KU SCHOOL OF BUSINESS WORKING PAPER NO. 330

A New Definition of Entropy of Belief Functions in the Dempster-Shafer Theory*

Radim Jiroušek[1]
radim@utia.cas.cz

Prakash P. Shenoy[2]
pshenoy@ku.edu

[1]Faculty of Management, University of Economics, and
Institute of Information Theory and Automation, Academy of Sciences
Jindřichův Hradec and Prague, Czech Republic

[2]University of Kansas School of Business
1300 Sunnyside Ave., Summerfield Hall
Lawrence, KS 66045-7601 USA

March 8, 2016

*Comments and suggestions for improvement are welcome.

Contents

A New Definition of Entropy of Belief Functions in the Dempster-Shafer Theory

Abstract

We propose a new definition of entropy for basic probability assignments (BPA) in the Dempster-Shafer (D-S) theory of belief functions, which is interpreted as a measure of total uncertainty in the BPA. Our definition is different from the definitions proposed by Höhle, Smets, Yager, Nguyen, Dubois-Prade, Lamata-Moral, Klir-Ramer, Klir-Parviz, Pal *et al.*, Maeda-Ichihashi, Harmanec-Klir, Jousselme *et al.*, and Pouly *et al.* We state a list of five desired properties of entropy for D-S belief functions theory that are motivated by Shannon's definition of entropy for probability functions together with the requirement that any definition should be consistent with the semantics of D-S belief functions theory. Three of our five desired properties are different from the five properties described by Klir and Wierman. We demonstrate that our definition satisfies all five properties in our list, and is consistent with the semantics of D-S theory, whereas none of the existing definitions do. Our new definition has two components. The first component is Shannon's entropy of an equivalent probability mass function obtained using plausibility transformation, which constitutes the conflict measure of entropy. The second component is Dubois-Prade's definition of entropy for basic probability assignments in the D-S theory, which constitutes the non-specificity measure of entropy. Our new definition is the sum of these two components. Our definition does not satisfy the subadditivity property. Whether there exists a definition that satisfies our five properties plus subadditivity, and that is consistent with the semantics for the D-S theory, remains an open question.

1 Introduction

The main goal of this paper is to generalize Shannon's definition of entropy of probability functions [45] to belief functions in the Dempster-Shafer's (D-S) theory [9, 40]. Shannon's definition of entropy of a probability function can be interpreted as the amount of uncertainty in the probability function. Yager [55] and Klir [20] argue that there are several characteristics of uncertainty such as conflict (or confusion or discord or strife or dissonance), and non-specificity (or vagueness or ambiguity or imprecision). In this paper, we are concerned only with entropy of belief functions as a measure of total uncertainty, as commonly understood. For example, suppose X is a discrete random variable with state space Ω_X, and probability mass function (PMF) P_X. Shannon's definition of entropy of P_X, denoted in this paper by $H_s(P_X)$, has the property that $H_s(P_X) \geq 0$, with equality if and only if there exists $x \in \Omega_X$ such that $P_X(x) = 1$. Such a PMF has no uncertainty. For another example, $H_s(P_X)$ achieves its maximum for the equiprobable PMF, which has maximum uncertainty among all PMFs for X. In this case $H_s(P_X) = \log(|\Omega_X|)$. Nevertheless, the equiprobable PMF is still the bearer of some information. If we know that an urn contains the same number of black and white balls, then we know more than in the case where we know that there are only white and black balls in the urn. The latter situation cannot be described in probability theory as well as it can be done in D-S theory, where it is described by the vacuous basic probability assignment (BPA), and therefore, naturally, we expect that the measure of uncertainty for the vacuous BPA should be higher than that for the equiprobable PMF.

We state five basic properties for entropy for the D-S belief functions theory based on our intuitive understanding of Shannon's entropy as a measure of uncertainty. First, if m is a basic probability assignment (BPA) for X with state space Ω_X, then $H(m) \geq 0$, with equality if and only if there exists $x \in \Omega_X$ such that $m(\{x\}) = 1$. We call this the *non-negativity* property. Second, based on our understanding that the vacuous BPA has the most uncertainty, if ι_X denotes the vacuous BPA for X, we require that $H(m) \leq H(\iota_X)$, with equality if and only if $m = \iota_X$. We call this the *maximum entropy* property. Third, $H(\iota_X)$ should be monotonically increasing function of $|\Omega_X|$. We call this the *monotonicity* property. The D-S theory can be considered as a generalization of Bayesian probability theory. A BPA for X in the D-S theory is called *Bayesian* if all its focal elements are singleton subsets of Ω_X. The positive values of such a BPA m can be considered as values of a PMF P_X of X such that $P_X(x) = m(\{x\})$. The fourth property states that if m is Bayesian, then $H(m) = H_s(P_X)$, where $P_X(x) = m(\{x\})$. We call this the *probability consistency* property. The fifth property is based on Dempster's rule of combination. If m_X is a BPA for X, and m_Y is a BPA for Y, and these are distinct, then $H(m_X \oplus m_Y) = H(m_X) + H(m_Y)$. We call this the *additivity* property. Finally, if m is a BPA for $\{X, Y\}$ with marginal BPAs $m^{\downarrow X}$ for X, and $m^{\downarrow Y}$ for Y, then $H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y})$. We call this the *subadditivity* property. This property is satisfied by Shannon's entropy for PMFs. It is difficult to satisfy this property for D-S belief functions. Therefore, we do not require this sixth property. We implicitly require that $H(m)$ exists (*existence*), and that $H(m)$ be a continuous function of m (*continuity*). Also, we are concerned with the D-S theory of belief function with Dempster's rule as the combination rule. Thus, any definition of $H(m)$ should be based on semantics of m that are consistent with the basic tenets of this theory (*consistency with D-S theory semantics*).

One of the earliest definitions of entropy for basic probability assignments (BPAs) in the D-S theory is due to Höhle [18]. Höhle's definition captures only the conflict measure of entropy. Höhle's definition of entropy for vacuous BPA is 0. Thus, it does not satisfy the non-negativity, maximum entropy, and monotonicity properties. It satisfies probability consistency and additivity, but not subadditivity.

Another definition of entropy for BPAs is due to Yager [55]. Like Höhle's definition, Yager's definition only captures the conflict measure of entropy. Like Höhle's, Yager's definition of entropy of the vacuous BPA is 0. Thus, it does not satisfy the non-negativity, maximum entropy, and monotonicity properties. It satisfies the probability consistency and additivity properties, but not the subadditivity property.

Another definition of entropy of BPAs is due to Nguyen [34]. Similar to Höhle's and Yager's definitions, Nguyen's entropy for vacuous BPA is 0, which again violates the non-negativity, maximum entropy, and monotonicity properties. It satisfies the probability consistency and additivity properties, but not the subadditivity property.

Another definition of entropy for belief functions is due to Dubois and Prade [12]. Dubois-Prade's definition captures only the non-specificity measure of uncertainty. It can be considered as a generalization of Hartley's entropy [17] for belief functions. Dubois-Prade's definition of entropy for a Bayesian BPA is 0, which does not satisfy the non-negativity and the probability consistency properties. Dubois-Prade's definition, however, does satisfy the maximum entropy, monotonicity, additivity and subadditivity properties.

Lamata and Moral [27] suggest a definition of entropy of BPA m where they combine Yager's conflict measure of uncertainty with Dubois-Prade's non-specificity measure of uncertainty. Lamata-Moral's definition assigns the same entropy value for the vacuous BPA and the Bayesian BPA with uniform probabilities. Thus, it does not satisfy the maximum entropy property. It does satisfy non-negativity, monotonicity, probability consistency, and additivity, but not subadditivity.

Klir and Ramer [23] propose a definition similar to Lamata-Moral's, but with an improved measure of conflict uncertainty. Klir and Parviz [22] tweak the Klir-Ramer definition a bit to improve the conflict component of entropy. Both of these definitions assign the same entropy for the vacuous BPA and the Bayesian uniform BPA and therefore do not satisfy the maximum entropy property. They do satisfy the non-negativity, monotonicity, probability consistency, and additivity properties, but not the subadditivity property.

