
$$p_2(x_1) = 0.8,\ p_2(x_2) = 0.1,\ p_2(x_3) = 0.1,$$

*Applying the BC operator to $p_1$ and $p_2$, i.e., $p_1 \otimes_{\mathcal{B}} p_2$, yields the following distribution:*

$$p_{1,2}(x_1) \approx 0.95,\ p_{1,2}(x_2) \approx 0.03,\ p_{1,2}(x_3) \approx 0.02,$$

*Hence, the result constitutes stronger evidence for $x_1$ than any of the operands.*

Now if we want to define an operator that generalizes the BC operator, in the sense of "point-wise Bayesianism", then one can substitute the operand single distributions to credal sets and apply the BC operator point-wise on every pair of distributions within the sets. Indeed, such an operator exists under the name *robust Bayesian combination operator* [1, 2]:

**Definition 4.** *The Robust Bayesian Combination (RBC) Operator* [4]*:*

$$\mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2 \triangleq CH \Big\{ p_i(X) \otimes_{\mathcal{B}} p_j(X) : p_i \in \mathcal{P}_X^1,\ p_j \in \mathcal{P}_X^2 \Big\},$$

*where $CH$ denotes the convex hull, $\mathcal{P}_X^1$ and $\mathcal{P}_X^2$ are interpreted as strongly conditionally independent evidences, i.e., convex sets of normalized likelihoods that are strongly conditionally independent given $X$ (see Def. 2). The operator is undefined if there exists $p_i \in \mathcal{P}_X^1$ and $p_j \in \mathcal{P}_X^2$ such that $\sum_{x \in \Omega_X} p_i(x)p_j(x) = 0$.*

The operator is both associative and commutative. Note that the case regarding division by zero is inherited from the BC operator (Def. 3). Discounting the operands (see further Sect. 5) using reliability weights strictly smaller than one, resolves such case (see further the discussion following Def. 3). Throughout the remainder of the paper we will assume that some technique, guaranteeing $\sum_{x \in \Omega_X} p_i(x)p_j(x) > 0$, for all $p_i \in \mathcal{P}_X^1$ and $p_j \in \mathcal{P}_X^2$, has been utilized (e.g., discounting).

The following theorem facilitates computation with the RBC operator (the theorem was implicitly mentioned in [2], with no proof, and explicitly stated in [1, Theorem 1], where only a "proof hint" was provided):

**Theorem 1.**

$$\mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2 = ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2)$$

*Proof.* The proof is partly inspired by Noack et al. [14, Theorem 2]. First note that $ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2) \subseteq \mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2$ is trivial. Assume that $ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2)$ is strictly smaller than $\mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2$, i.e., $ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2) \subset \mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2$. Then there must exists at least one $u \in ext(\mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2)$ such that $u \notin ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2)$, where $u$ has the following form: $u = p_1 p_2 / \sum_{x \in \Omega_X} p_1(x)p_2(x)$, $p_1 \in \mathcal{P}_X^1$ and

[3] Arnborg [2] referred to this operator as *Laplace's parallel composition*

[4] Arnborg [2] defined the operator without the inclusion of a convex-hull operator (however he mentioned in the discussion following his definition that such an operator should be utilized)

$p_2 \in \mathcal{P}_X^2$, where at least one of $p_1$ and $p_2$ is not an extreme point. We can express $p_1$ and $p_2$ as:

$$\begin{aligned} p_1 &= \sum_{i=1}^{m} \lambda_i v_i \\ p_2 &= \sum_{j=1}^{n} \alpha_j w_j, \end{aligned} \tag{7}$$

where $v_i \in ext(\mathcal{P}_X^1)$, $w_j \in ext(\mathcal{P}_X^2)$, $\lambda_i \geq 0$, $\alpha_j \geq 0$, $1 \leq i \leq m$, $1 \leq j \leq n$, $\sum_{i=1}^{m} \lambda_i = \sum_{j=1}^{n} \alpha_i = 1$. Therefore (remember that the denominator is assumed not to be equal to zero, see the discussion following Def. 3 and Def. 4):

$$u = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \alpha_j v_i w_j}{\sum_{x \in \Omega_X} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \alpha_j v_i(x) w_j(x) \right)} \tag{8}$$

Let us introduce the following notation:

$$\gamma_{i,j} \triangleq \frac{\lambda_i \alpha_j \sum_{x \in \Omega_X} v_i(x) w_j(x)}{\sum_{x \in \Omega_X} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \alpha_j v_i(x) w_j(x) \right)} \tag{9}$$

We can now rephrase $u$ as:

$$u = \sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{i,j} \frac{v_i w_j}{\sum_{x \in \Omega_X} v_i(x) w_j(x)} \tag{10}$$

Since:

$$\frac{v_i w_j}{\sum_{x \in \Omega_X} v_i(x) w_j(x)} \in ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2), \tag{11}$$

and $\gamma_{i,j} \geq 0$, $\sum_{i=1}^{m} \sum_{j=1}^{n} \gamma_{i,j} = 1$, we get $u \in ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2)$, which is a contradiction. Hence we must conclude that $\mathcal{P}_X^1 \otimes_{\mathcal{R}} \mathcal{P}_X^2 = ext(\mathcal{P}_X^1) \otimes_{\mathcal{R}} ext(\mathcal{P}_X^2)$. □

# 3 Imprecision and Conflict

We here define measures for degree of *imprecision* and *conflict*.

