

Characterization and Empirical Evaluation of Bayesian and Credal Combination Operators

ALEXANDER KARLSSON
RONNIE JOHANSSON
STEN F. ANDLER

We address the problem of combining independent evidences from multiple sources by utilizing the Bayesian and credal combination operators. We present measures for degree of conflict and imprecision, which we use in order to characterize the behavior of the operators through a number of examples. We introduce discounting operators that can be used whenever information about the reliability of sources is available. The credal discounting operator discounts a credal set with respect to an interval of reliability weights, hence, we allow for expressing reliability of sources imprecisely. We prove that the credal discounting operator can be computed by using the extreme points of its operands. We also perform two experiments containing different levels of risk where we compare the performance of the Bayesian and credal combination operators by using a simple score function that measures the informativeness of a reported decision set. We show that the Bayesian combination operator performed on centroids of operand credal sets outperforms the credal combination operator when no risk is involved in the decision problem. We also show that if a risk component is present in the decision problem, a simple cautious decision policy for the Bayesian combination operator can be constructed that outperforms the corresponding credal decision policy.

Manuscript received August 8, 2010; revised April 27, 2011 and August 17, 2011; released for publication August 22, 2011.

Refereeing of this contribution was handled by Huimin Chen.

Authors addresses: A. Karlsson, Informatics Research Centre, University of Skövde, Sweden, E-mail (alexander.karlsson@his.se); R. Johansson, Informatics Research Centre, University of Skövde, Sweden, E-mail (ronnie.johansson@his.se); S. F. Andler, Informatics Research Centre, University of Skövde, Sweden, E-mail (sten.f.andler@his.se).

1557-6418/11/\$17.00 © 2011 JAIF

1. INTRODUCTION

Bayesian theory [5] is one of the most commonly utilized theories for managing uncertainty in *information fusion* [20, 12]. The theory relies on two main assumptions: (1) a *probability function* should be used for representing *belief* and (2) *Bayes' theorem* should be used for *belief updating* when a new observation has been made. The main criticism of Bayesian theory that can be found in the literature (e.g., [14, 25]) is that the first assumption is unrealistically strong since one is forced to quantify belief precisely even if one only possesses scarce information about the environment of interest. For this reason, a family of alternative theories has been introduced that usually goes under the name *imprecise probability* [26], where belief can be expressed *imprecisely*.

One common theory that belongs to the family of imprecise probability is *credal set theory* [2, 3, 9, 10, 19], also known as “theory of credal sets” [11] and “quasi-Bayesian theory” [8], where one utilizes a *closed convex set of probability functions* (instead of a single function), denoted as a *credal set* [19], for representing belief. An attractive feature of credal set theory is that it reduces to Bayesian theory if singleton sets are adopted. Furthermore, credal set theory can be thought of as point-wise application of Bayes theorem on all probability (and likelihood) functions within operand sets (unlike, e.g., evidence theory [23], which is inconsistent with this point-wise Bayesian paradigm [2, 3]). Hence, credal set theory can be seen as the most straightforward generalization of Bayesian theory to imprecise probability.

In this paper, we are interested in contrasting Bayesian theory with credal set theory when used for *combining independent pieces of evidence*, known as the *combination problem* [16]. Arnborg [2, 3] has previously characterized the relation between *robust Bayesian theory*, which can be seen as a *sensitivity interpretation* [4, 14] of credal set theory, and *evidence theory* [23] when used for the combination problem. We extend Arnborg's work by characterizing the *Bayesian and credal combination operators*¹ in terms of *imprecision* and *conflict* and by introducing methods for accounting for *reliability* of sources. In addition, we also empirically evaluate the use of the operators for decision making regarding some state space of interest. Since the credal combination operator is considerably more computational demanding than the Bayesian counterpart, such an evaluation can reveal whether or not the additional computational expense yields an increase in decision performance.

¹Arnborg [2, 3] denoted this operator by “robust Bayesian combination operator.” We deliberately avoid using this terminology since robust Bayesianism imposes a sensitivity interpretation of the credal set [4, 14] and we do not want to exclude other interpretations (see e.g., Walley [25]).

The paper is organized in the following way: In Section 2, we derive the Bayesian and credal combination operators. In Section 3, we present measures for conflict and imprecision for the operators. Based on these measures, we present a number of examples that highlight the behavior of the operators. We introduce discounting operators for the Bayesian and credal combination operators, which can be used whenever information about reliability of sources is known. In Section 4, we present two experiments; one where no risk component is present in the decision problem, i.e., there is no cost for making an erroneous decision, and one where such a component exists. We discuss the design and analyze the result of each experiment. Lastly, in Section 5, we summarize the article and present the main conclusions.

2. PRELIMINARIES

We derive the Bayesian and credal combination operators and elaborate on how the credal combination operator can be computed.

2.1. Bayesian Combination Operator

Let X and Y_1, \dots, Y_n be discrete random variables with state spaces Ω_X and $\Omega_{Y_1}, \dots, \Omega_{Y_n}$, respectively. Assume that we have n sources and that source $i \in \{1, \dots, n\}$ has made observation $y_i \in \Omega_{Y_i}$ and reported a *likelihood function* $p(y_i | X)$ as a representation of the *evidence* provided by y_i regarding X . By assuming that the observations are *conditionally independent given X* , we can construct the *joint evidence* (or joint likelihood):

$$p(y_1, \dots, y_n | X) = p(y_1 | X) \dots p(y_n | X). \quad (1)$$

In principle, we can use (1) as a Bayesian way of *combining the evidences*, however, this is not convenient when implemented in an operational system since the joint evidence monotonically decreases with the number of sources n . Let us therefore elaborate on how this problem can be solved. Let

$$p_i(X) \triangleq \frac{p(y_i | X)}{\sum_{x \in \Omega_X} p(y_i | x)} \quad (2)$$

i.e., $p_i(X)$ are probability functions (normalized likelihood functions). By using Bayes' theorem and the assumption of conditional independence, we obtain

$$\begin{aligned} p(X | y_1, \dots, y_n) &= \frac{p(y_1, \dots, y_n | X)p(X)}{\sum_{x \in \Omega_X} p(y_1, \dots, y_n | x)p(x)} \\ &= \frac{p(y_1 | X) \dots p(y_n | X)p(X)}{\sum_{x \in \Omega_X} p(y_1 | x) \dots p(y_n | x)p(x)} \\ &= \frac{p_1(X) \dots p_n(X)p(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)p(x)} \\ &= \frac{\frac{p_1(X) \dots p_n(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)} p(X)}{\sum_{x \in \Omega_X} \frac{p_1(x) \dots p_n(x)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)} p(x)}. \end{aligned} \quad (3)$$

Let

$$\Phi(p_1(X), \dots, p_n(X)) \triangleq \frac{p_1(X) \dots p_n(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)}. \quad (4)$$

From (3) we see that the joint evidence $p(y_1, \dots, y_n | X)$ has the same effect on the *posterior* $p(X | y_1, \dots, y_n)$, irrespective of the prior $p(X)$, as $\Phi(p_1(X), \dots, p_n(X))$, i.e., $p(y_1, \dots, y_n | X)$ and $\Phi(p_1(X), \dots, p_n(X))$ are equivalent evidences. The following theorem allows us to *recursively combine* evidences into a joint evidence.

THEOREM 1

$$\begin{aligned} &\Phi(\dots \Phi(p_1(X), p_2(X)) \dots, p_n(X)) \\ &= \frac{p_1(X) \dots p_n(X)}{\sum_{x \in \Omega_X} p_1(x) \dots p_n(x)}. \end{aligned} \quad (5)$$

PROOF See Appendix.

Note that the normalization in each combination in the recursion eliminates the problem of a monotonically decreasing joint evidence when n increases. We use the recursive form of Φ as our basis for the definition of a *Bayesian combination operator* denoted by Φ_B (i.e., we define the operator for two operands) [3, 2]:

DEFINITION 1 The Bayesian combination operator is defined as

$$\Phi_B(p_1(X), p_2(X)) \triangleq \frac{p_1(X)p_2(X)}{\sum_{x \in \Omega_X} p_1(x)p_2(x)} \quad (6)$$

where $p_i(X)$, $i \in \{1, 2\}$, are conditionally independent evidences in the form of probability functions (normalized likelihood functions). The operator is undefined when $\sum_{x \in \Omega_X} p_1(x)p_2(x) = 0$.

Note that the operator is associative and commutative.

2.2. Credal Combination Operator

The credal combination operator, also known as the *robust Bayesian combination operator* (see Footnote 1) [2, 3], can be derived by using *credal set theory* [19, 9, 10, 2, 3]. As we mentioned in the introduction, in credal set theory one represents belief by a *closed convex set of probability functions*. However, one is also allowed to express *evidence* regarding some random variable imprecisely, i.e., instead of a *single likelihood function* as a representation of evidence, as in the Bayesian case, one can adopt a closed convex set of such functions. Combination of such evidences then amounts to applying the Bayesian combination operator point-wise on all possible combinations of functions from the sets. In order to enforce convexity of the posterior result one applies the *convex-hull operator*.

One important concept within credal set theory, which we will use extensively in the proofs is *convex combination* defined as [1].

DEFINITION 2 A convex combination of probability functions $p_1(X), \dots, p_n(X)$ is a probability function expressed in the following form

$$\lambda_1 p_1(X) + \dots + \lambda_n p_n(X) \quad (7)$$

where $(\forall i \in \{1, \dots, n\}) (0 \leq \lambda_i)$ and $\sum_{i=1}^n \lambda_i = 1$.

We can now define the *convex hull* of a finite set as the set of all convex combinations of points in the set [1].

DEFINITION 3 The convex hull of a finite set $\{p_1(X), \dots, p_n(X)\}$ is defined as

$$\begin{aligned} \mathcal{CH}(\{p_1(X), \dots, p_n(X)\}) \\ \triangleq \left\{ \sum_{i=1}^n \lambda_i p_i(X) : (\forall i \in \{1, \dots, n\}) \right. \\ \left. \times (\lambda_i \geq 0), \sum_{i=1}^n \lambda_i = 1 \right\}. \end{aligned} \quad (8)$$

Let $\mathcal{P}(X)$ denote a *prior credal set*, i.e., a closed convex set of probability functions of the form $p(X)$ and $\mathcal{P}(X | y)$ a *posterior credal set* of functions $p(X | y)$. Let $\mathcal{E}(\mathcal{P}(X))$ denote the set of *extreme points* of $\mathcal{P}(X)$, i.e., points that belong to the set and cannot be expressed as a convex combination of other points in the set. We are now ready to define the notion of independence for credal sets referred to as *strong independence* [7].

DEFINITION 4 The discrete random variables X and Y are strongly independent iff all $p(X, Y) \in \mathcal{E}(\mathcal{P}(X, Y))$ can be expressed as $p(X, Y) = p(X)p(Y)$, where $p(X) \in \mathcal{P}(X)$ and $p(Y) \in \mathcal{P}(Y)$. Similarly, X and Y are strongly conditionally independent given Z iff all $p(X, Y | z) \in \mathcal{E}(\mathcal{P}(X, Y | z))$ can be expressed as $p(X, Y | z) = p(X | z) \cdot p(Y | z)$, $\forall z \in \Omega_Z$, where $p(X | z) \in \mathcal{P}(X | z)$ and $p(Y | z) \in \mathcal{P}(Y | z)$.