Pal *et al.* [36] propose a modification of Nguyen's definition, and it satisfies the non-negativity, monotonicity, probability consistency, and additivity properties. It does not satisfy maximum entropy and subadditivity properties.

A BPA m for X can be regarded as an encoding of a set of probability distributions for X such that the probability of every event is bounded at the bottom by the value of the belief function corresponding to m for the event, and bounded at the top by the value of the plausibility function corresponding to m for the event. However, such a view of BPA m is incompatible with Dempster's rule of combination (see, e.g., [41, 43, 44, 14]).

Maeda and Ichihashi [30] propose a definition of entropy of m which consists of two components. The first component consists of the maximum entropy of the set of probability distributions that are encoded by m. The second component is the generalized Hartley entropy defined by Dubois-Prade [12]. It satisfies all properties including the subadditivity property. However, the interpretation of a BPA m as a set of probability distributions is incompatible with Dempster's combination rule, the main rule underlying the D-S theory of belief functions. This definition may be appropriate for a theory of belief functions with the Fagin-Halpern combination rule [13], but not for the D-S theory.

Harmanec-Klir [15] (independent of [30]) define entropy of m consisting of only the first component of Maeda-Ichihashi's definition. This definition assigns the same entropy value to the vacuous BPA as the Bayesian uniform BPA. Thus, it does not satisfy the maximum entropy property. It does satisfy all other properties, including the subadditivity property. Like Maeda-ichihashi's definition, it is based on semantics of belief functions that are incompatible with Dempster's combination rule.

Jousselme *et al.* [19] propose a definition of entropy of a BPA as the Shannon's entropy of the pignistic PMF transformation of the BPA. Like Maeda-Ichihashi's, this definition assigns the same entropy value to the vacuous BPA and the Bayesian uniform BPA, and therefore does not satisfy the maximum entropy property. It does not satisfy the subadditivity property either [21]. It does satisfy the remaining four properties. However, Cobb and Shenoy [7] have argued that the pignistic transformation of a BPA is incompatible with Dempster's rule of combination, and therefore its Shannon's entropy is not appropriate for the D-S theory of belief functions.

Finally, Pouly *et al.* [37] describe a definition of entropy for hints, which is a formalization of the multivalued semantics for BPA functions. If we adapt this definition to define entropy for BPAs, we would get the same definition as that by Jousselme *et al.* [19], with the same properties.

Our proposal for definition of entropy for D-S theory has two parts. First, given a BPA, we find an equivalent PMF using the plausibility transformation proposed by Cobb and Shenoy [7]. The plausibility transformation has the property that if we combine two BPAs by Dempster's rule and then find the equivalent PMF using the plausibility transformation, we get the same PMF if we find the plausibility transformation for each of the BPAs and then combine the two PMFs using Bayes combination rule (pointwise multiplication followed by normalization) [53]. The first part of our new definition of entropy of a BPA is Shannon's entropy of the equivalent PMF using the plausibility transformation. The second part is Dubois-Prade's definition of entropy of a BPA. Our definition consists of the sum of the two parts. Our definition satisfies all five properties we propose, but it does not satisfy the subadditivity property. It is, however, consistent with semantics for the D-S theory of belief functions. Whether there exists a definition of entropy for BPAs that satisfies the five properties we list, the subadditivity property, and that is consistent with semantics for the D-S theory remains an open question.

An outline of the remainder of the paper is as follows. In Section 2, we briefly review Shannon's definition of entropy for PMFs of discrete random variables, and its properties. In Section 3, we review the basic definitions in the D-S belief functions theory. In Section 4, we propose five properties that an entropy function for BPA should satisfy more formally. We compare our properties with those proposed by Klir and Wierman [24], and also with a monotonicity-like property proposed by

Abellán and Masegosa [3]. Then, we discuss the various definitions that have been proposed in the literature, and how they compare vis-a-vis our list of properties. In Section 5, we propose a new definition of entropy for D-S theory, and show that it satisfies all the properties proposed in Section 4. In Section 6, we discuss some additional properties of our definition. Finally, in Section 7, we summarize our findings, and conclude with some open questions.

2 Shannon's Entropy for PMFs of Discrete Random Variables

In this section we briefly review Shannon's definition of entropy for PMFs of discrete random variables, and its properties. Most of the material in this section is taken from [45, 29].

Information Content Suppose X is a discrete random variable with state space Ω_X and PMF P_X. Consider a state $x \in \Omega_X$ such that $P_X(x) > 0$. What is the information content of this state? Shannon [45] defines the *information* content of state $x \in \Omega_X$ as follows:

$$I(x) = \log_2\left(\frac{1}{P_X(x)}\right). \tag{1}$$

Information content has units of *bits*. Intuitively, the information content of a state is inversely proportional to its probability. If we have a state with probability one, then observing such a state has no information content (0 bits). Conversely, if we observe a rare state, e.g., $x \in \Omega_X$ such that $P_X(x) = \frac{1}{1{,}024}$, then observing such a state is very informative (10 bits). Notice that $I(x) \geq 0$, and $I(x) = 0$ if and only if $P_X(x) = 1$.

Although we have used logarithm to the base 2, we could use any base (e.g., e, or 10), but this will change the units as

$$\log_2(a) \quad = \quad \frac{\log_e(a)}{\log_e(2)} = 1.44\log_e(a).$$

Henceforth, we will simply write log for $\log_2$.

Shannon's Entropy Suppose X is a random variable with PMF P_X. The *entropy* of P_X is the expected information content of the possible states of X:

$$H_s(P_X) = \sum_{x\in\Omega_X} P_X(x)I(x) = \sum_{x\in\Omega_X} P_X(x)\log\left(\frac{1}{P_X(x)}\right) \tag{2}$$

Like information content, entropy is measured in units of bits. One can interpret entropy $H_s(P_X)$ as a measure of uncertainty in the PMF $P_X(x)$. If $P_X(x) = 0$, we will follow the convention that $P_X(x)\log(1/p_X(x)) = 0$ as $\lim_{\theta\to 0^+}\theta\log(1/\theta) = 0$.

Suppose Y is another random variable, and suppose that the joint PMF of X and Y is $P_{X,Y}$ with P_X and P_Y as the marginal PMFs of X and Y, respectively. If we observe $Y = a$ such that $P_Y(a) > 0$, then the posterior PMF of X is $P_{X|a}$ (where $P_{X|a}(x)\cdot P_Y(a) = P_{X,Y}(x,a)$), and the respective posterior entropy is $H_s(P_{X|a})$.

From our viewpoint, the following properties of Shannon's entropy function for PMFs are the most important:

1. $H_s(P_X) \geq 0$, with equality if and only if there exists $x \in \Omega_X$ such that $P_X(x) = 1$.

2. $H_s(P_X) \leq \log(|\Omega_X|)$, with equality if and only if $P_X(x) = \frac{1}{|\Omega_X|}$ for all $x \in \Omega_X$. $|\Omega_X|$ denotes the cardinality (i.e., number of elements) of set Ω_X.

3. The entropy of P_X does not depend on the labels attached to the states of X, only on their probabilities. This is in contrast with, e.g., variance of X, which is defined only for real-valued random variables. Thus, for a real-valued discrete random variable X, and $Y = 10\,X$, it is obvious that $H_s(P_Y) = H_s(P_X)$, whereas $V(P_Y) = 100\,V(P_X)$.

4. Shannon [45] derives the expression for entropy of X axiomatically using four axioms as follows.

   (a) Axiom 0 (*Existence*): $H(X)$ exists.

   (b) Axiom 1 (*Continuity*): $H(X)$ should be a continuous function of $P_X(x)$ for $x \in \Omega_X$.

   (c) Axiom 2 (*Monotonicity*): If we have an equally likely PMF, then $H(X)$ should be a monotonically increasing function of $|\Omega_X|$.