## 3.1 Degree of Imprecision

Obviously, since credal set theory belongs to the family of theories referred to as *imprecise probabilities* [23], *imprecision* is an important concept to define. Walley [21, Section 5.1.4] has introduced a measure which he refers to as *the degree of imprecision* for an event $x_i \in \Omega_X$:

$$\Delta(x_i) \triangleq \max_{p \in \mathcal{P}_X} p(x_i) - \min_{p \in \mathcal{P}_X} p(x_i) \tag{12}$$

However, the measure does not capture the imprecision of a credal set, since it only operates on single events. At first, one might be tempted to think of the imprecision of a credal set as its *volume*. However, the volume can be made arbitrarily small while a high degree of imprecision for some event is preserved, something that is counterintuitive. Let us therefore base our measure of degree of imprecision for a credal set on Walley's measure in the following way:

**Definition 5.** *Degree of Imprecision:*

$$\mathcal{I}(\mathcal{P}_X) \triangleq \frac{1}{n} \sum_{x \in \Omega_X} \Delta(x)$$

*where* $\mathcal{P}_X \subseteq \mathbb{R}^n$ *and* $n = |\Omega_X|$

The optimization problems involved in the definition of $\mathcal{I}$ are linear, hence, the solutions can be found by iterating through the extreme points.

## 3.2 Degree of Conflict

Assume that two sources report (strongly conditionally independent) evidence in the form of credal sets $\mathcal{P}_X^1$ and $\mathcal{P}_X^2$ and that one wants to formulate the combined evidence concerning $X$ based on these sets. If both sources report exactly the same credal set, then they are willing to act according to any distribution within any of their sets. In other cases, i.e., when the credal sets are not equal, then there exists a distribution which not both sources are willing to act upon, i.e., a certain *degree of conflict* is present. Intuitively, the degree of conflict between $\mathcal{P}_X^1$ and $\mathcal{P}_X^2$ should be related to some distance between the sets. Indeed, there exists such distance measure, which goes under the name of *Hausdorff distance* [10]. Let us therefore define a *degree of conflict* between two credal sets in the following way:

**Definition 6.** *Degree of Conflict:*

$$\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^2) \triangleq \frac{\mathcal{H}(\mathcal{P}_X^1, \mathcal{P}_X^2)}{\sqrt{2}},$$

*where the denominator is a constant constituting the* diameter *of the set* $\mathcal{P}_X^*$ *(see Def. 1), i.e.,* $\max_{p_i \in \mathcal{P}_X^*} \left\{ \max_{p_j \in \mathcal{P}_X^*} d(p_i, p_j) \right\} = \sqrt{2}$ *(the diameter of a credal set is found in the set of distances between extreme points [7, Theorem 12]) where* $d$ *denotes the Euclidean distance, and* $\mathcal{H}$ *is the Hausdorff distance defined by:*

$$\mathcal{H}(\mathcal{P}_X^1, \mathcal{P}_X^2) \triangleq \max \left\{ \overrightarrow{\mathcal{H}}(\mathcal{P}_X^1, \mathcal{P}_X^2), \overrightarrow{\mathcal{H}}(\mathcal{P}_X^2, \mathcal{P}_X^1) \right\},$$

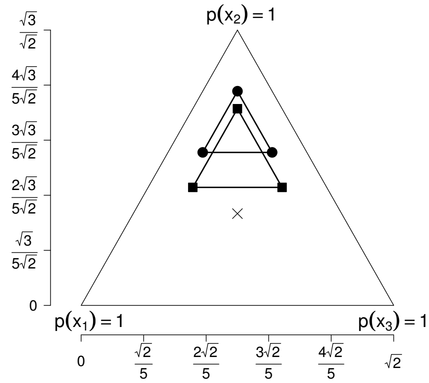



Figure 1: $\mathcal{P}_X^1$ (circles) and $\mathcal{P}_X^2$ (squares) projected on two-dimensional space. The triangle where extreme points $p(x_1) = p(x_2) = p(x_3) = 1$ have been, marked constitutes $\mathcal{P}_X^*$ (see Def. 1).