The intuition behind this definition is that each extreme point of a joint credal set should fulfill the same criteria for independence as in ordinary probability calculus, i.e., the extreme points should factorize [11]. By using this notion of independence, the credal combination operator² can be derived as a straightforward generalization of the Bayesian combination operator.

DEFINITION 5 The credal combination operator is defined as

$$\begin{aligned} \Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \\ \triangleq \mathcal{CH}(\{\Phi_B(p_1(X), p_2(X)) : p_1(X) \in \mathcal{P}_1(X), \\ p_2(X) \in \mathcal{P}_2(X)\}) \end{aligned} \quad (9)$$

²Arnborg [2, 3] defined the operator without the inclusion of a convex-hull operator, however, he mentions in the discussion following his definition that such operator should be utilized. See also Footnote 1.

where $\mathcal{P}_i(X)$, $i \in \{1, 2\}$, are strongly conditionally independent evidences in the form of credal sets (closed convex sets of normalized likelihood functions) and where \mathcal{CH} is the convex-hull operator. The Φ_C operator is undefined iff there exists $p_i(X) \in \mathcal{P}_i(X)$, $i \in \{1, 2\}$, such that Φ_B is undefined.

The operator is associative and commutative. Note that the operator is based on point-wise application of the Bayesian combination operator on all combinations of functions from the operand credal sets. Hence, the operator is equivalent to the Bayesian combination operator for singleton sets. One important credal set, that we will use extensively throughout the article, is the set of all probability functions for a given state space, denoted as a *probability simplex*.

DEFINITION 6 The probability simplex $\mathcal{P}^*(X)$ for a discrete random variable X with state space Ω_X is defined as

$$\mathcal{P}^*(X) \triangleq \left\{ p(X) : (\forall x \in \Omega_X) (p(x) \geq 0), \sum_{x \in \Omega_X} p(x) = 1 \right\} \quad (10)$$

In order to compute the credal combination operator, we only consider operand credal sets that has a finite number of extreme points. Such a property can be guaranteed by using credal sets in the form of *polytopes* [1]

DEFINITION 7 A credal set $\mathcal{P}(X)$ is a polytope iff

$$\mathcal{P}(X) = \mathcal{CH}(\{p_1(X), \dots, p_n(X)\}) \quad (11)$$

where $\{p_1(X), \dots, p_n(X)\} \subset \mathcal{P}^*(X)$ is a finite set and where \mathcal{CH} is the convex-hull operator.

The following theorem enables computation by extreme points of the credal combination operator when the operands are polytopes (the theorem was implicitly mentioned by Arnborg [3], with no proof, and explicitly stated by Arnborg [2], but only a ‘‘proof hint’’ was provided. A corresponding theorem has been stated and proved for filtering (continuous case) by Noack, et al. [21, Theorem 2]).

THEOREM 2

$$\begin{aligned} \Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \\ = \Phi_C(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))). \end{aligned} \quad (12)$$

PROOF See Appendix.

3. CHARACTERIZATION OF THE BAYESIAN AND CREDAL COMBINATION OPERATORS

In this section,³ we define measures for *degree of conflict* and *imprecision* and use these for characterizing the behavior the Bayesian and credal combination operators through a number of examples. We introduce

³This section includes material from Karlsson et al. [16].

discounting operators that can be used whenever information about the *reliability* of sources is known. We exemplify the utilization of the discounting operators by revisiting the examples.

3.1. Degree of Conflict

One important concept when combining evidences from multiple sources is *the degree of conflict* measured on the evidences reported by the sources. Intuitively, such measure can be thought of as an “inverse similarity measure,” i.e., the more similar the reported evidences are, the less conflict exists between the sources. Hence, for the Bayesian case, we simply use the *Euclidean norm* as the basis for a conflict measure.

DEFINITION 8 The degree of conflict between two evidences in the form of probability functions $p_1(X)$ and $p_2(X)$ is defined as

$$\Gamma_B(p_1(X), p_2(X)) \triangleq \frac{\|p_1(X) - p_2(X)\|}{\sqrt{2}} \quad (13)$$

where $\|\cdot\|$ is the Euclidean norm and where the denominator constitutes the diameter of the probability simplex $\mathcal{P}^*(X)$, i.e.,

$$\max \left\{ \max_{p_j(X) \in \mathcal{P}^*(X)} \|p_i(X) - p_j(X)\| : p_i(X) \in \mathcal{P}^*(X) \right\} = \sqrt{2}. \quad (14)$$

Similarly to the above Bayesian conflict measure, we base a credal conflict measure on the notion of similarity. A similarity measure for general closed convex sets exists under the name of *Hausdorff distance* [15]. The Hausdorff distance is the largest distance one can find between a point from any of the two sets to the closest point in the other set. By using the Hausdorff distance we can define the following conflict measure for credal sets:

DEFINITION 9 The degree of conflict between two credal sets $\mathcal{P}_1(X)$ and $\mathcal{P}_2(X)$ is defined as

$$\Gamma_C(\mathcal{P}_1(X), \mathcal{P}_2(X)) \triangleq \frac{\mathcal{H}(\mathcal{P}_1(X), \mathcal{P}_2(X))}{\sqrt{2}} \quad (15)$$

where the denominator constitutes the diameter of the probability simplex $\mathcal{P}^*(X)$ and where \mathcal{H} is the Hausdorff distance defined by [15]

$$\begin{aligned} \mathcal{H}(\mathcal{P}_1(X), \mathcal{P}_2(X)) \\ \triangleq \max \{ \vec{\mathcal{H}}(\mathcal{P}_1(X), \mathcal{P}_2(X)), \vec{\mathcal{H}}(\mathcal{P}_2(X), \mathcal{P}_1(X)) \} \end{aligned} \quad (16)$$

where $\vec{\mathcal{H}}$ is the forward Hausdorff distance:

$$\begin{aligned} \vec{\mathcal{H}}(\mathcal{P}_1(X), \mathcal{P}_2(X)) \\ \triangleq \max \left\{ \min_{p_2(X) \in \mathcal{P}_2(X)} \|p_1(X) - p_2(X)\| : p_1(X) \in \mathcal{P}_1(X) \right\} \end{aligned} \quad (17)$$

where $\|\cdot\|$ is the Euclidean norm.

Note that the credal conflict measure reduces to the Bayesian conflict measure for singleton sets. The forward Hausdorff-distance can be calculated in $O(|\mathcal{E}(\mathcal{P}_1(X))| |\mathcal{F}(\mathcal{P}_2(X))|)$ [15], where $\mathcal{F}(\mathcal{P}(X))$ is the set of faces of $\mathcal{P}(X)$.

3.2. Degree of Imprecision

Obviously, since credal set theory belongs to the family of theories referred to as *imprecise probabilities* [26], *imprecision* is an important concept to define. Walley [25, Section 5.1.4] has introduced a measure which he refers to as *the degree of imprecision* for an event $x \in \Omega_X$

$$\Delta(x, \mathcal{P}(X)) \triangleq \max_{p(X) \in \mathcal{P}(X)} p(x) - \min_{p(X) \in \mathcal{P}(X)} p(x). \quad (18)$$

However, the measure does not capture the imprecision of a credal set since it only operates on single events. Let us therefore base our measure of *degree of imprecision* for a credal set on a simple average of Walley’s measure in the following way.

DEFINITION 10 The degree of imprecision of a credal set $\mathcal{P}(X)$ is defined as

$$\mathcal{I}(\mathcal{P}(X)) \triangleq \frac{1}{|\Omega_X|} \sum_{x \in \Omega_X} \Delta(x, \mathcal{P}(X)) \quad (19)$$

where $\Delta(x, \mathcal{P}(X))$ is Walley’s measure for degree of imprecision for a single event [25, Section 5.1.4].

As an example, if we have a credal set $\mathcal{P}(X)$ where

$$\mathcal{P}(X) = \mathcal{CH}(\{(0.2, 0.1, 0.7)^T, (0.5, 0.2, 0.3)^T, (0.6, 0.3, 0.1)^T\}) \quad (20)$$

where $\Omega_X = \{x_1, x_2, x_3\}$ and the order of the probabilities is $(p(x_1), p(x_2), p(x_3))^T$, then

$$\begin{aligned} \mathcal{I}(\mathcal{P}(X)) &= \frac{1}{3} \sum_{i=1}^3 \Delta(x_i, \mathcal{P}(X)) \\ &= \frac{1}{3} (0.6 - 0.2 + 0.3 - 0.1 + 0.7 - 0.1) \\ &= 0.4. \end{aligned} \quad (21)$$

3.3. Examples

We here provide a number of examples, containing different degrees of imprecision and conflict, in order to characterize the behavior of the Bayesian and credal combination operators. Let us first elaborate on a convenient way of visualizing belief and evidence in cases where the state space consists of three elements. Assume that $\Omega_X = \{x_1, x_2, x_3\}$. In such case the probability simplex $\mathcal{P}^*(X)$ constitutes the plane orthogonally projected on two dimensional space, seen in Fig. 1, which is geometrically equivalent to the convex hull of the points $(1, 0, 0)^T$, $(0, 1, 0)^T$, and $(0, 0, 1)^T$. Each corner of the triangle represents an extreme point of $\mathcal{P}^*(X)$, i.e., a probability function where all probability mass lies on

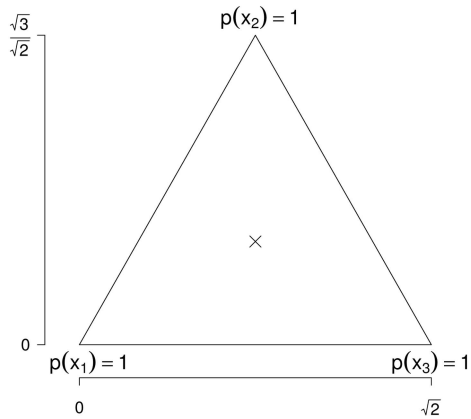


Fig. 1. Probability simplex $\mathcal{P}^*(X)$, where $\Omega_X = \{x_1, x_2, x_3\}$, projected on two dimensional space.

a single element of Ω_X . Each point in the triangle represents a probability function. As an example, the center of the triangle, indicated with a cross, is the uniform distribution over Ω_X . The closer a specific point is to one of the corners in the triangle the higher probability for the respective state. This type of visualization is commonly used within the imprecise probability community (see, e.g., [25]).

3.3.1. Bayesian Combination Operator

Let us start with the example seen in Fig. 2 where there only is a minor conflict among the sources. We see that since both sources suggest x_2 as most probable, seen in Fig. 2(a), the joint evidence, seen in Fig. 2(b), is reinforced towards this state. Now, consider the case where there is a strong conflict among the sources instead, seen in Fig. 3. Both sources have provided evidences that states that x_3 is unlikely to be the true state of X . However, there is a strong disagreement, i.e., conflict among the sources, regarding the states x_1 and x_2 . The joint evidence, seen in Fig. 3(b), is approximately uniformly distributed over $\{x_1, x_2\}$, i.e., from the result we cannot single out a best choice between these two states, however, x_3 is still highly unlikely due to the distance to that corner.