   (d) Axiom 3 (*Compound distributions*): If a PMF is factored into two PMFs, then its entropy should be the sum of entropies of its factors, e.g., $P_{X,Y}(x,y) = P_X(x)\,P_{Y|x}(y)$, then $H(P_{X,Y}) = H(P_X) + \sum_{x\in\Omega_X} P_X(x)\,H(P_{Y|x})$.

   Shannon [45] proves that the only function H_s that satisfies Axioms 0–3 is of the form $H_s(P_X) =$
   $K \sum_{x\in\Omega_X} P_X(x) \log\left(\frac{1}{P_X(x)}\right)$, where K is a constant depending on the choice of units of measurement.

Suppose X and Y are discrete random variables with joint PMF $P_{X,Y}$. In analogy to the one-dimensional case, the *joint entropy* of $P_{X,Y}$ is:

$$H_s(P_{X,Y}) = \sum_{x\in\Omega_X} \sum_{y\in\Omega_Y} P_{X,Y}(x,y) \log\left(\frac{1}{P_{X,Y}(x,y)}\right) \tag{3}$$

Let $P_{Y|X} : \Omega_{\{X,Y\}} \to [0,1]$ be a function such that $P_{Y|X}(x,y) = P_{Y|x}(y)$ for all $(x,y) \in \Omega_{\{X,Y\}}$. $P_{Y|X}$ is not a conditional PMF, but can be considered as a collection of conditional PMFs, one for each $x \in \Omega_X$. If we combine P_X and $P_{Y|X}$ using pointwise multiplication followed by normalization, an operation that we denote by $\otimes$, then we obtain $P_{X,Y}$, i.e., $P_{X,Y} = P_X \otimes P_{Y|X}$, i.e., $P_{X,Y}(x,y) = P_X(x)\,P_{Y|X}(x,y) = P_X(x)\,P_{Y|x}(y)$ for all $(x,y) \in \Omega_{\{X,Y\}}$. As P_X and $P_{Y|x}$ are PMFs, there is no need for normalization (or the normalization constant is 1).

Shannon defined entropy of $P_{Y|X}$ as follows:

$$H_s(P_{Y|X}) = \sum_{x\in\Omega_X} P_X(x)\,H_s(P_{Y|x}) \tag{4}$$

We call $H_s(P_{Y|X})$ the *conditional* entropy of Y given X.

It follows from Axiom 3 that

$$H_s(P_{X,Y}) = H_s(P_X \otimes P_{Y|X}) = H_s(P_X) + H_s(P_{Y|X}). \tag{5}$$

If we call $H_s(P_X)$ as *marginal* entropy of X, Eq. (5) is the compound distribution axiom underlying Shannon's entropy expressed in terms of marginal and conditional entropies. Eq. (5) is also called the *chain rule* of entropy.

If X and Y are independent with respect to $P_{X,Y}$, i.e., $P_{Y|x}(y) = P_Y(y)$ for all $(x, y) \in \Omega_{\{X,Y\}}$ such that $P_X(x) > 0$, then it follows from Eq. (4) that $H_s(P_{Y|X}) = H_s(P_Y)$. Thus, if X and Y are independent with respect to $P_{X,Y}$, then $H_s(P_{X,Y}) = H_s(P_X) + H_s(P_Y)$.

If X and Y are not independent with respect to joint PMF $P_{X,Y}$, then it can be shown that $H_s(P_{X|Y}) < H_s(P_X)$. This leads to the definition of *mutual information* between X and Y, denoted by $I(X;Y)$, defined as follows:

$$I(X;Y) = H_s(P_X) - H_s(P_{X|Y}) \tag{6}$$

Eq. (6) defining mutual information is the central concept in information theory. $I(X;Y) \geq 0$, with equality if and only if X and Y are independent (with respect to $P_{X,Y}$). From the chain rule of entropy, $H_s(P_{X,Y}) = H_s(P_X) + H_s(P_{Y|X}) = H_s(P_Y) + H_s(P_{X|Y})$. Rearranging terms, we have $H_s(P_X) - H_s(P_{X|Y}) = H_s(P_Y) - H_s(P_{Y|X})$, i.e., $I(X;Y) = I(Y;X)$. $I(X;Y)$ is a measure of total dependence between X and Y (with respect to $P_{X,Y}$).

$I(X;Y)$ is related to the concept of Kullback-Leibler divergence between two PMFs [26]. Suppose P_1 and P_2 are two PMFs for X. We assume that $P_2(x) > 0$ for all $x \in \Omega_X$. The *Kullback-Leibler divergence* between P_1 and P_2, denoted by $D(P_1, P_2)$, is defined as follows:

$$D(P_1, P_2) = \sum_{x \in \Omega_X} P_1(x) \log\left(\frac{P_1(x)}{P_2(x)}\right) \tag{7}$$

$D(P_1, P_2)$ has the following properties. $D(P_1, P_2) \geq 0$, with equality if and only if $P_1 = P_2$. Also, $D(P_1, P_2)$ is not symmetric, i.e., $D(P_1, P_2)$ may not be equal to $D(P_2, P_1)$. It can be easily shown that $I(X, Y) = D(P_{X,Y}, P_X \otimes P_Y)$.

Suppose P_X and P_Y are the marginal PMFs obtained from the joint PMF $P_{X,Y}$. As discussed above, it follows from the chain rule of entropy and the inequality $H_s(P_{Y|X}) \leq H_s(P_Y)$, that

$$H_s(P_{X,Y}) \leq H_s(P_X) + H_s(P_Y), \tag{8}$$

with equality if and only if X and Y are independent with respect to $P_{X,Y}$. The inequality in Eq. (8) is called *subadditivity* in the literature, see e.g., [12].

3 Basic Definitions of the D-S Belief Functions Theory

In this section we review the basic definitions in the D-S belief functions theory. Like the various uncertainty theories, D-S belief functions theory includes functional representations of uncertain knowledge, and operations for making inferences from such knowledge.

Basic Probability Assignment Suppose X is a random variable with state space Ω_X. Let 2^{Ω_X} denote the set of all *non-empty* subsets of Ω_X. A basic probability assignment (BPA) m for X is a function $m : 2^{\Omega_X} \to [0, 1]$ such that

$$\sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) = 1. \tag{9}$$

The non-empty subsets $\mathsf{a} \in 2^{\Omega_X}$ such that $m(\mathsf{a}) > 0$ are called *focal* elements of m. An example of a BPA for X is the vacuous BPA for X, denoted by ι_X, such that $\iota_X(\Omega_X) = 1$. We say m is *deterministic* if m has a single focal element (with probability 1). Thus, the vacuous BPA for X is deterministic with focal element Ω_X. If all focal elements of m are singleton subsets of Ω_X, then we say m is *Bayesian*. In this case, m is equivalent to the PMF P for X such that $P(x) = m(\{x\})$ for each $x \in \Omega_X$. Let m_u denote the Bayesian BPA with uniform probabilities, i.e., $m_u(\{x\}) = \frac{1}{|\Omega_X|}$ for all $x \in \Omega_X$. If Ω_X is a focal element, then we say m is *non-dogmatic*, and *dogmatic* otherwise. Thus, a Bayesian BPA is dogmatic.

Plausibility Function The information in a BPA m can be represented by a corresponding plausibility function Pl_m that is defined as follows.

$$Pl_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} :\, \mathsf{b} \cap \mathsf{a} \neq \emptyset} m(\mathsf{b}) \tag{10}$$

for all $\mathsf{a} \in 2^{\Omega_X}$. For an example, suppose $\Omega_X = \{x, \bar{x}\}$. Then, the plausibility function Pl_{ι_X} corresponding to BPA ι_X is given by $Pl_{\iota_X}(\{x\}) = 1$, $Pl_{\iota_X}(\{\bar{x}\}) = 1$, and $Pl_{\iota_X}(\Omega_X) = 1$.

Belief Function The information in a BPA m can also be represented by a corresponding belief function Bel_m that is defined as follows.