*where $\overrightarrow{\mathcal{H}}$ is the forward Hausdorff distance defined by:*

$$\overrightarrow{\mathcal{H}}(\mathcal{F}_1, \mathcal{F}_2) \triangleq \max_{f_i \in \mathcal{F}_1} \left\{ \min_{f_j \in \mathcal{F}_2} d(f_i, f_j) \right\},$$

*where $\mathcal{F}_1$ and $\mathcal{F}_2$ are general closed convex sets in $\mathbb{R}^n$.*

The forward Hausdorff-distance can be calculated in $O(|ext(\mathcal{F}_1)||fac(\mathcal{F}_2)|)$ [10], where $fac$ denotes the set of faces. Let us demonstrate the conflict measure by a simple example:

**Example 2.** *Consider Fig. 1 where two credal sets, $\mathcal{P}_X^1$ and $\mathcal{P}_X^2$, for a random variable $X$ with $\Omega_X = \{x_1, x_2, x_3\}$, has been plotted. From the figure, it is seen that $\overrightarrow{\mathcal{H}}(\mathcal{P}_X^2, \mathcal{P}_X^1) > \overrightarrow{\mathcal{H}}(\mathcal{P}_X^1, \mathcal{P}_X^2)$, since there exists at least one point in $\mathcal{P}_X^2$ (e.g., the lower right extreme point) from where the minimum distance to $\mathcal{P}_X^1$ is larger than the distance from any point in $\mathcal{P}_X^1$ to a point in $\mathcal{P}_X^2$. Hence, the Hausdorff distance $\mathcal{H}(\mathcal{P}_X^1, \mathcal{P}_X^2)$ must be equal to the forward Hausdorff distance $\overrightarrow{\mathcal{H}}(\mathcal{P}_X^2, \mathcal{P}_X^1)$, which is the maximum of the set of distances from the set of extreme points of $\mathcal{P}_X^2$ to $\mathcal{P}_X^1$'s faces [10]. In this example, the maximum such distance, approximately equal to $0.16$, is found among the distances between the lower extreme points of $\mathcal{P}_X^2$ to the lower extreme points of $\mathcal{P}_X^1$, i.e., $\mathcal{H}(\mathcal{P}_X^1, \mathcal{P}_X^2) \approx \overrightarrow{\mathcal{H}}(\mathcal{P}_X^2, \mathcal{P}_X^1) \approx 0.16$, yielding a degree of conflict $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^2) \approx 0.11$.*

Notice that if $\mathcal{P}_X^1 = \mathcal{P}_X^2$ then $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^1) = 0$. Also, if $ext(\mathcal{P}_X^1) \subseteq ext(\mathcal{P}_X^*)$ and $ext(\mathcal{P}_X^2) \subseteq ext(\mathcal{P}_X^*)$, and $ext(\mathcal{P}_X^1) \cap ext(\mathcal{P}_X^2) = \emptyset$ then $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^1) = 1$ (since the distance between two different extreme points of the set $\mathcal{P}_X^*$ is $\sqrt{2}$).

# 4 Examples

We will here give some examples of utilizing the robust Bayesian combination (RBC) operator in scenarios where there are different degrees of conflict present. For simplicity, let us utilize the family of credal sets that can be obtained by the *imprecise Dirichlet model* (IDM) [22] for constructing the operand credal sets. Note that these sets stem from a credal set of priors (hence not from a set of likelihoods) and that we are only utilizing the IDM as a convenient way of constructing different geometrical shapes of credal sets for the examples. Consider a random variable $X$ with state space $\Omega_X = \{x_1, x_2, x_3\}$. A credal set obtained from the IDM for this state space can be parameterized according to:

$$\begin{aligned} &\mathrm{IDM}(\boldsymbol{\alpha}, s) \triangleq \\ &\left\{ p : \frac{\alpha_i}{\sum_{i=1}^{3} \alpha_i + s} \leq p(x_i) \leq \frac{\alpha_i + s}{\sum_{i=1}^{3} \alpha_i + s}, \right. \\ &\left. 1 \leq i \leq 3, \ \sum_{i=1}^{3} p(x_i) = 1 \right\}, \end{aligned} \tag{13}$$

where $\alpha_i$ denotes the $i^{th}$ component of $\boldsymbol{\alpha}$.

## 4.1 Low Conflict

Let us start with an example where there exists a low degree of conflict between the sources. We define the example by utilizing Eq. (13) on the following parameters:

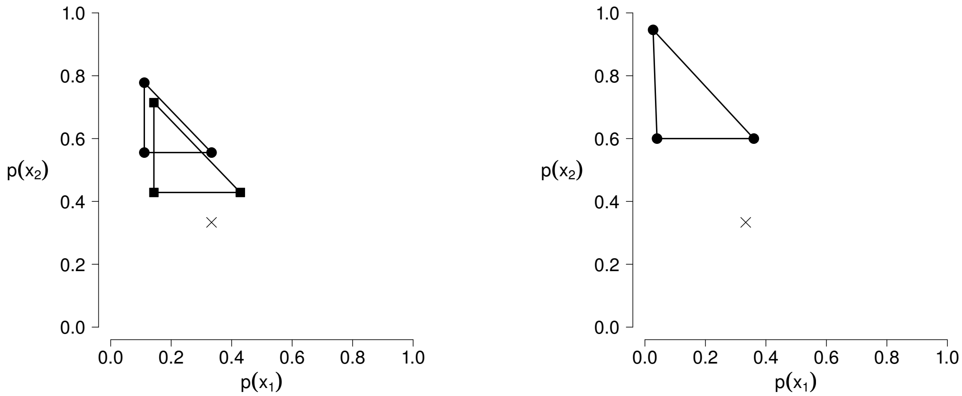
$$\begin{aligned} \mathcal{P}_X^1 &= \mathrm{IDM}((1, 5, 1), 2) \\ \mathcal{P}_X^2 &= \mathrm{IDM}((1, 3, 1), 2) \end{aligned} \tag{14}$$