3.3.2. Credal Combination Operator

Let us again start with an example where there is a low degree of conflict between the sources, seen in Fig. 4. The operand credal sets in Fig. 4(a) have been constructed by expanding equilateral triangles around the operands in Fig. 2(a). From the figure we see that both sources essentially agree on the state x_2 as being most probable. Therefore the combined evidence

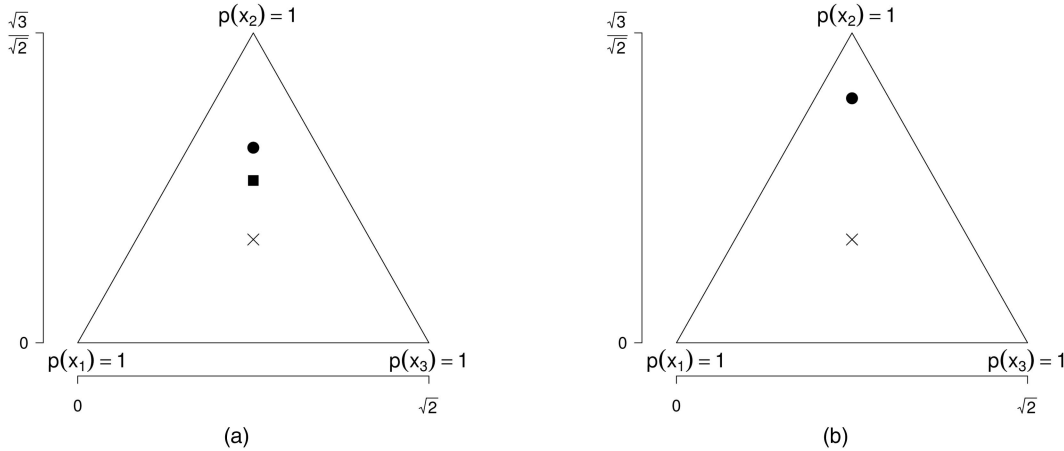


Fig. 2. $p_1(X)$, $p_2(X)$, and $p_{1,2}(X)$ when a low degree of conflict is present. (a) $p_1(X)$ (circle) and $p_2(X)$ (square). (b) $p_{1,2}(X)$.

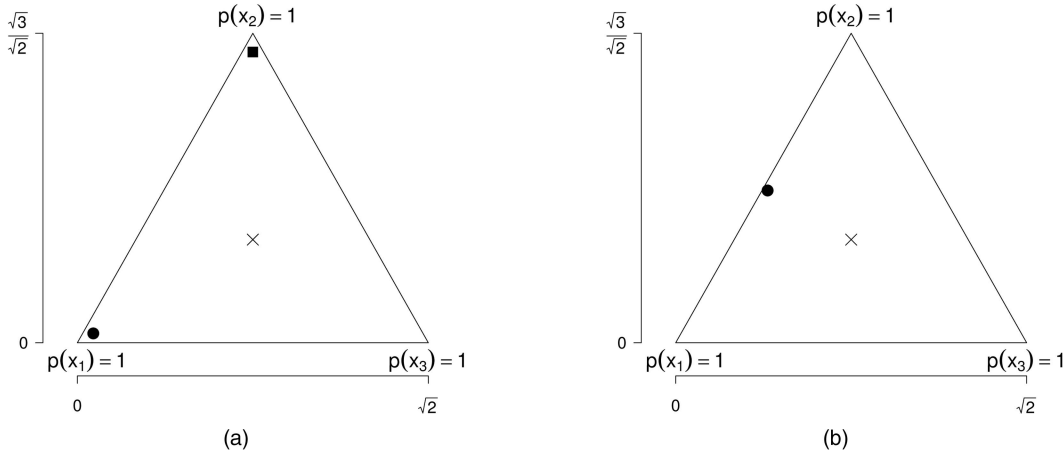


Fig. 3. $p_1(X)$, $p_2(X)$, and $p_{1,2}(X)$ when a high degree of conflict is present. (a) $p_1(X)$ (circle) and $p_2(X)$ (square). (b) $p_{1,2}(X)$.

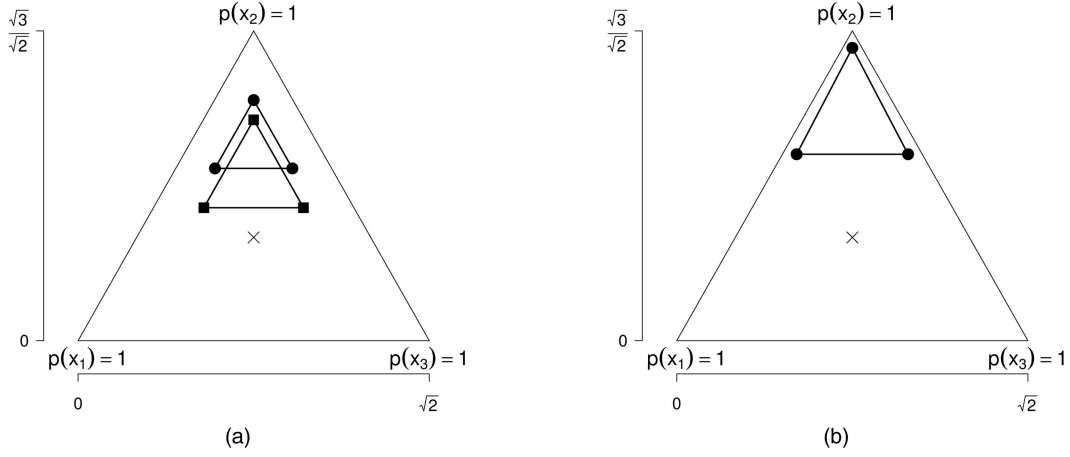


Fig. 4. $\mathcal{P}_1(X)$, $\mathcal{P}_2(X)$, and $\mathcal{P}_{1:2}(X)$ when a low degree of conflict is present. (a) $\mathcal{P}_1(X)$ (circles) and $\mathcal{P}_2(X)$ (squares). (b) $\mathcal{P}_{1:2}(X)$.

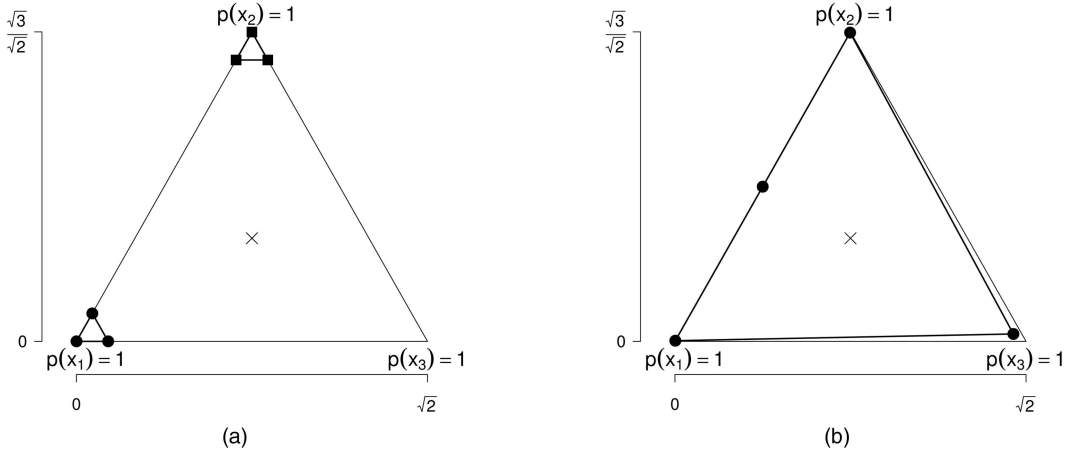


Fig. 5. $\mathcal{P}_1(X)$, $\mathcal{P}_2(X)$, and $\mathcal{P}_{1:2}(X)$ when a high degree of conflict is present. (a) $\mathcal{P}_1(X)$ (circles) and $\mathcal{P}_2(X)$ (squares). (b) $\mathcal{P}_{1:2}(X)$.

$\mathcal{P}_{1:2}(X)$ is reinforced towards a high probability for the state x_2 , as is seen in Fig. 4(b). Note that $\mathcal{P}_{1:2}(X)$ preserves the property of not favoring any of the states x_1 and x_3 .

Consider again an example where evidences are strongly conflicting (a similar example has been presented by Arnborg [2]). The evidences provided by the sources can be seen in Fig. 5. We see that the resulting joint evidence has a high degree of imprecision. Note that it is the combination of the lower right extreme points of the operand credal sets that is the cause for the lower right extreme point of the joint evidence; a case that has similarities with the well-known *Zadeh's example for Dempster's combination rule* [27]. This is due to that the extreme points componentwise suppress each other for the states x_1 and x_2 , i.e., if we denote the lower right extreme point of $\mathcal{P}_1(X)$, $\mathcal{P}_2(X)$, and $\mathcal{P}_{1:2}(X)$ by $p_1(X)$, $p_2(X)$, and $p_{1:2}(X)$, respectively, where

$$p_1(x_1) = 1 - \epsilon - \vartheta, \quad p_1(x_2) = \epsilon, \quad p_1(x_3) = \vartheta \quad (22)$$

$$p_2(x_1) = \epsilon, \quad p_2(x_2) = 1 - \epsilon - \vartheta, \quad p_2(x_3) = \vartheta \quad (23)$$

then we obtain the following expression for $p_{1:2}(x_3)$

$$p_{1:2}(x_3) = \frac{\vartheta^2}{(1 - \epsilon - \vartheta)\epsilon + \epsilon(1 - \epsilon - \vartheta) + \vartheta^2} \quad (24)$$

which approaches one when $\epsilon \rightarrow 0$ (in the figure, $\epsilon > 0$, which is why the lower right extreme point of $\mathcal{P}_{1:2}(X)$ is not exactly positioned at the lower right corner of the probability simplex $\mathcal{P}^*(X)$).

Lastly, let us consider another type of conflict that can appear in the credal case, seen in Fig. 6, where one of the sources expresses a credal set that is highly imprecise, i.e., approximately equivalent to the probability simplex $\mathcal{P}^*(X)$ (there is a small distance between the extreme points of the credal set and the extreme points of the probability simplex which cannot be seen from the figure), and the other source expresses a credal set that constitutes strong evidence for the state x_2 . Since the highly imprecise credal set contains probability functions that constitute strong evidence for each of the states in Ω_X , such a credal set is not significantly affected by other operands, unless these contain probability functions that are considerably stronger.⁴

⁴Arnborg [3] denotes the probability simplex as “total scepticism,” since such a set is impossible to affect.

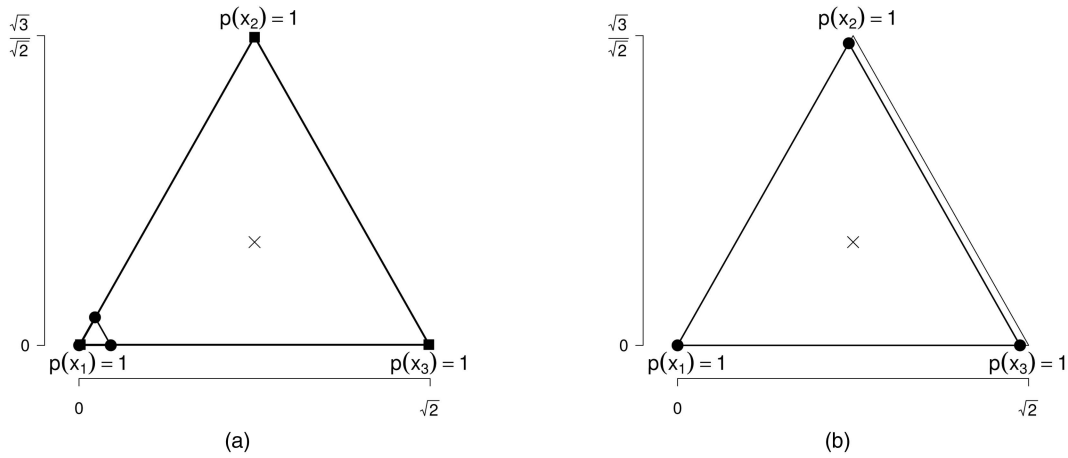


Fig. 6. $\mathcal{P}_1(X)$, $\mathcal{P}_2(X)$ and $\mathcal{P}_{1,2}(X)$ when a high degree of conflict is present. (a) $\mathcal{P}_1(X)$ (circles) and $\mathcal{P}_2(X)$ (squares). (b) $\mathcal{P}_{1,2}(X)$.