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} :\, \mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}) \tag{11}$$

for all $\mathsf{a} \in 2^{\Omega_X}$. For the example above with $\Omega_X = \{x, \bar{x}\}$, the belief function Bel_{ι_X} corresponding to BPA ι_X is given by $Bel_{\iota_X}(\{x\}) = 0$, $Bel_{\iota_X}(\{\bar{x}\}) = 0$, and $Bel_{\iota_X}(\Omega_X) = 1$.

Commonality Function The information in a BPA m can also be represented by a corresponding commonality function Q_m that is defined as follows.

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_X} :\, \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}) \tag{12}$$

for all $\mathsf{a} \in 2^{\Omega_X}$. For the example above with $\Omega_X = \{x, \bar{x}\}$, the commonality function Q_{ι_X} corresponding to BPA ι_X is given by $Q_{\iota_X}(\{x\}) = 1$, $Q_{\iota_X}(\{\bar{x}\}) = 1$, and $Q_{\iota_X}(\Omega_X) = 1$. If m is non-dogmatic, then $Q_m(\mathsf{a}) > 0$ for all $\mathsf{a} \in 2^{\Omega_X}$. Notice also that for singleton subsets $\mathsf{a} \in 2^{\Omega_X}$ ($|\mathsf{a}| = 1$), $Q_m(\mathsf{a}) = Pl_m(\mathsf{a})$. This is because for singleton subsets a, the set of all subsets that have non-empty intersection with a coincide with the set of all supersets of a.

All four representations—BPA, belief, plausibility, and commonality—have exactly the same information. Given any one, we can transform it to another [40]. However, they have different semantics. Next, we describe the two main operations for making inferences.

Dempster's Rule of Combination In the D-S theory, we can combine two BPAs m_1 and m_2 representing distinct pieces of evidence by Dempster's rule [9] and obtain the BPA $m_1 \oplus m_2$, which represents the combined evidence. Dempster referred to this rule as the product-intersection rule, as the product of the BPA values are assigned to the intersection of the focal elements, followed by normalization. Normalization consists of discarding the probability assigned to $\emptyset$, and normalizing the remaining values so that they add to 1. In general, Dempster's rule of combination can be used to combine two BPAs for arbitrary sets of variables. In this paper, we are interested only in two special situations: combination of BPAs defined for the same variable, and combination of BPAs defined for disjoint sets of variables.

Suppose m_1 and m_2 are two BPAs for X. In this case, formally,

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{b}_1,\, \mathsf{b}_2 \in 2^{\Omega_X} :\, \mathsf{b}_1 \cap \mathsf{b}_2 = \mathsf{a}} m_1(\mathsf{b}_1)\, m_2(\mathsf{b}_2), \tag{13}$$

for all $\mathsf{a} \in 2^{\Omega_X}$, where K is a normalization constant given by

$$K = 1 - \sum_{\mathsf{b}_1,\, \mathsf{b}_2 \in 2^{\Omega_X} :\, \mathsf{b}_1 \cap \mathsf{b}_2 = \emptyset} m_1(\mathsf{b}_1)\, m_2(\mathsf{b}_2). \tag{14}$$

The definition of Dempster's rule assumes that the normalization constant K is non-zero. If $K = 0$, then the two BPAs m_1 and m_2 are said to be in *total conflict* and cannot be combined. If $K = 1$, we say m_1 and m_2 are *non-conflicting*.

Dempster's rule can also be described in terms of commonality functions [40]. Suppose Q_{m_1} and Q_{m_2} are commonality functions corresponding to BPAs m_1 and m_2, respectively. The commonality function corresponding to BPA $Q_{m_1 \oplus m_2}$ is as follows:

$$Q_{m_1 \oplus m_2}(\mathsf{a}) = K^{-1} Q_{m_1}(\mathsf{a})\, Q_{m_2}(\mathsf{a}), \tag{15}$$

for all $\mathsf{a} \in 2^{\Omega_X}$, where the normalization constant K is exactly the same as in Eq. (14). In terms of commonality functions, Dempster's rule is pointwise multiplication of commonality functions followed by normalization.

Suppose that m_X and m_Y are two BPAs for X and Y, respectively. In this case, $m_X \oplus m_Y$ is a BPA for $\{X, Y\}$. To define this two-dimensional BPA formally we have to first define projection of states, and then projection of subsets of states.

Projection of states simply means dropping extra coordinates; for example, if (x, y) is a state of $\{X, Y\}$, then the projection of (x, y) to X, denoted by $(x, y)^{\downarrow X}$, is simply x, which is a state of X.

Projection of subsets of states is achieved by projecting every state in the subset. Suppose $\mathsf{b} \in 2^{\Omega_{\{X,Y\}}}$. Then $\mathsf{b}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{b}\}$. Notice that $\mathsf{b}^{\downarrow X} \in 2^{\Omega_X}$.

The combination of BPAs m_X and m_Y for X and Y, respectively, simplifies to the formula

$$(m_X \oplus m_Y)(\mathsf{c}) = m_X(\mathsf{c}^{\downarrow X})\, m_Y(\mathsf{c}^{\downarrow Y}), \tag{16}$$

for all $\mathsf{c} \in 2^{\Omega_{\{X,Y\}}}$. Notice that in this case there is no need for normalization as there is no mass on the empty set, i.e., m_X and m_Y are always non-conflicting.

Marginalization Marginalization in D-S theory is addition of values of BPAs. Suppose m is a BPA for $\{X, Y\}$. Then, the marginal of m for X, denoted by $m^{\downarrow X}$, is a BPA for X such that for each $\mathsf{a} \in 2^{\Omega_X}$,

$$m^{\downarrow X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_{\{X,Y\}}} :\, \mathsf{b}^{\downarrow X} = \mathsf{a}} m(\mathsf{b}). \tag{17}$$

Conditional belief functions In probability theory, it is common to construct joint probability mass functions for a set of discrete variables by using conditional probability distributions. For example, $P_{X,Y} = P_X \otimes P_{Y|X}$. We can construct joint BPA for $\{X, Y\}$ in a similar manner.

Consider a BPA m_X for X with corresponding plausibility function Pl_{m_X}. Let $m_{Y|x}$ denote a conditional BPA for Y given that $X = x$ where $x \in \Omega_X$ is such that $Pl_{m_X}(\{x\}) > 0$, i.e., $m_{Y|x} : 2^{\Omega_Y} \to [0, 1]$ such that $\sum_{\mathsf{b} \in 2^{\Omega_Y}} m_{Y|x}(\mathsf{b}) = 1$. $m_{Y|x}$ represents our belief about Y if we know that $X = x$. We can embed this conditional BPA for Y into a (unconditional) BPA for

$\{X, Y\}$, say $m_{x,Y}$, such that the following two conditions hold. First, $m_{x,Y}$ tells us nothing about X, i.e., $m_{x,Y}^{\downarrow X} = \iota_X$. Second, if we combine $m_{x,Y}$ with the deterministic BPA $m_{X=x}$ for X such that $m_{X=x}(\{x\}) = 1$ using Dempster's rule, and marginalize the result to Y we obtain $m_{Y|x}$, i.e., $(m_{x,Y} \oplus m_{X=x})^{\downarrow Y} = m_{Y|x}$. One way to obtain such an embedding is suggested by Smets [46, 42], called *conditional embedding*, and it consists of taking each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of $m_{Y|x}$, and converting it to a corresponding focal element of $m_{x,Y}$ (with the same mass) as follows: $(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y)$. It is easy to confirm that this method of embedding satisfies both conditions mentioned above.