The corresponding credal sets are shown in Fig. 2(a), where the sets have been projected on the components $p(x_1)$ and $p(x_2)$ (this enables one to see the probabilities directly from the plot). The line segment defined by coordinates $(0, 1)$ and $(1, 0)$ corresponds to the set of distributions where $p(x_3) = 0$. From the figure we see that there is only a slight conflict, $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^2) \approx 0.11$, and that both sources essentially agree on "$x_2$" as being most probable. Therefore the result, denoted by $\mathcal{P}_X^{1,2}$ ($\mathcal{I}(\mathcal{P}_X^{1,2}) \approx 0.34$), is reinforced towards a high probability for "$x_2$", as is seen in Fig. 2(b).

## 4.2 Balanced Conflict

Consider an example where the evidences from the sources are strongly conflicting:

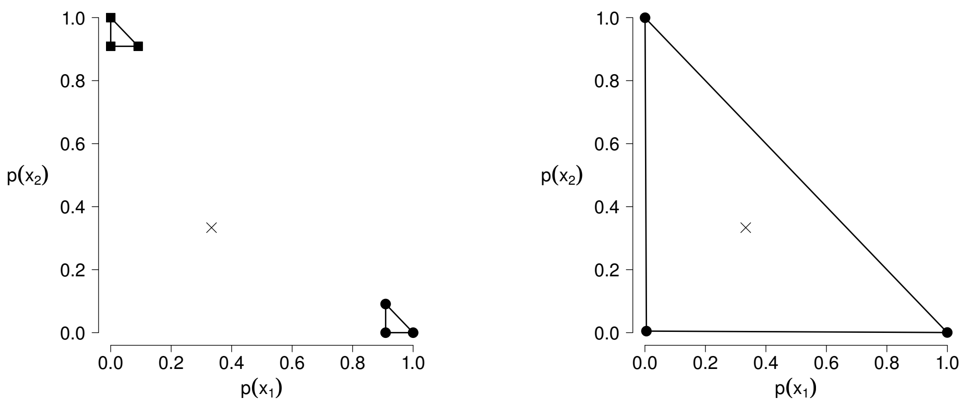
$$\begin{aligned} \mathcal{P}_X^1 &= \mathrm{IDM}((20, 10^{-3}, 10^{-3}), 2) \\ \mathcal{P}_X^2 &= \mathrm{IDM}((10^{-3}, 20, 10^{-3}), 2) \end{aligned} \tag{15}$$




(a) $\mathcal{P}_X^1$ (circles) and $\mathcal{P}_X^2$ (squares)

(b) $\mathcal{P}_X^{1,2}$

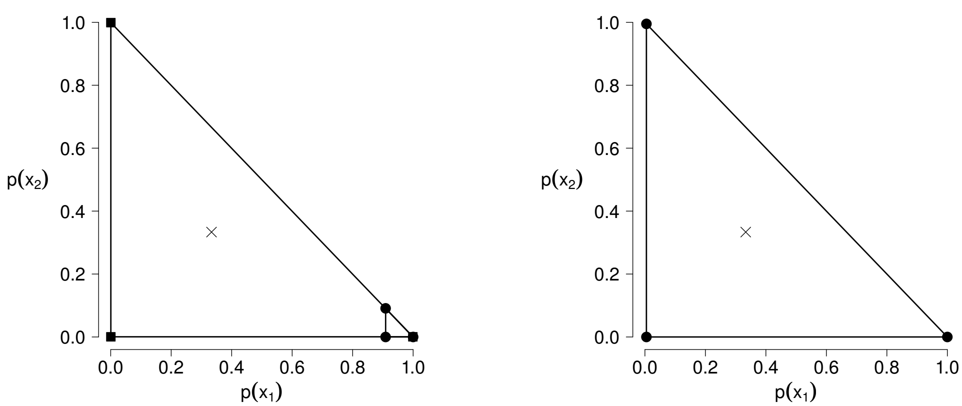
Figure 2: $\mathcal{P}_X^i$, $i \in \{1,2\}$, and $\mathcal{P}_X^{1,2}$ for Example 1 – Low Conflict.