3.4. Discounting

In cases where a strong conflict is present among the sources that provide evidences, as was the case for several examples in the previous section, it may be beneficial to account for the sources' *reliability*. If one has obtained information regarding the reliability of sources, e.g., in terms of sensor quality, then it would be reasonable to compensate for such information prior to the combination. Intuitively, if a source is less reliable, then that source should have less effect on the end result, i.e., the joint evidence should be less influenced by that source. Accounting for the reliability of sources is commonly referred to as *discounting* in the literature [23].

3.4.1. Bayesian Discounting Operator

Discounting with respect to the Bayesian combination operator is performed by transforming an operand to a "more uniform" probability function. The reason for this is that the uniform probability function represents an evidence that does not affect the joint evidence in any way:⁵

DEFINITION 11 The Bayesian discounting operator for an evidence in the form of a probability function $p(X)$ with state space Ω_X is defined as

$$\Psi_B(p(X), w) \triangleq wp(X) + (1 - w)p_u(X) \quad (25)$$

where $w \in [0, 1]$ is a reliability weight, describing a degree of reliability of the discounted source, and $p_u(X)$ is the uniform distribution over Ω_X .

Let us now revisit the example in Fig. 3 but where we have obtained the following reliability weights for

⁵Arnborg [2] adopted another interpretation of discounting, which amounts to increase the imprecision for an operand single probability function. We have based our interpretation of discounting on evidence theory [23], i.e., that a discounted operand should have less effect on the end result.

the sources⁶

$$w_1 = 0.85, \quad w_2 = 0.95. \quad (26)$$

Let us introduce the following short-hand notation

$$p_{w_i}(X) \triangleq \Psi_B(p_i(X), w_i) \quad (27)$$

where $i \in \{1, 2\}$ and

$$p_{w_1, w_2}(X) \triangleq \Phi_B(\Psi_B(p_1(X), w_1), \Psi_B(p_2(X), w_2)). \quad (28)$$

The result of applying the Bayesian discounting operator with the reliability weights in (26) is seen in Fig. 7. In contrast to the former case, where no discounting was performed, we can here see that due to that the first source is slightly more unreliable, the result is less influenced by that source. This is seen from the figure, since the resulting probability function is closer to the corner $p(x_2) = 1$ than $p(x_1) = 1$.

3.4.2. Credal Discounting Operator

Consider discounting a source that reports an operand credal set for the credal combination operator. Instead of using a single reliability weight, we here allow reliability weights to be expressed imprecisely⁷ by a *convex set of reliability weights* $\mathcal{W} \triangleq [\underline{w}, \bar{w}]$ where $[\underline{w}, \bar{w}] \subseteq [0, 1]$, i.e., an interval. If we generalize the Bayesian discounting operator to the credal case, we obtain an operator that point-wise discounts each distribution in the credal set with respect to each reliability

⁶We introduce the Bayesian and credal discounting operators without specifying an exact interpretation of the reliability weights. In principle such a interpretation can differ depending on the application. Exploring different modeling schemas and interpretations for reliabilities is a topic for future research.

⁷Imprecision in reliability weights was inspired by Troffaes [24].

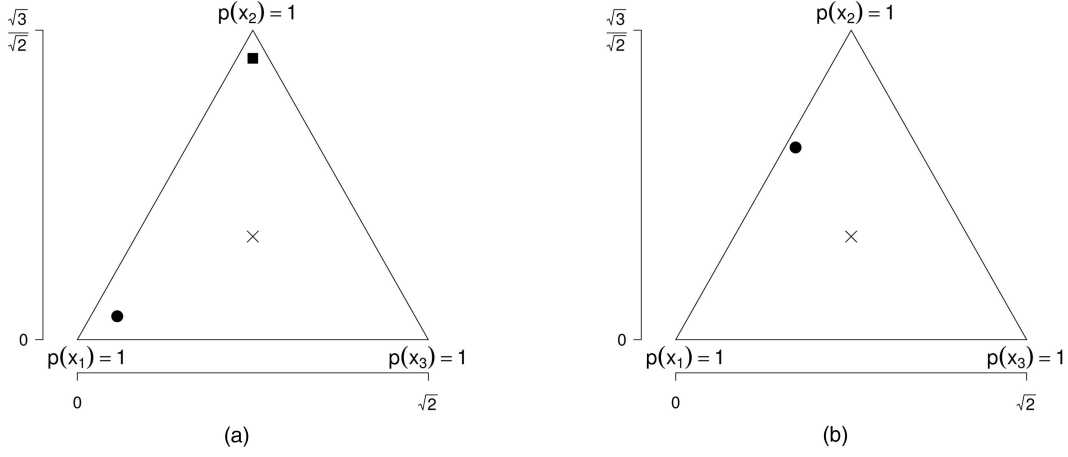


Fig. 7. $p_{w_1}(X)$, $p_{w_2}(X)$, and the discounted combined result $p_{w_1, w_2}(X)$. (a) $p_{w_1}(X)$ (circle) and $p_{w_2}(X)$ (square). (b) $p_{w_1, w_2}(X)$.

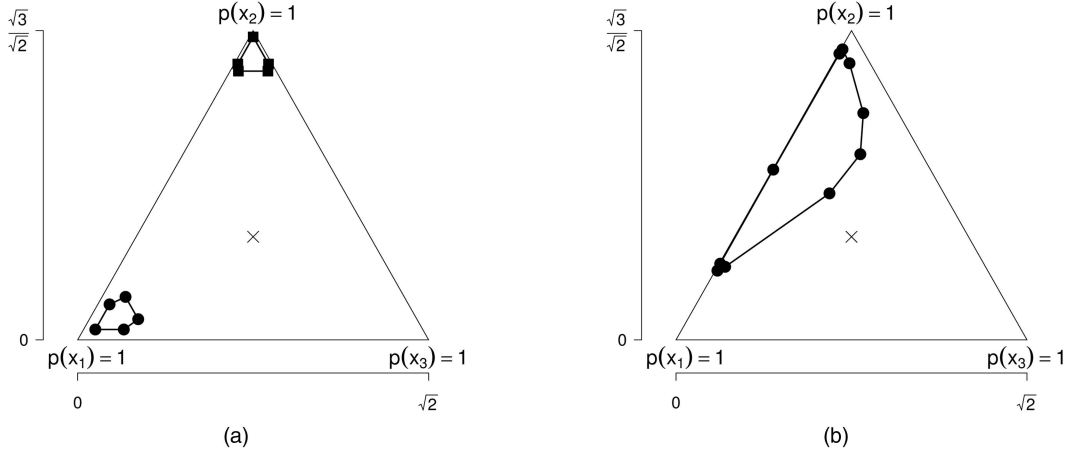


Fig. 8. $\mathcal{P}_{w_1}(X)$, $\mathcal{P}_{w_2}(X)$, and $\mathcal{P}_{w_1, w_2}(X)$. (a) $\mathcal{P}_{w_1}(X)$ (circles) and $\mathcal{P}_{w_2}(X)$ (squares). (b) $\mathcal{P}_{w_1, w_2}(X)$.

weight in \mathcal{W} :

DEFINITION 12 The discounting operator for a credal set $\mathcal{P}(X)$ given a set of reliability weights $\mathcal{W} = [\underline{w}, \bar{w}]$, where $[\underline{w}, \bar{w}] \subseteq [0, 1]$, is defined as

$$\begin{aligned} \Psi_C(\mathcal{P}(X), \mathcal{W}) \\ \triangleq \mathcal{CH}(\{\Psi_B(p(X), w) : w \in \mathcal{W}, p(X) \in \mathcal{P}(X)\}) \end{aligned} \quad (29)$$

where $\Psi_B(p(X), w)$ is the Bayesian discounting operator.

The discounting operator collapses a credal set point-wise towards the uniform distribution. The following theorem allows computation of the credal discounting operator by using the extreme points of the operand sets:

THEOREM 3

$$\Psi_C(\mathcal{P}(X), \mathcal{W}) = \Psi_C(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})). \quad (30)$$

PROOF See Appendix.

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

have obtained the following reliability weights⁶ for the example in Fig. 5

$$\mathcal{W}_1 = [0.80, 0.90], \quad \mathcal{W}_2 = [0.93, 0.98]. \quad (31)$$

Let us introduce some short-hand notation

$$\mathcal{P}_{\mathcal{W}_i}(X) \triangleq \Psi_C(\mathcal{P}_i(X), \mathcal{W}_i) \quad (32)$$

where $i \in \{1, 2\}$ and

$$\mathcal{P}_{\mathcal{W}_1, \mathcal{W}_2}(X) \triangleq \Phi_C(\Psi_C(\mathcal{P}_1(X), \mathcal{W}_1), \Psi_C(\mathcal{P}_2(X), \mathcal{W}_2)). \quad (33)$$

The results of applying the discounting operator is seen in Fig. 8. We see that there is a significant difference in terms of imprecision compared to the non-discounted case in Fig. 5(b).

Let us also revisit the example shown in Fig. 6. Assume that one has obtained the following reliabilities for the sources

$$\mathcal{W}_1 = [1.00, 1.00], \quad \mathcal{W}_2 = [0.75, 0.80]. \quad (34)$$

The result of discounting the sources with respect to these weights is seen in Fig. 9. The lower bound of

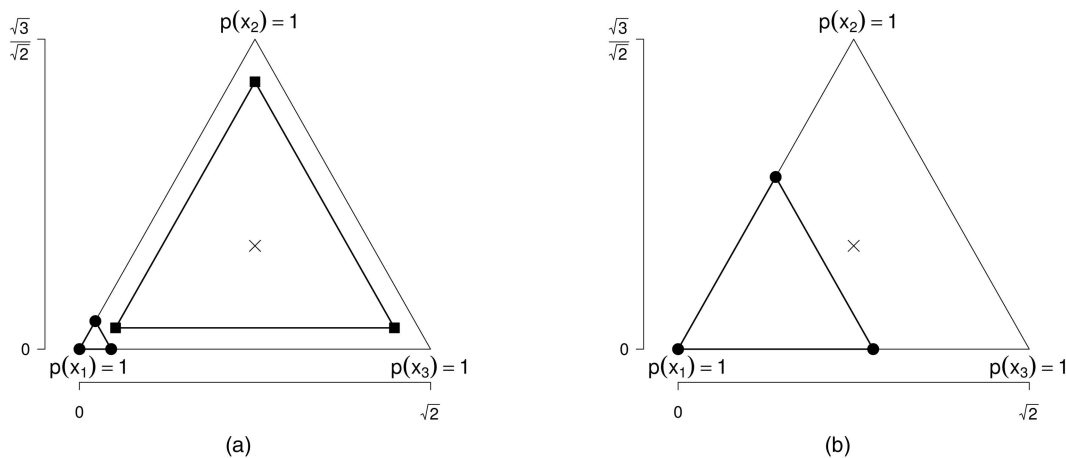


Fig. 9. $\mathcal{P}_{\mathcal{W}_1}(X)$, $\mathcal{P}_{\mathcal{W}_2}(X)$, and $\mathcal{P}_{\mathcal{W}_1, \mathcal{W}_2}(X)$. (a) $\mathcal{P}_{\mathcal{W}_1}(X)$ (circles) and $\mathcal{P}_{\mathcal{W}_2}(X)$ (squares). (b) $\mathcal{P}_{\mathcal{W}_1, \mathcal{W}_2}(X)$.