Example 1. *Consider discrete variables X and Y, with $\Omega_X = \{x, \bar{x}\}$ and $\Omega_Y = \{y, \bar{y}\}$. If we have a conditional BPA $m_{Y|x}$ for Y given $X = x$ as follows: $m_{Y|x}(y) = 0.8$, and $m_{Y|x}(\Omega_Y) = 0.2$, then its conditional embedding into BPA $m_{x,Y}$ for $\{X, Y\}$ is as follows: $m_{x,Y}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) = 0.8$, and $m_{x,Y}(\Omega_{\{X,Y\}}) = 0.2$. Similarly, if we have a conditional BPA $m_{Y|\bar{x}}$ for Y given $X = \bar{x}$ as follows: $m_{Y|\bar{x}}(\bar{y}) = 0.9$, and $m_{Y|\bar{x}}(\Omega_Y) = 0.1$, then its conditional embedding into BPA $m_{\bar{x},Y}$ for $\{X, Y\}$ is as follows: $m_{\bar{x},Y}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.9$, $m_{\bar{x},Y}(\Omega_{\{X,Y\}}) = 0.1$. Assuming we have these two conditional BPAs, and their corresponding embeddings, it is clear that the two BPA $m_{x,Y}$ and $m_{\bar{x},Y}$ are distinct, and can be combined with Dempster's rule of combination, resulting in the BPA $m_{Y|X} = m_{x,Y} \oplus m_{\bar{x},Y}$ for $\{X, Y\}$ as follows: $m_{Y|X}(\{(x, y), (\bar{x}, \bar{y})\}) = 0.72$, $m_{Y|X}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.08$, $m_{Y|X}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) = 0.18$, and $m_{Y|X}(\Omega_{\{X,Y\}}) = 0.02$. $m_{Y|X}$ has the following properties. First, $m_{Y|X}^{\downarrow Y} = \iota_Y$. Second, if we combine $m_{Y|X}$ with deterministic BPA $m_{X=x}(\{x\}) = 1$ for X, and marginalize the combination to Y, then we get $m_{Y|x}$, i.e., $(m_{Y|X} \oplus m_{X=x})^{\downarrow Y} = m_{Y|x}$. Third, $(m_{Y|X} \oplus m_{X=\bar{x}})^{\downarrow Y} = m_{Y|\bar{x}}$.*

There are some differences with conditional probability distributions. First, in probability theory, $P_{Y|X}$ consists of *all* conditional distributions $P_{Y|x}$ that are well-defined, i.e., for all $x \in \Omega_X$ such that $P_X(x) > 0$. In D-S belief function theory, there are no such constraints. We can include only those conditionals that we have (non-vacuous) knowledge of. Also, if we have more than one conditional BPA, say for $X = x_1$, and $X = x_2$, we embed these conditionals into unconditional BPAs and combine them using Dempster's rule of combination to obtain a (unconditional) BPA $m_{Y|X}$. Second, given any joint PMF $P_{X,Y}$ for $\{X, Y\}$, we can always factor this into P_X for X, and $P_{Y|X}$ for $\{X, Y\}$, such that $P_{X,Y} = P_X \otimes P_{Y|X}$. This is not true in D-S belief function theory. Given a joint BPA $m_{X,Y}$ for $\{X, Y\}$, we cannot always find a belief function $m_{Y|X}$ for $\{X, Y\}$ such that $m_{X,Y} = m_{X,Y}^{\downarrow X} \oplus m_{Y|X}$. However, we can always *construct* joint BPA $m_{X,Y}$ for $\{X, Y\}$ by first assessing m_X for X, and assessing conditionals $m_{Y|x_i}$ for those x_i that we have knowledge about, embed these conditionals into unconditional BPAs, and combine all such BPAs to obtain the BPA $m_{Y|X}$ for $\{X, Y\}$. We can then construct $m_{X,Y} = m_X \oplus m_{Y|X}$.

This completes our brief review of the D-S belief function theory. For further details, the reader is referred to [40].

4 Entropy of BPAs in the D-S Theory

First, we propose five properties that an entropy function for BPAs in the D-S theory should satisfy. Next, we compare these properties with those proposed by Klir and Wierman [24] for the same purposes. Finally, we discuss some definitions that have been proposed in the literature, and how they compare vis-a-vis the list of properties.

First, we are concerned in this paper only with the D-S belief functions theory that includes Dempster's rule of combination as the operation for aggregating uncertain knowledge. There are theories of belief functions that use other combination rules. For example, a BPA m for X can be considered as an encoding of a collection of PMFs $\mathcal{P}_m$ for X such that for all $\mathsf{a} \in 2^{\Omega_X}$ we have

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}) = \min_{P \in \mathcal{P}_m} \sum_{x \in \mathsf{a}} P(x). \tag{18}$$

$\mathcal{P}_m$ is referred to as a *credal* set corresponding to m in the imprecise probability literature (see, e.g. [54]). For such a theory of belief functions, a combination rule is proposed by Fagin and Halpern [13] that is different from Dempster's combination rule. Thus, a BPA m in the D-S theory cannot be interpreted as a collection of PMFs satisfying Eq. (18) [41, 14]. There are, of course, semantics that are consistent with D-S theory, such as multivalued mappings [9], random codes [41], transferable beliefs [50], and hints [25].

Second, given a BPA m for X in the D-S theory, there are many ways to transform m to a corresponding PMF P_m for X [8]. However, only one of these ways, called *plausibility transform*, is consistent with m in the D-S theory in the sense that $P_{m_1} \otimes P_{m_2} = P_{m_1 \oplus m_2}$, where $\otimes$ is the combination rule in probability theory, and $\oplus$ is Dempster's combination rule in D-S theory [7]. Thus, any method for defining entropy of m in the D-S theory by first transforming m to a corresponding PMF should use the plausibility transformation method.

Thus, one requirement we implicitly assume is that any definition of entropy of m should be based on semantics for m that are consistent with the basis tenets of D-S theory. Also, we implicitly assume existence and continuity—given a BPA m, $H(m)$ should always exist, and $H(m)$ should be a continuous function of m. We do not list these three requirements explicitly.

Desired Properties of Entropy for D-S Theory The following list of desired properties of entropy $H(m)$, where m is a BPA, are motivated by the properties of Shannon's entropy for PMFs described in Section 2.

Let X and Y denote random variables with state spaces Ω_X and Ω_Y, respectively. Let m_X and m_Y denote distinct BPAs for X and Y, respectively. Let ι_X and ι_Y denote the vacuous BPAs for X and Y, respectively.

1. (*Non-negativity*) $H(m_X) \geq 0$, with equality if and only if there is a $x \in \Omega_X$ such that $m_X(\{x\}) = 1$. This is similar to the probabilistic case.

2. (*Maximum entropy*) $H(m_X) \leq H(\iota_X)$, with equality if and only if $m_X = \iota_X$. This makes sense as the vacuous BPA ι_X for X has the most uncertainty among all BPAs for X. Such a property is advocated in [4].

3. (*Monotonicity*) If $|\Omega_X| < |\Omega_Y|$, then $H(\iota_X) < H(\iota_Y)$. This is similar to Axiom 2 of Shannon.

4. (*Probability consistency*) If m_X is a Bayesian BPA for X, then $H(m_X) = H_s(P_X)$, where P_X is the PMF of X corresponding to m_X, i.e., $P_X(x) = m_X(\{x\})$ for all $x \in \Omega_X$, and $H_s(P_X)$ is Shannon's entropy of PMF P_X. In other words, if m_X is a Bayesian BPA for X, then $H(m_X) = \sum_{x \in \Omega_X} m_X(\{x\}) \log\left(\frac{1}{m_X(\{x\})}\right)$.

5. (*Additivity*) As m_X and m_Y are distinct, we can combine them using Dempster's rule yielding BPA $m_X \oplus m_Y$ for $\{X, Y\}$. Then,

$$H(m_X \oplus m_Y) = H(m_X) + H(m_Y). \tag{19}$$

   This is similar to the compound axiom for Shannon's entropy of a PMF in case of independent random variables.

Klir and Wierman [24] also describe a set of properties that they believe should be satisfied by any meaningful measure of uncertainty based on intuitive grounds. Some of the properties that they suggest are also included in the above list. For example, probability consistency and additivity appear in both sets of requirements. Two of the properties that they require do not make intuitive sense to us.