(a) $\mathcal{P}_X^1$ (circles) and $\mathcal{P}_X^2$ (squares)

(b) $\mathcal{P}_X^{1,2}$

Figure 3: $\mathcal{P}_X^i$, $i \in \{1,2\}$, and $\mathcal{P}_X^{1,2}$ for Example 2 – Balanced Conflict.




(a) $\mathcal{P}_X^1$ (circles) and $\mathcal{P}_X^2$ (squares)

(b) $\mathcal{P}_X^{1,2}$

Figure 4: $\mathcal{P}_X^i$, $i \in \{1,2\}$, and $\mathcal{P}_X^{1,2}$ for Example 3 – Unbalanced Conflict.

Since the sources expresses the same degree of imprecision, we refer to the conflict as *balanced*. The operand credal sets and result can be seen in Fig. 3. We see that there is a high degree of conflict, $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^2) \approx 0.91$ and that the resulting credal set $\mathcal{P}_X^{1,2}$ has a high degree of imprecision, $\mathcal{I}(\mathcal{P}_X^{1,2}) \approx 1$. The main reason for this is due to that the "point-wise" combination of the lower left extreme points of $\mathcal{P}_X^1$ and $\mathcal{P}_X^2$ results in the lower left extreme point of $\mathcal{P}_X^{1,2}$; a case that is similar to the well-known *Zadeh's (counter) example* for Dempster's rule [24]. The reason for such behavior is due to that the extreme points component-wise suppress each other for events $x_1$ and $x_2$.

### 4.3 Unbalanced Conflict

Now consider an example where one of the operand credal set is highly imprecise while the other is not:

$$\begin{aligned}\mathcal{P}_X^1 &= \mathrm{IDM}((20, 10^{-3}, 10^{-3}), 2) \\ \mathcal{P}_X^2 &= \mathrm{IDM}((10^{-3}, 10^{-3}, 10^{-3}), 2)\end{aligned} \tag{16}$$

The corresponding credal sets can be seen in Fig. 4. We see that the resulting credal set $\mathcal{P}_X^{1,2}$ has been strongly affected by the second source since $\mathcal{I}(\mathcal{P}_X^{1,2}) \approx 1$. However, since there exist distributions in $\mathcal{P}_X^2$ that are positioned at a large distance from any distribution in $\mathcal{P}_X^1$, there is a strong conflict present: $\mathcal{K}(\mathcal{P}_X^1, \mathcal{P}_X^2) \approx 0.91$. Since the conflict in this case is due to differences in imprecision, we will refer to the conflict as *unbalanced*.

## 5 Discounting

Assume that one possesses information concerning the *reliability* of the sources and that one encodes this information via a convex set of *reliability weights* $\mathcal{W}$ [5], i.e., an interval. If one knows that some source is not fully reliable, e.g., a sensor of low quality, then one should suppress the statement from that source accordingly, i.e., the source should have less influence on the end result. Such procedure is commonly referred to as *discounting* in the literature [16]. If both the credal set and set of reliability weights are singleton, then discounting is achieved by transforming the single distribution, with respect to the weight, to a new distribution that is more similar to the uniform distribution. The reason for this is that the uniform distribution represents evidence that has no influence on the end result when combined with another distribution, i.e., the latter is always returned as result in such case.

[5] Imprecision in reliability weights was inspired by Troffaes [20]

Now, if we generalize the above approach to credal sets and set of reliability weights, preserving the idea of "point-wise Bayesianism", we obtain the following discounting operator:

**Definition 7.** *The RBC Discounting Operator:*

$$\mathcal{D}(\mathcal{P}_X, \mathcal{W}) \triangleq CH\left\{wp + (1-w)p_u : w \in \mathcal{W}, p \in \mathcal{P}_X\right\},$$

*where $\mathcal{P}_X \subseteq \mathbb{R}^n$, $\mathcal{W} \subseteq [0,1]^2$ is an interval of reliability weights, and $p_u \in \mathbb{R}^n$, $n = |\Omega_X|$, is the uniform distribution over $\Omega_X$.*

The RBC discounting operator collapses a credal set "towards" the uniform distribution. Note that when the uniform distribution is combined, using the RBC operator, with any other credal set, the latter is obtained as result. Hence, by applying the RBC discounting operator on an operand, the end result will be less influenced by that operand, depending on $\mathcal{W}$ (the collapse towards the uniform distribution should therefore not be interpreted as a "bias" towards the uniform distribution as it would have been for an aggregation operator). The following theorem allows one to perform computation with the RBC discounting operator:

**Theorem 2.**

$$\mathcal{D}(\mathcal{P}_X, \mathcal{W}) = \mathcal{D}(ext\,(\mathcal{P}_X)\,, ext\,(\mathcal{W}))$$