\mathcal{W}_2 will in this case not have any effect since $\mathcal{P}_2(X)$ is centered around the uniform distribution.

4. EMPIRICAL EVALUATION OF BAYESIAN AND CREDAL COMBINATION OPERATORS

In a previous study⁸ [17], we explored the performance of the Bayesian and credal combination operators when a single decision is made, i.e., a single state is chosen. In order to select a single decision in the credal case, we selected a representative function from the joint evidence to base the decision upon. We explored three main ways of selecting such a function, often found in the literature [3, 8], namely: a random function, the maximum entropy function, and the centroid function. By assuming that the sources have an implicit uniform second-order distribution over the operand credal set, as a representation of not favoring any probability function within the set, we found that using the Bayesian combination operator on centroids of operand credal sets significantly outperform⁹ any of the credal decision schemas (i.e., using the credal combination operator and any of the representative function previously mentioned). The reason for this result is that the second-order distribution can be considerably skewed over the joint evidence and using centroid distributions of operand credal sets as operands for the Bayesian combination operator constitutes a better approximation of the expected value with respect to this skewed second-order distribution over the joint evidence.

Even though it may be better to utilize the Bayesian combination operator than the credal counterpart when a single decision has to be made, the question remains whether or not the credal combination operator can be beneficial to utilize when a set of decisions is allowed. In principle, an optimal method for “set-output”

should only output a non-singleton set when the singleton decision output from a Bayesian method is erroneous.

Exploring the performance of utilizing the Bayesian and credal combination operators when decision sets are allowed (i.e., a set of states) is the main aim with our experiments that we will present in the coming sections.¹⁰ In the experiments, we use a simple state space consisting of three states since we then can perform exact computation of the credal combination operator. In real-world applications one is likely to be forced to use some approximations technique in order to limit the number of extreme points of the involved credal sets, since this number can grow exponentially in worst case (with respect to the number of combinations). We present two main experiments for combining evidences reported by a number of sources where there exists some degree of conflict between them. We motivate the utilization of a decision set by a risk component, i.e., there is a large negative cost if one reports a set that does not contain the true state. In the first experiment there is no risk component whereas in the second experiment such a component exists.

4.1. Experiment A—No Risk

Let us start with a scenario that does not contain a risk component in the sense that there is no cost of reporting an erroneous state. Assume that we are interested in determining the state of a random variable X with a state space consisting of three possible states, i.e., $\Omega_X = \{x_1, x_2, x_3\}$, and that we base our decision regarding X on n sources that provide us with pieces of evidence regarding X in the form of strongly conditionally independent credal sets $\mathcal{P}_1(X), \dots, \mathcal{P}_n(X)$. Assume that the true state of X is x_2 . Obviously, if we have selected the sources well, a majority of these provides us with credal sets $\hat{\mathcal{P}}(X)$ that constitute evidence for the

⁸In Karlsson, et al. [17] we considered the problem of belief updating instead of evidence combination, however, it is the same basic operator that is used for both cases.

⁹Two score functions (i.e., performance metrics) were used for comparison: (1) *accuracy* and (2) *Brier loss* [6].

¹⁰The sections includes material from Karlsson, et al. [18].

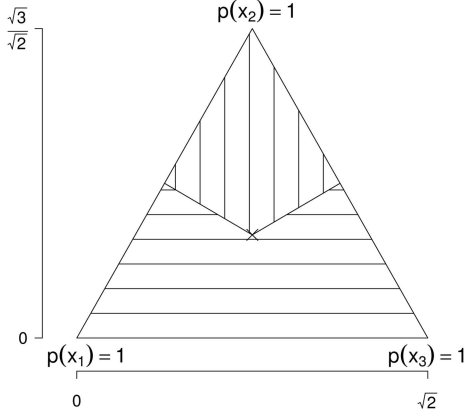


Fig. 10. The probability simplex $\mathcal{P}^*(X)$ partitioned into evidence region (vertical lines) and counter-evidence region (horizontal line) with respect to the true state x_2 .

truth solely, i.e.,

$$\hat{p}(X) \in \hat{\mathcal{P}}(X) \Rightarrow x_2 = \arg \max_{x \in \Omega_X} \hat{p}(x). \quad (35)$$

Such type of credal set is completely contained in the region with vertical lines shown in Fig. 10. Let us assume that there is a possibility of obtaining a counter evidence $\tilde{\mathcal{P}}(X)$ with respect to the truth from some of the sources, i.e.,

$$\tilde{p}(X) \in \tilde{\mathcal{P}}(X) \Rightarrow x_2 \neq \arg \max_{x \in \Omega_X} \tilde{p}(x). \quad (36)$$

The counter evidence $\tilde{\mathcal{P}}(X)$ is completely contained in the region with horizontal lines shown in Fig. 10. The imprecision of the credal evidence and counter evidence can be thought of as a *second-order uncertainty* regarding the strength of an evidence in the form of a probability function (i.e., a Bayesian evidence). Let us assume that the sources have no reason to favor any probability function in the credal evidence, i.e., the sources are *indifferent* regarding the probability functions.

Now, assume that we want to combine all the evidences obtained from the sources into a joint evidence. In the Bayesian case, since we cannot apply the Φ_B operator directly on the operand credal sets, we need to select a single representative probability function from the operands to be utilized for combination. Since the sources are indifferent regarding the probability functions in the operand credal sets, we can assume an implicit uniform distribution over the sets. It is therefore reasonable to utilize the *expected value* of this distribution as a representative function, i.e., the *centroid distribution*. Consequently we obtain the following joint Bayesian evidence

$$p_{1:n}(X) \triangleq \Phi_B(\dots \Phi_B(\Upsilon(\mathcal{P}_1(X)), \Upsilon(\mathcal{P}_2(X))), \dots, \Upsilon(\mathcal{P}_n(X))) \quad (37)$$

where the operator Υ is defined as

$$\Upsilon(\mathcal{P}(X)) \triangleq E_{\text{Un}(\mathcal{P}(X))}[\mathcal{P}(X)] \quad (38)$$

where $\text{Un}(\mathcal{P}(X))$ denotes the uniform distribution over $\mathcal{P}(X)$ (i.e., $\Upsilon(\mathcal{P}(X))$ gives the centroid distribution of $\mathcal{P}(X)$). In the credal case, the joint evidence is straightforwardly obtained by utilizing the Φ_C operator

$$P_{1:n}(X) \triangleq \Phi_C(\dots \Phi_C(\mathcal{P}_1(X), \mathcal{P}_2(X)), \dots, \mathcal{P}_n(X)). \quad (39)$$

Now, based on the joint Bayesian and credal evidences, we want to make a decision regarding the true state of the variable X . In the Bayesian case this is simply performed by reporting the most probable state(s)

$$\begin{aligned} \mathcal{D}_B(p_{1:n}(X)) \\ \triangleq \{x_i \in \Omega_X : (\forall x_j \in \Omega_X)(p_{1:n}(x_i) \geq p_{1:n}(x_j))\}. \end{aligned} \quad (40)$$

From the above equation, we see that the *Bayesian decision set* $\mathcal{D}_B(p_{1:n}(X))$ is singleton in a majority of the cases. In the credal case, however, it is quite likely, depending on the degree of imprecision reported by the sources, that the decision set is non-singleton

$$\mathcal{D}_C(\mathcal{P}_{1:n}(X)) \triangleq \bigcup_{p_{1:n}(X) \in \mathcal{P}_{1:n}(X)} \mathcal{D}_B(p_{1:n}(X)). \quad (41)$$

Note that unless all probability functions within $\mathcal{P}_{1:n}(X)$ agree on the most probable state, the decision set is non-singleton. Let us also introduce a credal method where the Φ_C operator is used for constructing the joint credal evidence but where the centroid distribution of the evidence is used for decision making in the same way as in the Bayesian case (see (40)), i.e.,

$$\mathcal{D}_C^c(\mathcal{P}_{1:n}(X)) \triangleq \mathcal{D}_B(\Upsilon(\mathcal{P}_{1:n}(X))). \quad (42)$$

An example of using the Φ_B and Φ_C operators is seen in Fig. 11. In Fig. 11(a), we see that one of the sources has reported a quite strong evidence for the truth x_2 while the other source has reported a counter evidence to this state (an evidence for x_1). Fig. 11(b) shows the results of the Bayesian and credal methods. We see that

$$\begin{aligned} \mathcal{D}_B(p_{1:2}(X)) &= \{x_2\} \\ \mathcal{D}_C(\mathcal{P}_{1:2}(X)) &= \{x_1, x_2\} \\ \mathcal{D}_C^c(\mathcal{P}_{1:2}(X)) &= \{x_2\}. \end{aligned} \quad (43)$$

Note that the centroid of the joint credal evidence differs from the joint Bayesian evidence.

Now, obviously a decision set $\mathcal{D} \subseteq \Omega_X$ that contains two states where one of them is the true state, i.e., x_2 , should be less valued in comparison to a decision set that is singleton with the true state. Moreover, a decision set that is equal to the state space is clearly non-informative about X since we have already modeled the set of possibilities for X by Ω_X . Hence such decision set is not regarded to be of any value. Based on this

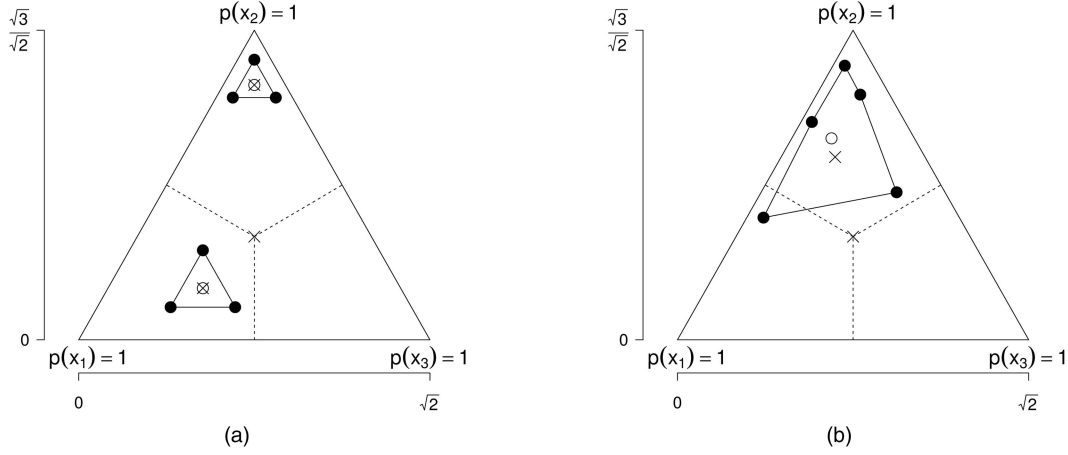


Fig. 11. The example shows the probability simplex $\mathcal{P}^*(X)$ where one operand constitutes evidence for the true state x_2 and the other counter evidence to the true state (in this case evidence for x_1). The dashed lines show the decision regions for the state space. The extreme points of the credal operands $\mathcal{P}_i(X)$, $i \in \{1, 2\}$ and joint credal evidence $\mathcal{P}_{1:2}(X)$ are depicted by filled circles. The centroid of $\mathcal{P}^*(X)$ and $\mathcal{P}_{1:2}(X)$ is depicted by a cross. The Bayesian joint evidence $p_{1:2}(X)$ is depicted by an unfilled circle. (a) $\mathcal{P}_i(X)$ and $E_{\text{Un}(\mathcal{P}_i(X))}[\mathcal{P}_i(X)]$, $i \in \{1, 2\}$. (b) $\mathcal{P}_{1:2}(X)$ and $p_{1:2}(X)$.

reasoning we adopt the following *score function* for our experiment

$$f_\alpha(\mathcal{D}) \triangleq \begin{cases} \frac{1}{|\mathcal{D}|}, & \text{if } x_2 \in \mathcal{D}, \mathcal{D} \neq \Omega_X \\ 0, & \text{if } \mathcal{D} = \Omega_X \\ -\alpha, & \text{otherwise} \end{cases} \quad (44)$$

where $\mathcal{D} \subseteq \Omega_X$ and α models a risk component. As we stated in the beginning of this section, we first explore the performance of the methods when no risk is involved in the decision problem, hence, we instantiate the score function with $\alpha = 0$, i.e., $f_0(\mathcal{D})$.