First, Klir and Weirman suggest a property that they call "set consistency" as follows: $H(m) = \log(|\mathsf{a}|)$ whenever m is deterministic with focal set a. This property would require that $H(\iota_X) = \log(|\Omega_X|)$. The probability consistency property would require that for the Bayesian uniform BPA m_u, $H(m_u) = \log(|\Omega_X|)$. Thus, these two requirements would entail that $H(\iota_X) = H(m_u) = \log(|\Omega_X|)$. We disagree. Consider the following experiment. Suppose we have an urn with 2 balls. Each of the two balls could be either black (b) or white (w), i.e., they can be both black, or both white, or one of each color. One ball is drawn at random from the urn. Let X denote the color of the ball drawn. For this experiment, $\Omega_X = \{b, w\}$. If we have no further information of this experiment beyond what is already stated, we can represent our uncertainty of X by the vacuous BPA ι_X for X. The set consistency requirement would require that $H(\iota_X) = \log(2) = 1$. If we also know that we have exactly one black ball and exactly one white ball in the urn, then we have more information than we had earlier, and consequently, less uncertainty. Given this additional information, we can represent our uncertainty about X by a Bayesian uniform BPA m_u, $m_u(\{b\}) = m_u(\{w\}) = \frac{1}{2}$. The Shannon entropy of m_u is $\log(2) = 1$. Thus, the set consistency property together with the probability consistency property would imply that ι_X and Bayesian BPA m_u have the same uncertainty! According to our requirements, $H(\iota_X) > H(m_u)$, which make more intuitive sense than requiring $H(\iota_X) = H(m_u)$.

Second, Klir and Wierman require a property they call "range" as follows: For any BPA m_X for X, $0 \leq H(m_X) \leq \log(|\Omega_X|)$. The probability consistency property requires that $H(m_u) = \log(|\Omega_X|)$. Also including the range property prevents us, e.g., from having $H(\iota_X) > H(m_u)$. So we do not include it in our list as it violates our intuition.

Finally, Klir and Wierman require the subadditivity property defined as follows. Suppose m is a BPA for $\{X, Y\}$, with marginal BPAs $m^{\downarrow X}$ for X, and $m^{\downarrow Y}$ for Y. Then,

$$H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y}) \tag{20}$$

We agree that this property is important, and the only reason we do not include it in our list is because we are unable to meet this requirement in addition to the five requirements that we do include, and our implicit requirement that any definition be consistent with the semantics of D-S theory of belief functions.

Abellán and Moral [4] interpret a BPA m as a credal set of PMFs as in Eq. (18). With this interpretation, they propose a monotonicity-like property as follows:

- (*Abellán-Moral's monotonicity-like property*) If m_1 and m_2 are BPA functions for X with credal sets $\mathcal{P}_{m_1}$ and $\mathcal{P}_{m_2}$, respectively, such that $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$, then $H(m_1) \leq H(m_2)$.

We can construct such a pair of BPAs by enlarging a focal set of m_1 while keeping everything else the same [3]. For example, if m_1 is a BPA for X with $\Omega_X = \{x_1, x_2, x_3\}$ such that $m_1(x_1, x_2) = \frac{1}{3}$, $m_1(x_1, x_3) = \frac{1}{2}$, and $m_1(x_2, x_3) = \frac{1}{6}$, then consider BPA m_2 such that the first focal element $\{x_1, x_2\}$ is enlarged to $\{x_1, x_2, x_3\}$, i.e., $m_2(x_1, x_2, x_3) = \frac{1}{3}$, $m_2(x_1, x_3) = \frac{1}{2}$, and $m_2(x_2, x_3) = \frac{1}{6}$. It is easy to confirm that $Bel_{m_1}(\mathsf{a}) \geq Bel_{m_2}(\mathsf{a})$ for all $\mathsf{a} \in 2^{\Omega_X}$. Consequently, we have the following relations between the credal sets of the two BPAs: $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$. Assuming that the credal set semantics of a BPA function are appropriate, it is reasonable to adopt such a monotonicity-like property. However, Dempster's combination rule is not consistent with the credal set semantics [41, 43, 44, 14]. If our current knowledge of X is represented by BPA m_1, and we obtain an piece of evidence represented by BPA m_2 for X that is distinct from m_1, then in the D-S theory, our new knowledge is now represented by $m_1 \oplus m_2$. In general, it is not possible to formulate any relationship between $\mathcal{P}_{m_1}$ and $\mathcal{P}_{m_1 \oplus m_2}$. For this reason, we do not adopt Abellán-Moral's monotonicity-like property. Abellán et al. [2] and Bronevich and Klir [6] describe some measures of uncertainty in credal sets.

The most important axiom that characterizes Shannon's definition of entropy is the compound axiom $H(P_{X,Y}) = H(P_X \otimes P_{Y|X}) = H(P_X) + H(P_{Y|X})$. This translated to the D-S theory of belief function would require factorizing a BPA m for $\{X, Y\}$ into BPA $m^{\downarrow X}$ for X, and $m_{Y|X}$ for $\{X, Y\}$ such that $m = m^{\downarrow X} \oplus m_{Y|X}$. This cannot be done for all BPA m for $\{X, Y\}$. But, as we discussed in Section 3, we could construct m for $\{X, Y\}$ such that $m = m_X \oplus m_{Y|X}$, where m_X is a BPA for X, and $m_{Y|X}$ is a BPA for $\{X, Y\}$ such that $m_{Y|X}^{\downarrow X} = \iota_X$, and m_X and $m_{Y|X}$ are non-conflicting. Notice that such a constructed BPA m would have the property $m^{\downarrow X} = (m_X \oplus m_{Y|X})^{\downarrow X} = m_X$. For such constructive BPAs m, we could require a *compound* property as follows:

$$H(m) = H(m_X \oplus m_{Y|X}) = H(m_X) + H(m_{Y|X}). \tag{21}$$

However, we are unable to formulate a definition of $H(m)$ to satisfy such a compound property. So like the subadditivity property, we do not include a compound property in our list of properties. The additivity property included in Klir-Wierman's and our list is so weak that it is satisfied by any definition on a log scale. All definitions of entropy for belief functions in the literature are defined on a log scale, and, thus, they all satisfy the additivity property.

Literature Review One of the earliest definitions of entropy for D-S theory is due to Höhle [18], who defines entropy of BPA m as follows. Suppose m is a BPA for X with state space Ω_X.

$$H_o(m) = \sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) \log\left(\frac{1}{Bel_m(\mathsf{a})}\right), \tag{22}$$

where Bel_m denotes the belief function corresponding to m as defined in Eq. (11). $H_o(m)$ captures only the conflict measure of uncertainty. $H_o(\iota_X) = 0$. Thus, H_o does not satisfy non-negativity, maximum entropy, and monotonicity properties. For Bayesian BPA, $m(\{x\}) = Bel_m(\{x\})$, and therefore, H_o does satisfy the probability consistency property. It satisfies the additivity property but not the subadditivity property [12].

Smets [47] defines entropy of BPA m as follows. Suppose m is a non-dogmatic BPA for X, i.e., $m(\Omega_X) > 0$. Let Q_m denote the commonality function corresponding to BPA m. As m is non-dogmatic, it follows that $Q_m(\mathsf{a}) > 0$ for all $\mathsf{a} \in 2^{\Omega_X}$. The entropy of m is as follows:

$$H_t(m) = \sum_{\mathsf{a} \in 2^{\Omega_X}} \log\left(\frac{1}{Q_m(\mathsf{a})}\right) \tag{23}$$

Smets' definition $H_t(m)$ is designed to measure "information content" of m, rather than uncertainty. Like Höhle's definition, $H_t(\iota_X) = 0$, and therefore, H_t does not satisfy the non-negativity, maximum entropy, and monotonicity properties. As a Bayesian BPA is not non-dogmatic, the probabilistic consistency property is not satisfied either. If m_1 and m_2 are two non-conflicting (i.e., normalization constant in Dempster's combination rule $K = 1$) and non-dogmatic BPAs, then $H_t(m_1 \oplus m_2) = H_t(m_1) + H_t(m_2)$. Thus, it satisfies the additivity property for the restricted class of non-dogmatic BPAs. It does not satisfy the subadditivity property [12].