*Proof.* First note that $\mathcal{D}\,(ext\,(\mathcal{P}_X)\,, ext\,(\mathcal{W})) \subseteq \mathcal{D}\,(\mathcal{P}_X, \mathcal{W})$ is trivial. Assume that $\mathcal{D}\,(ext\,(\mathcal{P}_X)\,, ext\,(\mathcal{W}))$ is strictly smaller than $\mathcal{D}\,(\mathcal{P}_X, \mathcal{W})$, i.e., $\mathcal{D}\,(ext\,(\mathcal{P}_X)\,, ext\,(\mathcal{W})) \subset \mathcal{D}\,(\mathcal{P}_X, \mathcal{W})$. Then there must exists at least one $u \in ext(\mathcal{D}\,(\mathcal{P}_X, \mathcal{W}))$ such that $u \notin \mathcal{D}\,(ext\,(\mathcal{P}_X)\,, ext\,(\mathcal{W}))$ where $u$ has the following form: $u = wp + (1-w)p_u$, $w \in \mathcal{W}$, and $p \in \mathcal{P}_X$, where at least one of $w$ and $p$ is not an extreme point. There are three cases:

*Case 1 – $p \in ext\,(\mathcal{P}_X)$, $w \notin ext\,(\mathcal{W})$:*
We know that $w$ can be expressed as:

$$w = \lambda w_1 + (1-\lambda)w_2, \tag{17}$$

where $w_1 \neq w_2$, $w_1, w_2 \in ext\,(\mathcal{W})$, $\lambda \in (0,1)$. We get:

$$\begin{aligned}u &= wp + (1-w)p_u \\ &= p_u + (\lambda w_1 + (1-\lambda)w_2)(p - p_u) \\ &= p_u + \lambda w_1(p - p_u) + (1-\lambda)w_2(p - p_u) \\ &\quad + \lambda p_u - \lambda p_u \\ &= \lambda(p_u + w_1(p - p_u)) + (1-\lambda)p_u \\ &\quad + (1-\lambda)w_2(p - p_u) \\ &= \lambda(p_u + w_1(p - p_u)) \\ &\quad + (1-\lambda)(p_u + w_2(p - p_u)) \\ &= \lambda(w_1 p + (1-w_1)p_u) \\ &\quad + (1-\lambda)(w_2 p + (1-w_2)p_u)\end{aligned} \tag{18}$$

Hence $u \in \mathcal{D}(ext(\mathcal{P}_X), ext(\mathcal{W}))$, which is a contradiction.

*Case 2 – $p \notin ext(\mathcal{P}_X)$, $w \in ext(\mathcal{W})$:*
We know that $p$ can be expressed as:

$$p = \sum_{i=1}^{n} \alpha_i p_i, \tag{19}$$

where $p_i \in ext(\mathcal{P}_X)$, $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i = 1$. Therefore:

$$\begin{aligned} u &= w\left(\sum_{i=1}^{n} \alpha_i p_i\right) + (1-w)p_u \\ &= \left(\sum_{i=1}^{n} w\alpha_i p_i\right) + (1-w)p_u \\ &\quad + \left(\sum_{i=1}^{n} \alpha_i (1-w)p_u\right) \\ &\quad - \left(\sum_{i=1}^{n} \alpha_i (1-w)p_u\right) \\ &= \left(\sum_{i=1}^{n} \alpha_i (wp_i + (1-w)p_u)\right) \\ &\quad + (1-w)p_u - \left(\sum_{i=1}^{n} \alpha_i (1-w)p_u\right) \\ &= \sum_{i=1}^{n} \alpha_i (wp_i + (1-w)p_u) \end{aligned} \tag{20}$$

Hence $u \in \mathcal{D}(ext(\mathcal{P}_X), ext(\mathcal{W}))$, which is a contradiction.

*Case 3 – $p \notin ext(\mathcal{P}_X))$, $w \notin ext(\mathcal{W})$:*
As is explained in case 1 and 2, we know that:

$$\begin{aligned} w &= \lambda w_1 + (1-\lambda)w_2 \\ p &= \sum_{i=1}^{n} \alpha_i p_i, \end{aligned} \tag{21}$$

We get:

$$\begin{aligned} u = (\lambda w_1 + (1-\lambda)w_2)\left(\sum_{i=1}^{n} \alpha_i p_i\right) \\ + (1 - (\lambda w_1 + (1-\lambda)w_2))p_u \end{aligned} \tag{22}$$

From Case 1 we know that Eq. (22) is equivalent to:

$$\begin{aligned} u = \lambda\left(w_1\left(\sum_{i=1}^{n} \alpha_i p_i\right) + (1-w_1)p_u\right) \\ + (1-\lambda)\left(w_2\left(\sum_{i=1}^{n} \alpha_i p_i\right) + (1-w_2)p_u\right) \end{aligned} \tag{23}$$

From Case 2 we know that Eq. (23) is equivalent to:

$$\begin{aligned} u = \lambda\left(\sum_{i=1}^{n} \alpha_i (w_1 p_i + (1-w_1)p_u)\right) \\ + (1-\lambda)\left(\sum_{i=1}^{n} \alpha_i (w_2 p_i + (1-w_2)p_u)\right) \end{aligned} \tag{24}$$

Hence $u \in \mathcal{D}(ext(\mathcal{P}_X), ext(\mathcal{W}))$, which is a contradiction.