Let the probability for the event that a source reports an evidence with respect to the truth (i.e., x_2) be denoted by β . Note that if we sum the degree of conflict for both the Bayesian and credal conflict measures for all $n - 1$ combinations, i.e.,

$$\Gamma_{\bullet}^{1:n-1} \triangleq \sum_{i=1}^{n-1} \Gamma^i \quad (45)$$

where Γ_{\bullet}^i (Definitions 8 and 9) denotes the conflict in the i th combination, then we expected $\Gamma_{\bullet}^{1:n-1}$ to increase when β monotonically decreases in the interval $[0.5, 1]$, i.e., the total amount of conflict among the sources increases. The experiment can now be defined by the following step-wise description:

1. Sample the number of sources $n \sim \text{Un}([5, 10])$.
2. Sample the probability for obtaining an evidence for the true state $\beta \sim \text{Un}([0.7, 0.9])$.
3. Sample evidences $\mathcal{P}_1(X), \dots, \mathcal{P}_n(X)$ where the probability of sampling an evidence $\hat{\mathcal{P}}_i(X)$ for the truth is β (see (35)) and $1 - \beta \triangleq$ for a counter evidence $\tilde{\mathcal{P}}_i(X)$ (see (36)), $i \in \{1, \dots, n\}$.
4. Calculate joint evidences $p_{1:n}(X)$ and $\mathcal{P}_{1:n}(X)$.

5. Calculate decision sets $\mathcal{D}_B(p_{1:n}(X))$, $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$, and $\mathcal{D}_C^c(\mathcal{P}_{1:n}(X))$.

6. Calculate the score $f_\alpha(\cdot)$ for each decision set in the previous step.

7. Repeat $m = 10^5$ times.

Remember that we have instantiated $\alpha = 0$ for this first experiment, i.e., there is no risk component involved. Let us elaborate somewhat on the implementation detail of the above description. In step three, we sample evidences by first deciding, utilizing β , if a specific source should report an evidence or a counter evidence for the truth. Then, when we know if it is an evidence or counter evidence that we should sample, we sample a centroid from the corresponding region (see Fig. 10), uniformly. Given the centroid, we sample imprecision by considering the distance from the centroid to the corner points of an equilateral triangle, under the condition that all corner points should reside in the same evidence region. Hence, the credal operand that we sample are all equilateral triangles (simplices) that are completely contained in the evidence or counter-evidence region with respect to the truth (e.g., Fig. 11(a)). Credal sets of this form can be obtained by interval constraints on marginal probabilities.

4.4.1. Results

The results of the experiment is seen in Table I. We see that the expected score of the Bayesian method \mathcal{D}_B is clearly better than the credal method \mathcal{D}_C . This means that the credal method does not isolate the cases for which the Bayesian method performs poorly in an optimal way, since we would then have expected a higher score for the credal method than the Bayesian one. This is seen from Table I since in 21.9% of the cases the credal method outputs a non-singleton set while the Bayesian method only outputs an erroneous state in 7.2% of the cases. The credal method outputs a

TABLE I
Expected Score $E[f_0(\cdot)]$, with 95% Confidence Intervals, for $\mathcal{D}_B(p_{1:n}(X))$, $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$ and $\mathcal{D}_C^c(\mathcal{P}_{1:n}(X))$

Method	$E[f_0(\cdot)]$	$f_0(\cdot) > 0$ (%)		$f_0(\cdot) = 0$ (%)		
		$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 3$
$\mathcal{D}_B(p_{1:n}(X))$	0.93 ± 0.002	92.8	0.0	7.2	0.0	0.0
$\mathcal{D}_C(\mathcal{P}_{1:n}(X))$	0.85 ± 0.002	78.1	13.1	1.8	0.4	6.6
$\mathcal{D}_C^c(\mathcal{P}_{1:n}(X))$	0.92 ± 0.002	91.6	0.0	8.4	0.0	0.0

decision set of no value (i.e., x_2 is not in the decision set or Ω_X is reported) in 8.8% of the cases. In fact, even if we would let the credal method obtain a reward of one in cases where two states is reported and one of them is the truth, the credal method would still perform worse than the Bayesian method (78.1% + 13.1% = 91.2% compared to 92.8%). Also note that the Bayesian method \mathcal{D}_B performs better than the credal centroid method \mathcal{D}_C^c , however, the difference is not as high compared to the former case.

4.2. Experiment B—Risk

One argument that one might have for using the credal method \mathcal{D}_C is that even though it cannot optimally isolate the cases where the Bayesian method \mathcal{D}_B performs poorly, it can still be an interesting choice when there exists a risk component in the decision problem, i.e., reporting an erroneous state is coupled with a negative cost. Indeed, if we use the result from Table I, we see that the Bayesian method reports an erroneous state in 7.2% of the cases while the credal method only makes erroneous reports in 1.8% + 0.4% = 2.2% of the cases. Hence, if we would have set $\alpha = 10$ in the score function in (44), we would have obtained an expected score $E[f_{10}(\mathcal{D}_B(p_{1:n}(X)))] \approx 0.21$ for the Bayesian method and $E[f_{10}(\mathcal{D}_C(\mathcal{P}_{1:n}(X)))] \approx 0.63$ for the credal correspondence. However, when risk is incorporated in the decision problem, there clearly exist cases when using the Bayesian method for which one would not simply output the single state that maximizes the probability, e.g., whenever the joint Bayesian evidence $p_{1:n}(X)$ is close to the uniform distribution. Let us therefore modify \mathcal{D}_B to a *cautious Bayesian method* \mathcal{D}_B^δ in the following way

$$\mathcal{D}_B^\delta(p_{1:n}(X)) \triangleq \{x \in \Omega_X : p_{1:n}(x) > \delta\} \quad (46)$$

where $\delta \in [0, |\Omega_X|^{-1}]$. The method partitions the probability simplex into decision regions, seen in Fig. 12. Note that a high value of δ yields a less cautious method and vice versa. Note that when $\delta = 0$ we get $\mathcal{D}_B^0(p_{1:n}(X)) = \Omega_X$ for all joint evidences $p_{1:n}(X)$ that do not reside on the boundary of the probability simplex. Also note that when $\delta = |\Omega_X|^{-1}$ we still have decision regions that are non-singleton.

Now let us use the same simulation settings as in Experiment A (Section 4.1) but where we now introduce a risk component by setting $\alpha = 10$, yielding a

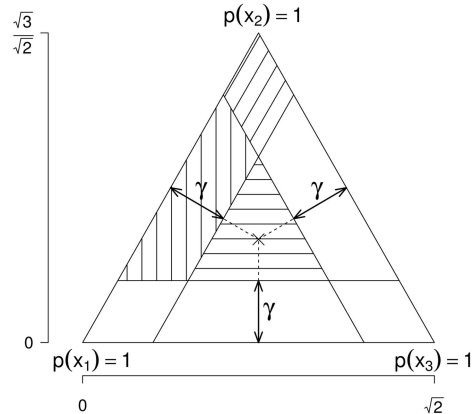


Fig. 12. An example of the cautious Bayesian method in (46) where $\delta = 0.2$. The parameter δ imposes decision regions by planes that are parallel to the (proper) faces of the simplex with a distance $\gamma = \delta(\sqrt{3}/\sqrt{2})$. The horizontal lines depict the decision region for Ω_X , the vertical lines depict $\{x_1, x_2\}$, and lastly the region with diagonal lines depicts $\{x_2\}$.

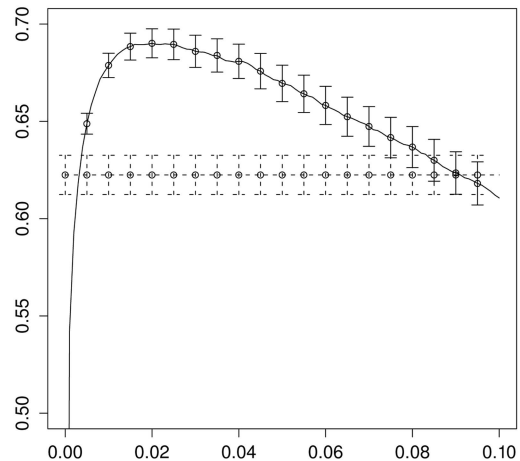


Fig. 13. The solid line shows the cautious Bayesian method $\mathcal{D}_B^\delta(p_{1:n}(X))$ and the dashed line the credal method $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$. The x-axis depicts δ and the y-axis $E[f_{10}(\cdot)]$. Confidence intervals on the 95%-level are also shown.

score function f_{10} . We perform the simulation for at set of values of the parameter $\delta \in [0, |\Omega_X|^{-1}]$ to see if there exist parameters that cause the cautious Bayesian method to outperform the credal method.

4.2.1. Results

The result is shown in Fig. 13. The cautious Bayesian method outperforms the credal method when $\delta \in [0.005,$

TABLE II
Expected Score $E[f_{10}(\cdot)]$, with 95% Confidence Intervals, for the Methods $\mathcal{D}_B^{0.02}(p_{1:n}(X))$ and $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$

Method	$E[f_{10}(\cdot)]$	$f_{10}(\cdot) \geq 0$ (%)			$f_{10}(\cdot) < 0$ (%)	
		$ \cdot = 1$	$ \cdot = 2$	$ \cdot = 3$	$ \cdot = 1$	$ \cdot = 2$
$\mathcal{D}_B^{0.02}(p_{1:n}(X))$	0.69 ± 0.006	69.6	22.1	7.1	0.7	0.5
$\mathcal{D}_C(\mathcal{P}_{1:n}(X))$	0.62 ± 0.010	77.9	13.2	6.7	1.8	0.4

TABLE III

The table shows the cautious Bayesian parameter δ in $\mathcal{D}_B^\delta(p_{1:n}(X))$ for different risks α . The intervals for δ depicts the region for which $\mathcal{D}_B^\delta(p_{1:n}(X))$ outperforms $\mathcal{D}_C(\mathcal{P}_{1:n}(X))$ with respect to $E[f_\alpha(\cdot)]$ and where the 95% confidence intervals are non-overlapping for the methods.