Another definition of entropy of BPA m is due to Yager [55]:

$$H_y(m) = \sum_{\mathsf{a}\in 2^{\Omega_X}} m(\mathsf{a}) \log\left(\frac{1}{Pl_m(\mathsf{a})}\right), \tag{24}$$

where Pl_m is the plausibility function corresponding to m as defined in Eq. 10. Yager's definition $H_y(m)$ measures only conflict in m, not total uncertainty. Like Höhle's and Smets' definitions, $H_y(\iota_X) = 0$, and therefore, H_y does not satisfy the non-negativity, maximum entropy, and monotonicity properties. It does satisfy the probability consistency property because for Bayesian BPA, $Pl_m(\{x\}) = m(\{x\})$. It satisfies the additivity property, but not the subadditivity property [12].

Nguyen [34] defines entropy of BPA m for X as follows:

$$H_n(m) = \sum_{\mathsf{a}\in 2^{\Omega_X}} m(\mathsf{a}) \log\left(\frac{1}{m(\mathsf{a})}\right) \tag{25}$$

The same definition is stated in [32]. Like all previous definitions, it captures only the conflict portion of uncertainty. As in the previous definitions, $H_n(\iota_X) = 0$. Thus, H_n does not satisfy the non-negativity, maximum entropy, and monotonicity properties. However, as it immediately follows from the properties of Shannon's entropy, it does satisfy the probabilistic consistency property. The fact that it also satisfies the additivity property follows from the fact that log of a product is the sum of the logs. Thus, $H(m_X \oplus m_Y) = \sum_{\mathsf{a}\in 2^{\Omega_{\{X,Y\}}}} m_X(\mathsf{a}^{\downarrow X})\, m_Y(\mathsf{a}^{\downarrow Y}) \log\left(\frac{1}{m_X(\mathsf{a}^{\downarrow X})\, m_Y(\mathsf{a}^{\downarrow Y})}\right) = \left(\sum_{\mathsf{a}^{\downarrow X}\in 2^{\Omega_X}} m_X(\mathsf{a}^{\downarrow X}) \log\left(\frac{1}{m_X(\mathsf{a}^{\downarrow X})}\right)\right) + \left(\sum_{\mathsf{a}^{\downarrow Y}\in 2^{\Omega_Y}} m_Y(\mathsf{a}^{\downarrow Y}) \log\left(\frac{1}{m_Y(\mathsf{a}^{\downarrow Y})}\right)\right) = H(m^{\downarrow X}) + H(m^{\downarrow Y})$. It does not satisfy the subadditivity property as can be seen from Example 2.

Example 2. *Consider BPA m for $\{X, Y\}$ as follows: $m(\{(x, y), (\bar{x}, \bar{y})\}) = m(\{(x, \bar{y}), (\bar{x}, y)\}) = \frac{1}{2}$. For this BPA, $H_n(m) = 1$. Also, $m^{\downarrow X} = \iota_X$, and $m^{\downarrow Y} = \iota_Y$. Therefore, $H_n(m^{\downarrow X}) = 0$, and $H_n(m^{\downarrow Y}) = 0$. Thus, subadditivity is not satisfied.*

Dubois and Prade [12] define entropy of BPA m for X as follows:

$$H_d(m) = \sum_{\mathsf{a}\in 2^{\Omega_X}} m(\mathsf{a}) \log(|\mathsf{a}|). \tag{26}$$

Recall that $|\mathsf{a}|$ denotes the cardinality of a. Dubois-Prade's definition captures only the non-specificity portion of uncertainty. If X is a random variable with state space Ω_X, Hartley [17] defines a measure of entropy of X as $\log(|\Omega_X|)$. Dubois-Prade's definition $H_d(m)$ can be regarded as the mean Hartley entropy of m. If ι_X denotes the vacuous BPA for X, then $H_d(\iota_X) = \log(|\Omega_X|)$. If m is a Bayesian BPA, then $H_d(m) = 0$ as all the focal elements of m are singletons. Thus, H_d satisfies the maximum entropy and monotonicity properties, but it does not satisfy the non-negativity and probabilistic consistency properties. However, it does satisfy the additivity and subadditivity properties [12]. Ramer [38] proves that H_d is the unique definition of non-specificity entropy of m that satisfies additivity and the subadditivity properties.

Lamata and Moral [27] suggest a definition of entropy of BPA m as follows:

$$H_l(m) = H_y(m) + H_d(m), \tag{27}$$

which combines Yager's definition $H_y(m)$ as a measure of conflict, and Dubois-Prade's definition $H_d(m)$ as a measure of non-specificity. It is easy to verify that $H_l(\iota_X) = H_l(m_u) = \log(|\Omega_X|)$, which violates the maximum entropy property. It satisfies the non-negativity, monotonicity, probability consistency, and additivity properties. It does not satisfy the subadditivity property [12].

Klir and Ramer [23] define entropy of BPA m for X as follows:

$$H_k(m) = \sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) \log \left(\frac{1}{1 - \sum_{\mathsf{b} \in 2^{\Omega_X}} m(\mathsf{b}) \frac{|\mathsf{b} \setminus \mathsf{a}|}{|\mathsf{b}|}} \right) + H_d(m) \tag{28}$$

The first component in Eq. (28) is designed to measure conflict, and the second component is designed to measure non-specificity. It is easy to verify that $H_k(\iota_X) = H_k(m_u) = \log(|\Omega_X|)$, which violates the maximum entropy property. It satisfies the non-negativity, monotonicity, probability consistency, and additivity properties. It does not satisfy the subadditivity property [51].

Klir and Parviz [22] modify Klir and Ramer's definition $H_k(m)$ slightly to measure conflict in a more refined way. The revised definition is as follows:

$$H_p(m) = \sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) \log \left(\frac{1}{1 - \sum_{\mathsf{b} \in 2^{\Omega_X}} m(\mathsf{b}) \frac{|\mathsf{a} \setminus \mathsf{b}|}{|\mathsf{a}|}} \right) + H_d(m) \tag{29}$$

Klir and Parviz argue that the first component in Eq. (29) is a better measure of conflict than the first component in Eq. (28). Like $H_k(m)$, $H_p(m)$ satisfies the non-negativity, monotonicity, probability consistency, and additivity properties, but not the maximum entropy, and subadditivity [52] properties.

Pal *et al.* [35, 36] define entropy $H_b(m)$ as follows:

$$H_b(m) = \sum_{\mathsf{a} \in 2^{\Omega_X}} m(\mathsf{a}) \log \left(\frac{|\mathsf{a}|}{m(\mathsf{a})} \right) \tag{30}$$

$H_b(m)$ satisfies non-negativity, monotonicity, probability consistency, and additivity [36]. $H_b(\iota_X) = H_b(m_u) = \log(|\Omega_X|)$. Thus, it does not satisfy the maximum entropy property. The maximum value of $H_b(m)$ is attained for m such that $m(\mathsf{a}) \propto |\mathsf{a}|$, for all $\mathsf{a} \in 2^{\Omega_X}$. Thus, for a binary-valued variable X, the maximum value of $H_b(m)$ is 2 whereas $H_b(\iota_X) = 1$.

Suppose m is a BPA for X. Then, there exists a set of PMFs $\mathcal{P}_m$ such that each PMF $P_X \in \mathcal{P}_m$ satisfies Eq. (18) [14]. Thus, a BPA m can be interpreted as an encoding of a set of PMFs as described in Eq. (18). If $m = \iota_X$, then $\mathcal{P}_{\iota_X}$ includes the set of all PMFs for X. If m is a Bayesian BPA for X, then $\mathcal{P}_m$ includes a single PMF P_X corresponding to the Bayesian BPA m.