Since all cases lead to contradictions we must conclude that $\mathcal{D}(\mathcal{P}_X, \mathcal{W}) = \mathcal{D}(ext(\mathcal{P}_X), ext(\mathcal{W}))$. $\square$

Let us now revisit the previous presented examples where a strong conflict was present.

### 5.1 Balanced Conflict – Revisited

Assume that the following set of reliability weights regarding the sources is available:

$$\begin{aligned} \mathcal{W}_1 &= [0.80, 0.90] \\ \mathcal{W}_2 &= [0.90, 0.95] \end{aligned} \tag{25}$$

The result of applying the RBC discounting operator on the operands in Sect. 4.2, utilizing the above set of reliability weights, is seen in Fig. 5, where we denote the discounted resulting credal set as $\mathcal{P}_X^{1_d,2_d}$. We get $\mathcal{I}(\mathcal{P}_X^{1_d,2_d}) \approx 0.53$, hence, a significant difference compared to the non-discounted case in Fig. 3(b).

### 5.2 Unbalanced Conflict – Revisited

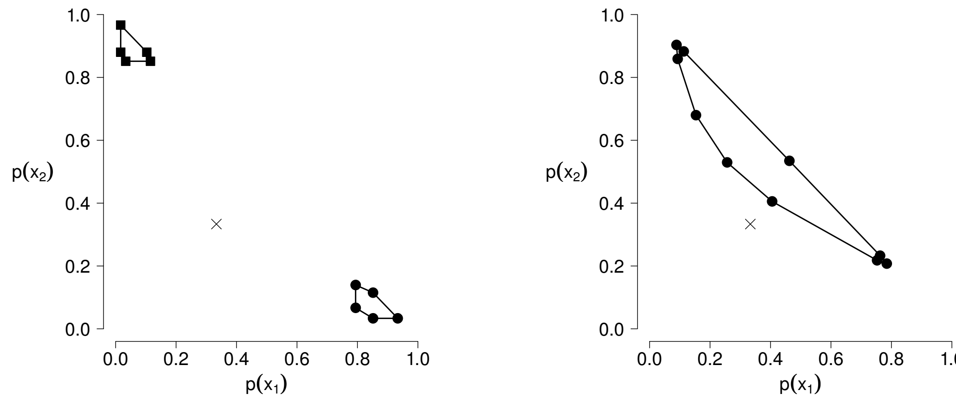
Assume that the following reliability weights regarding the sources are available:

$$\begin{aligned} \mathcal{W}_1 &= [1.00, 1.00] \\ \mathcal{W}_2 &= [0.75, 0.80], \end{aligned} \tag{26}$$

The result of applying the RBC discounting operator on the operands in Sect. 4.3, utilizing the above set of reliability weights, is seen in Fig. 6, where $\mathcal{I}(\mathcal{P}_X^{1_d,2_d}) \approx 0.56$. The lower bound of $\mathcal{W}_2$ will in this case not have any effect since $\mathcal{P}_X^2$ is centered on the uniform distribution.

## 6 Summary and Conclusions

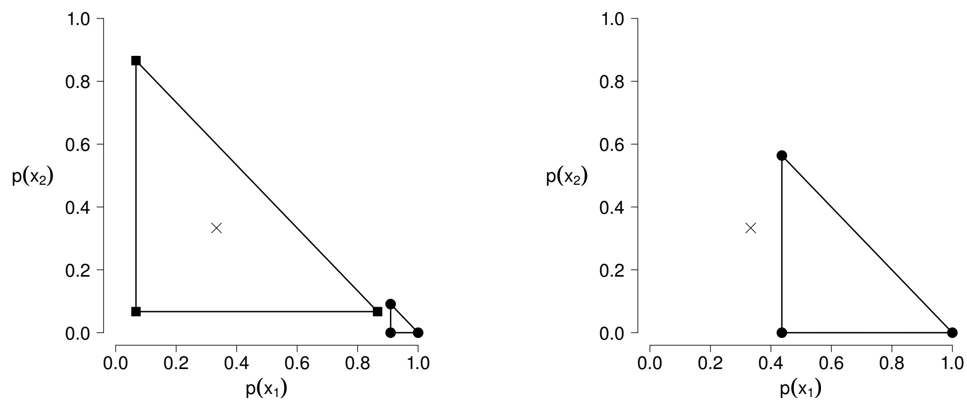
We have studied the combination problem for credal sets via the robust Bayesian combination operator. We extended Walley's notion of degree of imprecision and introduced a measure for degree of conflict between two credal sets. We investigated the behavior of the operator through a number of examples where different degrees of conflict between the operands were




(a) $\mathcal{D}\left(\mathcal{P}_X^1, \mathcal{W}_1\right)$ (circles) and $\mathcal{D}\left(\mathcal{P}_X^2, \mathcal{W}_2\right)$ (squares)

(b) $\mathcal{P}_X^{1_d, 2_d}$

Figure 5: $\mathcal{D}\left(\mathcal{P}_X^i, \mathcal{W}_i\right)$, $i \in \{1, 2\}$, and $\mathcal{P}_X^{1_d, 2_d}$ for Example 2 – Revisited.