α	0	2	4	6	8	10
δ	[0.05, 0.33]	[0.04, 0.33]	[0.03, 0.14]	[0.02, 0.10]	[0.01, 0.08]	[0.01, 0.07]

0.07]. Let us explore the cautious Bayesian method at its peak performance, which approximately occur at $\delta = 0.02$. The result of this parameter value is seen in Table II. We see that the cautious Bayesian method tends to output a non-singleton set more often than the credal counterpart. However, the Bayesian method only reports an erroneous state in $0.7\% + 0.5\% = 1.2\%$ of the cases compared to $1.8\% + 0.4\% = 2.2\%$ in the credal case and due to the high risk component this yields a better score for the Bayesian method. Note that since δ is quite low and the cautious Bayesian method only outputs Ω_X in approximately 7.1% of the cases, we can conclude that the joint Bayesian evidence $p_{1:n}(X)$ is close to the boundary of the probability simplex in a majority of the cases, which is quite natural since we have assumed that a majority ($\beta \in [0.7, 0.9]$) of the sources output evidence for the true state x_2 . Let us further study the sensitivity of the parameter δ by exploring how a parameter set for δ changes with respect to the risk for the case where the cautious Bayesian method outperforms the credal method (where the 95% confidence intervals are non-overlapping), seen in Table III. From the table we see that when there is a low risk, $\alpha \in \{0, 2\}$, the parameter sets for δ where \mathcal{D}_B^δ outperforms \mathcal{D}_C are quite large. When the risk increases, $\alpha \in \{4, 6, 8, 10\}$, the parameter sets becomes considerably smaller. From the table we see that \mathcal{D}_B^δ with $\delta = 0.05$ performs better than \mathcal{D}_C irrespective of the risk α .

5. SUMMARY AND CONCLUSIONS

We have characterized the behavior of the Bayesian and credal combination operators through a number of examples. We introduced measures for degree of conflict and imprecision and explored the behavior of the Bayesian and credal combination operators for a number of examples where different degrees of conflict and imprecision were present. We highlighted that when a strong conflict is present between the sources that report credal sets, the joint evidence can be highly imprecise. We therefore introduced Bayesian and credal discounting operators that can be utilized whenever

information about the reliability of sources is available. We showed that the credal discounting operator can be computed by utilizing the extreme points of the operands (credal set and interval of reliability weights). Finally, we showed that the credal discounting operator can have a significant impact on the combined result when used.

Both the Bayesian and credal discounting operators have been introduced so that they are consistent with the underlying paradigm in the respective theories, i.e., the Bayesian discounting operator takes a single reliability weight as an argument while the credal such operator takes a convex set of reliability weights in the form of an interval. Hence the Bayesian discounting operator assumes that a precise reliability weight can always be formulated while in the credal case imprecision is allowed. Moreover, the credal discounting operator preserves the intuitive paradigm of being a point-wise version of the Bayesian counterpart, i.e., the operator discounts each probability function within the credal set with respect to each of the reliability weight in the set (interval) of such weights. The credal discounting operator has three properties that makes it unique [13]: (1) it can discount any credal set (i.e., it is not restricted to a particular type of credal set), (2) a credal set can be discounted with respect to a set of reliability weights, i.e., one can express reliability imprecisely, and (3) a discounted credal set can be reversed to its original form if the set of reliability weights, used for the discounting, is known.

We also performed two experiments where we evaluate the Bayesian and credal combination operators. In both experiments the sources report credal sets with an implicit uniform second-order distribution as a representation of not favoring any probability function within the sets. For the Bayesian combination operator, we have utilized the expected value of the operand credal sets, i.e., centroids, for obtaining the joint evidence. We evaluated the operators by using a simple score function that gives a reward corresponding to the informativeness

of the joint evidence and a loss according to a specified risk.

In the first experiment, we showed that in scenarios where there exists no risk component, i.e., no cost of reporting an erroneous decision set, it is clearly beneficial to utilize the Bayesian combination operator instead of the credal correspondence. This is true even if one would maintain imprecision by using the credal combination operator and lastly utilize the centroid for decision making. However, this difference in performance was not as clear as in the previous case. Nevertheless, the latter results show that nothing is gained by maintaining imprecision and then using the centroid for constructing the decision set.

By using the result from the experiment, we concluded that if a large risk component is present, the credal method is preferred due to a lower number of erroneous decision sets. However, we introduced a simple cautious Bayesian method, using a single parameter that partitions the probability simplex into regions corresponding to different decision sets, and we showed that such a method can outperform the credal correspondence. One potential problem with the cautious Bayesian method is that one needs to choose an appropriate parameter value. However, we showed that there exist values for which the method outperforms the credal method for a set of risk components.

In essence our results tells us that if there is no risk component in the scenario of interest, then one should use the Bayesian combination operator, even if the sources choose to report imprecision by credal sets. Furthermore, if a risk component does exist in the scenario, then one should use the cautious Bayesian method that we introduced. Hence for both cases it is sufficient to use a single probability function and the Bayesian combination operator for representing respectively combining evidences. From the perspective of computational complexity this is indeed positive results, considering that the number of extreme points of the joint credal evidence in the worst case can grow exponentially with the number of combinations.

The question then is if there exist cases where one might want to maintain imprecision by using the credal combination operator? One possible such scenario could be when there is a human decision maker involved in the scenario, in particular when there exists a risk component. In such cases the decision maker might want to use the credal combination operator in order to maintain imprecision for the purpose of keeping track of worst-case scenarios with respect to the risk.

ACKNOWLEDGMENTS

Computations have been performed with R [22]. 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.infusion.se>).

APPENDIX

THEOREM 1

$$\begin{aligned} & \Phi(\dots\Phi(p_1(X), p_2(X))\dots, p_n(X)) \\ &= \frac{p_1(X)\dots p_n(X)}{\sum_{x\in\Omega_X} p_1(x)\dots p_n(x)}. \end{aligned} \quad (47)$$

PROOF The proof is by induction. Let us introduce the following shorthand notation

$$p_{1:n}(X) \triangleq \Phi(\Phi(\dots\Phi(p_1(X), p_2(X))\dots, p_{n-1}(X)), p_n(X)). \quad (48)$$

The base case

$$p_{1:2}(X) = \frac{p_1(X)p_2(X)}{\sum_{x\in\Omega_X} p_1(x)p_2(x)} \quad (49)$$

holds by (4). Let the induction hypothesis be

$$p_{1:n-1}(X) = \frac{p_1(X)\dots p_{n-1}(X)}{\sum_{x\in\Omega_X} p_1(x)\dots p_{n-1}(x)}. \quad (50)$$

We need to show that such assumption implies

$$p_{1:n}(X) = \frac{p_1(X)\dots p_n(X)}{\sum_{x\in\Omega_X} p_1(x)\dots p_n(x)}. \quad (51)$$

We have that

$$p_{1:n}(X) = \frac{p_{1:n-1}(X)p_n(X)}{\sum_{x\in\Omega_X} p_{1:n-1}(x)p_n(x)}. \quad (52)$$

By using the induction hypothesis in (50), we get

$$\begin{aligned} p_{1:n}(X) &= \frac{\frac{p_1(X)\dots p_{n-1}(X)}{\sum_{x\in\Omega_X} p_1(x)\dots p_{n-1}(x)} p_n(X)}{\sum_{x\in\Omega_X} \frac{p_1(x)\dots p_{n-1}(x)}{\sum_{x\in\Omega_X} p_1(x)\dots p_{n-1}(x)} p_n(x)} \\ &= \frac{p_1(X)\dots p_n(X)}{\sum_{x\in\Omega_X} p_1(x)\dots p_n(x)}. \end{aligned} \quad (53)$$

By (49)–(53), the proof is complete.

THEOREM 2

$$\begin{aligned} & \Phi_c(\mathcal{P}_1(X), \mathcal{P}_2(X)) \\ &= \Phi_c(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))). \end{aligned} \quad (54)$$

PROOF The proof is inspired by Noack, et al. [21, Theorem 2]. First note that

$$\Phi_c(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))) \subseteq \Phi_c(\mathcal{P}_1(X), \mathcal{P}_2(X)) \quad (55)$$

is trivial. Assume that

$$\Phi_c(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))) \subset \Phi_c(\mathcal{P}_1(X), \mathcal{P}_2(X)). \quad (56)$$

Then there must exist at least one

$$u(X) \in \Phi_c(\mathcal{P}_1(X), \mathcal{P}_2(X)) \quad (57)$$

such that

$$u(X) \notin \Phi_c(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))) \quad (58)$$

where $u(X)$ has the following form

$$u(X) = \frac{p_1(X)p_2(X)}{\sum_{x \in \Omega_X} p_1(x)p_2(x)} \quad (59)$$

and where $p_1(X) \in \mathcal{P}_1(X)$ and $p_2(X) \in \mathcal{P}_2(X)$, where at least one of $p_1(X)$ and $p_2(X)$ is not an extreme point. We can express $p_1(X)$ and $p_2(X)$ as

$$p_1(X) = \sum_{i=1}^m \lambda_i v_i(X) \quad (60)$$

$$p_2(X) = \sum_{j=1}^n \alpha_j w_j(X) \quad (61)$$

where $v_i(X) \in \mathcal{E}(\mathcal{P}_1(X))$, $w_j(X) \in \mathcal{E}(\mathcal{P}_2(X))$, $\lambda_i \geq 0$, $\alpha_j \geq 0$, $1 \leq i \leq m$, $1 \leq j \leq n$, $\sum_{i=1}^m \lambda_i = \sum_{j=1}^n \alpha_j = 1$, and where there exists at least one $\lambda_i \in (0, 1)$ or $\alpha_j \in (0, 1)$. By using (60) and (61) in (59), we obtain

$$u(X) = \frac{\sum_{i=1}^m \sum_{j=1}^n \lambda_i \alpha_j v_i(X) w_j(X)}{\sum_{x \in \Omega_X} \left(\sum_{i=1}^m \sum_{j=1}^n \lambda_i \alpha_j v_i(x) w_j(x) \right)}. \quad (62)$$

Let us introduce the following notation

$$\gamma_{i,j} \triangleq \frac{\lambda_i \alpha_j \sum_{x \in \Omega_X} v_i(x) w_j(x)}{\sum_{x \in \Omega_X} \left(\sum_{i=1}^m \sum_{j=1}^n \lambda_i \alpha_j v_i(x) w_j(x) \right)}. \quad (63)$$

We can now rephrase $u(X)$ as

$$u(x) = \sum_{i=1}^m \sum_{j=1}^n \gamma_{i,j} \frac{v_i(X) w_j(X)}{\sum_{x \in \Omega_X} v_i(x) w_j(x)}. \quad (64)$$

Since

$$\frac{v_i(X) w_j(X)}{\sum_{x \in \Omega_X} v_i(x) w_j(x)} \in \Phi_c(\mathcal{E}(\mathcal{P}_1(X)), \mathcal{E}(\mathcal{P}_2(X))) \quad (65)$$

and $\gamma_{i,j} \geq 0$, $\sum_{i=1}^m \sum_{j=1}^n \gamma_{i,j} = 1$, we get (cf (8)):

$$u(X) \in \Phi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{P}(X))) \quad (66)$$

which is a contradiction.

THEOREM 3

$$\Psi_c(\mathcal{P}(X), \mathcal{W}) = \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})). \quad (67)$$

PROOF First note that

$$\Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})) \subseteq \Psi_c(\mathcal{P}(X), \mathcal{W}) \quad (68)$$

is trivial. Assume that

$$\Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})) \subset \Psi_c(\mathcal{P}(X), \mathcal{W}). \quad (69)$$

Then there must exist at least one

$$u(X) \in \mathcal{E}(\Psi_c(\mathcal{P}(X), \mathcal{W})) \quad (70)$$

such that

$$u(X) \notin \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})) \quad (71)$$

where $u(X)$ has the following form

$$u(X) = wp(X) + (1-w)p_u(X), \quad (72)$$

where $w \in \mathcal{W}$, and $p(X) \in \mathcal{P}(X)$, and where at least one of w and $p(X)$ is not an extreme point. There are three cases:

Case 1— $p(X) \in \mathcal{E}(\mathcal{P}(X))$, $w \notin \mathcal{E}(\mathcal{W})$: We know that $w = \lambda w_1 + (1-\lambda)w_2$ where $w_1 \neq w_2$, $w_1, w_2 \in \mathcal{E}(\mathcal{W})$, $\lambda \in (0, 1)$. We get

$$\begin{aligned} u(X) &= wp(X) + (1-w)p_u(X) \\ &= p_u(X) + (\lambda w_1 + (1-\lambda)w_2)(p(X) - p_u(X)) \\ &= p_u(X) + \lambda w_1(p(X) - p_u(X)) \\ &\quad + (1-\lambda)w_2(p(X) - p_u(X)) \\ &\quad + \lambda p_u(X) - \lambda p_u(X) \\ &= \lambda(p_u(X) + w_1(p(X) - p_u(X))) \\ &\quad + (1-\lambda)p_u(X) + (1-\lambda)w_2(p(X) - p_u(X)) \\ &= \lambda(p_u(X) + w_1(p(X) - p_u(X))) \\ &\quad + (1-\lambda)(p_u(X) + w_2(p(X) - p_u(X))) \\ &= \lambda(w_1 p(X) + (1-w_1)p_u(X)) \\ &\quad + (1-\lambda)(w_2 p(X) + (1-w_2)p_u(X)). \end{aligned} \quad (73)$$

Hence, $u(X) \in \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W}))$ (cf (8)), which is a contradiction.

Case 2— $p(X) \notin \mathcal{E}(\mathcal{P}(X))$, $w \in \mathcal{E}(\mathcal{W})$: We know that $p(X) = \sum_{i=1}^n \alpha_i p_i(X)$, where $p_i(X) \in \mathcal{E}(\mathcal{P}(X))$, $\alpha_i \geq 0$, $\sum_{i=1}^n \alpha_i = 1$ where there exists at least one $\alpha_i \in (0, 1)$. We get

$$\begin{aligned} u(X) &= w \left(\sum_{i=1}^n \alpha_i p_i(X) \right) + (1-w)p_u(X) \\ &\quad + \left(\sum_{i=1}^n \alpha_i (1-w)p_u(X) \right) \\ &\quad - \left(\sum_{i=1}^n \alpha_i (1-w)p_u(X) \right) \\ &= \left(\sum_{i=1}^n \alpha_i (w p_i(X) + (1-w)p_u(X)) \right) \\ &\quad + (1-w)p_u(X) \\ &\quad - \left(\sum_{i=1}^n \alpha_i (1-w)p_u(X) \right) \\ &= \sum_{i=1}^n \alpha_i (w p_i(X) + (1-w)p_u(X)). \end{aligned} \quad (74)$$

Hence, $u(X) \in \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W}))$ (cf (8)), which is a contradiction.

Case 3— $p(X) \notin \mathcal{E}(\mathcal{P}(X))$, $w \notin \mathcal{E}(\mathcal{W})$: Similar to Case 1 and 2, we have that

$$\begin{aligned} w &= \lambda w_1 + (1-\lambda)w_2 \\ p(X) &= \sum_{i=1}^n \alpha_i p_i(X). \end{aligned} \quad (75)$$

We get

$$u(X) = (\lambda w_1 + (1 - \lambda)w_2) \left(\sum_{i=1}^n \alpha_i p_i(X) \right) + (1 - (\lambda w_1 + (1 - \lambda)w_2)) p_u(X). \quad (76)$$

From Case 1 we know that the above equation is equivalent to

$$u(X) = \lambda \left(w_1 \left(\sum_{i=1}^n \alpha_i p_i(X) \right) + (1 - w_1) p_u(X) \right) + (1 - \lambda) \left(w_2 \left(\sum_{i=1}^n \alpha_i p_i(X) \right) + (1 - w_2) p_u(X) \right). \quad (77)$$

From Case 2 we know that the above equation is equivalent to

$$u(X) = \lambda \left(\sum_{i=1}^n \alpha_i (w_1 p_i(X) + (1 - w_1) p_u(X)) \right) + (1 - \lambda) \left(\sum_{i=1}^n \alpha_i (w_2 p_i(X) + (1 - w_2) p_u(X)) \right). \quad (78)$$

Hence, $u(X) \in \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W}))$ (cf (8)), which is a contradiction. Since all possible cases lead to contradictions we must conclude that

$$\Psi_c(\mathcal{P}(X), \mathcal{W}) = \Psi_c(\mathcal{E}(\mathcal{P}(X)), \mathcal{E}(\mathcal{W})). \quad (79)$$

REFERENCES

- [1] N. Andréasson, A. Evgrafov, and M. Patriksson
An Introduction to Continuous Optimization.
Studentlitteratur, Lund, Sweden, 2005.
- [2] S. Arnborg
Robust Bayesianism: Imprecise and paradoxical reasoning.
In *Proceedings of the 7th International Conference on Information fusion*, June 2004, 407–414.
- [3] S. Arnborg
Robust Bayesianism: relation to evidence theory.
Journal of Advances in Information Fusion, **1**, 1 (Apr. 2006), 75–90.
- [4] J. O. Berger
An overview of robust Bayesian analysis.
Test, **3** (1994), 5–124.
- [5] J. M. Bernardo and A. F. M. Smith
Bayesian Theory.
John Wiley and Sons, 2000.
- [6] G. W. Brier
Verification of forecasts expressed in terms of probability.
Monthly Weather Review, **78** (Feb. 1950), 1–3.
- [7] I. Couso, S. Moral, and P. Walley
A survey of concepts of independence for imprecise probabilities.
Risk Decision and Policy, **5** (Sept. 2000), 165–181.
- [8] F. Cozman
Decision Making Based on Convex Sets of Probability Distributions: Quasi-Bayesian Networks and Outdoor Visual Position Estimation.
Ph.D. thesis, The Robotics Institute, Carnegie Mellon University, 1997.
- [9] F. Cozman
A derivation of quasi-Bayesian theory.
Technical Report CMU-RI-TR-97-37, Robotics Institute, Carnegie Mellon University, 1997.
- [10] F. G. Cozman
Credal networks.
Artificial Intelligence, **120**, 2 (July 2000), 199–233.
- [11] F. G. Cozman
Graphical models for imprecise probabilities.
International Journal of Approximate Reasoning, **39**, 2–3 (June 2005), 167–184.
- [12] S. Das
High-Level Data Fusion.
Artech House, 2008.
- [13] S. Destercke
A new contextual discounting rule for lower probabilities.
In *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, June–July 2010, 198–207.
- [14] D. R. Insua and F. Ruggeri, (Eds.)
Robust Bayesian Analysis.
Springer, 2000.
- [15] A. Iripino and V. Tontodonato
Cluster reduced interval data using Hausdorff distance.
Computational Statistics, **21**, 2 (2006), 241–288.
- [16] A. Karlsson, R. Johansson, and S. F. Andler
On the behavior of the robust Bayesian combination operator and the significance of discounting.
In *Proceedings of the 6th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*, 2009.
- [17] A. Karlsson, R. Johansson, and S. F. Andler
An empirical comparison of Bayesian and credal set theory for discrete state estimation.
In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, June–July 2010, 80–89.
- [18] A. Karlsson, R. Johansson, and S. F. Andler
An empirical comparison of Bayesian and credal combination operators.
In *Proceedings of The 13th International Conference on Information Fusion*, July 2010.
- [19] I. Levi
The Enterprise of Knowledge.
The MIT Press, 1983.
- [20] M. E. Liggins, D. L. Hall, and J. Llinas (Eds.)
Multisensor Data Fusion, Second Edition.
CRC Press, 2009.
- [21] B. Noack, V. Klumpp, D. Brunn, and U. D. Hanebeck
Nonlinear Bayesian estimation with convex sets of probability densities.
In *Proceedings of the 11th International Conference on Information Fusion*, June–July, 2008, 1–8.
- [22] R Development Core Team
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2009.
- [23] G. Shafer
A Mathematical Theory of Evidence.
Princeton University Press, 1976.
- [24] M. C. Troffaes
Generalizing the conjunction rule for aggregating conflict expert opinions.
International Journal of Intelligent Systems, **21**, 3 (Mar. 2006), 361–380.
- [25] P. Walley
Statistical Reasoning with Imprecise Probabilities.
Chapman and Hall, 1991.

- [26] P. Walley
Towards a unified theory of imprecise probability.
International Journal of Approximate Reasoning, **24**, 2–3
(May 2000), 125–148.

- [27] L. A. Zadeh
Review of books: a mathematical theory of evidence.
AI Magazine, **5**, 3 (1984), 81–83.



Alexander Karlsson holds a Ph.D. in computer science at Örebro University, Sweden, 2010, and a M.Sc. in computer science and engineering at Chalmers University of Technology, Sweden, 2004.

Dr. Karlsson is a researcher at the University of Skövde, Sweden. His main research interest is theory for reasoning in the presence of uncertainty.



Ronnie Johansson received his Ph.D. in computer science from the Royal Institute of Technology (KTH), Sweden, in 2006.

Dr. Johansson is a researcher at the Swedish Defence Research Agency (FOI) in Stockholm. He is also a part-time employee at the University of Skövde, Sweden, where he teaches and conducts research. His research has since 2000 focused on autonomous systems and information fusion. He is currently interested in fusion algorithms, management of uncertainty, knowledge representation and information acquisition. He has served on the programme committees of the Information Fusion Conference and the Multisensor Fusion and Integration for Intelligent Systems Conference. In 2000, he spent six months at the RIKEN institute in Saitama, Japan, while working on multi-robot path planning.

Sten F. Andler received his Ph.D. in computer science in 1979 from Carnegie Mellon University, Pittsburgh, PA, and a Ph.D. in computer science from Chalmers University of Technology, Göteborg, Sweden, also in 1979.

He is a Professor of Computer Science at University of Skövde, Sweden, and Program Director of Infusion, the Research Program in Information Fusion at Skövde. Infusion is funded for 6+2 years by grants from the Swedish Knowledge Foundation and matching grants from industry and the university, totaling SEK 120+15 million (approx USD \$21 million). He has served three years as Dean of Research at the University of Skövde and six years on the Faculty Board at Chalmers University of Technology, Göteborg, Sweden, and is currently serving a second three-year term on the International Society of Information Fusion (ISIF) Board of Directors. He was previously affiliated with the IBM Almaden Research Center and IBM Software Solutions, San Jose, CA, for fourteen years, with brief periods as visiting professor at University of California at Berkeley and Research Intern at Xerox Palo Alto Research Center (PARC). In addition to the Information Fusion Program, now in its seventh year, he heads a research group in Distributed Real-Time Systems (DRTS), with activities in information fusion infrastructures, distributed real-time databases, and model-based software testing.

Dr. Andler's interests are in the areas of information fusion, distributed systems, real-time systems, and databases. He is a Member of the ACM and the IEEE Computer Society. He is a Member of the Editorial Board of *Innovations in Systems and Software Engineering*.