(a) $\mathcal{D}\left(\mathcal{P}_X^1, \mathcal{W}_1\right)$ (circles) and $\mathcal{D}\left(\mathcal{P}_X^2, \mathcal{W}_2\right)$ (squares)

(b) $\mathcal{P}_X^{1_d, 2_d}$

Figure 6: $\mathcal{D}\left(\mathcal{P}_X^i, \mathcal{W}_i\right)$, $i \in \{1, 2\}$, and $\mathcal{P}_X^{1_d, 2_d}$ for Example 3 – Revisited.

present. We proposed the RBC discounting operator to be used with the combination operator when a set of reliability weights for the sources are available. We showed that the result of the operators can be computed by utilizing the extreme points of the operand sets. Both operators preserve the intuitive paradigm of "point-wise Bayesianism".

An important aspect to recognize when using the robust Bayesian combination operator is that a source, which reports a credal set that is highly imprecise, can considerably affect the result of the combination (see Fig. 4). If a strong conflict is present among the sources, then additional information about the reliability of the sources can be encoded as reliability weights to be used by the RBC discounting operator. If no such information is available, the conflict may be regarded as irrelevant, if a sufficient number of sources make strong statements that are not in conflict (i.e., the sources have made similar observations). For example, if a large number of credal sets similar to $\mathcal{P}_X^1$ in Fig. 4(a), are combined with $\mathcal{P}_X^2$ in the same figure, then the conflict can be sufficiently suppressed to be regarded as irrelevant

Our next step is to evaluate the robust Bayesian combination and discounting operators against other combination operators, e.g., the Bayesian combination operator and Dempster's rule. Such an evaluation must also concern different modeling strategies for obtaining the reliability weights. We are convinced that if credal set theory is going to be accepted by a broader body of researchers and practitioners, it is necessary to thrust towards research where it can be shown that the theory yields measurable advantages in comparison to other broadly accepted theories, e.g., Bayesian theory.

## Acknowledgements

Computations have been performed with *R* [15] with the package *rcdd* [8]. This work was supported by the Information Fusion Research Program (University of Skövde, Sweden) in partnership with the Swedish Knowledge Foundation under grant 2003/0104 (URL: http://www.infofusion.se).

## References

[1] Stefan Arnborg. Robust Bayesianism: Imprecise and paradoxical reasoning. In *Proceedings of the 7th International Conference on Information fusion*, 2004.

[2] Stefan Arnborg. Robust Bayesianism: Relation to evidence theory. *Journal of Advances in Information Fusion*, 1(1):63–74, 2006.

[3] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.

[4] Fabio Cozman. *Decision Making Based on Convex Sets of Probability Distributions: Quasi-Bayesian Networks and Outdoor Visual Position Estimation.* PhD thesis, The Robotics Institute, Carnegie Mellon University, 1997.

[5] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[6] Fabio Gagliardi Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39:167–184, 2005.

[7] Harold Gordon Eggleston. *Convexity.* CUP Archive, 1958.

[8] Charles J. Geyer, Glen D. Meeden, and incorporates code from cddlib (ver 0.94f) written by Komei Fukuda. *rcdd: rcdd (Computational Geometry)*, 2008. R package version 1.1-1.

[9] Joseph Y. Halpern and Ronald Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[10] Antonio Irpino and Valentino Tontodonato. Cluster reduced interval data using Hausdorff distance. *Computational Statistics*, 21:241–288, 2006.

[11] Isaac Levi. *The enterprise of knowledge.* The MIT press, 1983.

[12] Serafín Moral and José del Sagrado. Aggregation of imprecise probability. Technical report, Department of Computer Science and Artificial Intelligence, 1997.

[13] Robert F. Nau. The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference*, 105:265–282, 2002.

[14] Benjamin Noack, Vesa Klumpp, Dietrich Brunn, and Uwe D. Hanebeck. Nonlinear Bayesian estimation with convex sets of probability densities. In *Proceedings of the 11th International Conference on Information fusion*, 2008.

[15] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[16] Glenn Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[17] Philippe Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8:387–412, 2006.

[18] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[19] Vicenç Torra and Yasuo Narukawa. *Modeling Decisions.* Springer, 2007.

[20] Matthias C.M. Troffaes. Generalizing the conjunction rule for aggregating conflict expert opinions. *International Journal of Intelligent Systems*, 21:361–380, 2006.

[21] Peter Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, 1991.

[22] Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58:3–57, 1996.

[23] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[24] Lofti A. Zadeh. Review of books: A mathematical theory of evidence. *AI Magazine*, 5:81–83, 1984.