Unfortunately, this interpretation of a BPA function is incompatible with Dempster's rule of combination [41, 43, 44, 14]. Fagin and Halpern [13] propose a new rule for updating beliefs, which is referred to as the Fagin-Halpern combination rule. If we start with a set of PMFs characterized by BPA m for X, and we observe some event $\mathsf{b} \subset \Omega_X$, then one possible updating rule is to condition each PMF in the set $\mathcal{P}$ on event b, and then find a BPA m' that corresponds to the lower envelope of the revised set of PMFs. The Fagin-Halpern rule [13] does precisely this, and is different from Dempster's rule of conditioning, which is a special case of Dempster's rule of combination.

Example 3. *Consider a BPA m_1 for X with state space $\Omega_X = \{x_1, x_2, x_3\}$ as follows: $m_1(\{x_1\}) = 0.5$, $m_1(\Omega_X) = 0.5$. With the credal set semantics of a BPA function, m_1 corresponds to a set of PMFs $\mathcal{P}_{m_1} = \{P \in \mathcal{P} : P(x_1) \geq 0.5\}$, where $\mathcal{P}$ denotes the set of all PMFs for X. Now suppose we get a distinct piece of evidence m_2 for X such that $m_2(\{x_2\}) = 0.5$, $m_2(\Omega_X) = 0.5$. m_2 corresponds to $\mathcal{P}_{m_2} = \{P \in \mathcal{P} : P(x_2) \geq 0.5\}$. The only PMF that is in both $\mathcal{P}_{m_1}$ and $\mathcal{P}_{m_2}$ is $P \in \mathcal{P}$ such that $P(x_1) = P(x_2) = 0.5$, and $P(x_3) = 0$. Notice that if we use Dempster's rule to combine m_1 and m_2, we have: $(m_1 \oplus m_2)(\{x_1\}) = \frac{1}{3}$, $(m_1 \oplus m_2)(\{x_2\}) = \frac{1}{3}$, and $(m_1 \oplus m_2)(\Omega_X) = \frac{1}{3}$. The set of PMFs $\mathcal{P}_{m_1 \oplus m_2} = \{P \in \mathcal{P} : P(x_1) \geq \frac{1}{3}, P(x_2) \geq \frac{1}{3}\}$ is not the same as $\mathcal{P}_{m_1} \cap \mathcal{P}_{m_2}$. Thus, credal set semantics of belief functions are not compatible with Dempster's rule of combination.*

Maeda-Ichihashi [30] define $H_i(m)$ as follows:

$$H_i(m) = \max_{P_X \in \mathcal{P}_m} \{H_s(P_X)\} + H_d(m), \tag{31}$$

where the first component is interpreted as a measure of conflict only, and the second component is interpreted as a measure of non-specificity. $H_i(m)$ satisfies all properties including the subadditivity property described in Eq. (20) [30]. $H_i(m)$ may be appropriate for a theory of belief functions interpreted as a credal set with the Fagin-Halpern combination rule. It is, however, inappropriate for the Dempster-Shafer theory of belief functions with Dempster's rule as the rule for combining (or updating) beliefs.

Maeda-Ichihashi's definition $H_i(m)$ does not satisfy the monotonicity-like property suggested by Abellán-Moral [4]. They suggest a modification of the Maeda-Ichihashi's definition where they add a third component so that the modified definition satisfies the monotonicity-like property in addition to the six properties satisfied by the Maeda-Ichihashi's definition. Abellán [1] describes a decomposition of H_i into two components such that the first component (different from the first component in Eq. (31)) is a measure of conflict, and the second component is a measure of non-specificity. Regarding numerical computation of the first component of $H_i(m)$, which involves nonlinear optimization, some algorithms are described in [31, 33, 16, 28, 5].

Harmanec-Klir [15] define $H_h(m)$ as follows:

$$H_h(m) = \max_{P_X \in \mathcal{P}_m} H_s(P_X), \tag{32}$$

where they interpret $H_h(m)$ as a measure of total uncertainty. They interpret $\min_{P_X \in \mathcal{P}_m} H_s(P_X)$ as a measure of conflict, and the difference between $H_h(m)$ and $\min_{P_X \in \mathcal{P}_m} H_s(P_X)$ as a measure of non-specificity. $H_h(\iota_X) = H_h(m_u) = \log(|\Omega_X|)$. Thus, it doesn't satisfy the maximum entropy property. It does, however, satisfy all other properties including subadditivity. Like, Maeda-Ichihashi's definition, Harmanec-Klir's definition (based on credal set semantics) is inconsistent with Dempster's rule of combination.

Jousselme *et al.* [19] define $H_j(m)$ based on first transforming a BPA m to a PMF $BetP_m$ using the so-called *pignistic* transformation [11, 48], and then using Shannon entropy of $BetP_m$. First, let's define $BetP_m$. Suppose m is a BPA for X. Then $BetP_m$ is a PMF for X defined as follows:

$$BetP_m(x) = \sum_{\mathsf{a} \in 2^{\Omega_X} :\, x \in \mathsf{a}} \frac{m(\mathsf{a})}{|\mathsf{a}|} \tag{33}$$

for all $x \in \Omega_X$. It is easy to verify that $BetP_m$ is a PMF.

Example 4. *This example is taken from [49]. Consider a BPA m for X with $\Omega_X = \{x_1, \ldots, x_{70}\}$ as follows: $m(\{x_1\}) = 0.30$, $m(\{x_2\}) = 0.01$, and $m(\{x_2, \ldots, x_{70}\}) = 0.69$. Then $BetP_m$ is as follows: $BetP_m(x_1) = 0.30$, $BetP_m(x_2) = 0.02$, and $BetP_m(x_3) = \ldots = BetP_m(x_{70}) = 0.01$.*

Jousselme *et al.*'s definition of entropy of BPA m for X is as follows.

$$H_j(m) = H_s(BetP_m) = \sum_{x \in \Omega_X} BetP_m(x) \log\left(\frac{1}{BetP_m(x)}\right) \tag{34}$$

A similar definition, called *pignistic* entropy, appears in [10] in the context of the Dezert-Smarandache theory, which can be considered as a generalization of the D-S belief functions theory.

$H_j(m)$ satisfies the non-negativity, monotonicity, probability consistency, and additivity properties [19]. It does not satisfy the maximum entropy property as $H_j(\iota_X) = H_j(m_u) = \log(|\Omega_X|)$.

Although Jousselme *et al.* claim that $H_j(m)$ satisfies the subadditivity property (Eq. (20)), a counter-example is provided in [21].

One basic assumption behind $H_j(m)$ is that $BetP_m$ is an appropriate probabilistic representation of the uncertainty in m in the D-S theory. However, it is argued in [7] that $BetP_m$ is an inappropriate probabilistic representation of m in the D-S theory. Consider a situation where we have vacuous prior knowledge of X and we receive evidence represented as BPA m as described in Example 4. If $BetP_m$ were appropriate for m, then x_1 would be 15 times more likely than x_2. Now suppose we receive another distinct piece of evidence that is also represented by m. As per the D-S theory, our total evidence is now $m \oplus m$. If on the basis of m (or $BetP_m$), x_1 was 15 times more likely than x_2, then now that we have evidence $m \oplus m$, x_1 should be 225 times more likely than x_2. But $BetP_{m\oplus m}(x_1) \approx 0.156$ and $BetP_{m\oplus m}(x_2) \approx 0.036$. So according to $BetP_{m\oplus m}$, x_1 is only 4.33 more likely than x_2. Thus, $BetP_m$ is not consistent with Dempster's combination rule. For this reason, we don't believe $H_j(m)$ is an appropriate measure of entropy of m in the D-S theory.

Pouly *et al.* [37] define entropy of a "hint" associated with a BPA m. A hint is a formalization of the multivalued mapping semantics for BPAs, and is more fine-grained than a BPA. Consider the following example. A witness claims that he saw the defendant commit a crime. Suppose that we have a PMF on the reliability R of the witness as follows. Let r and $\bar{r}$ denote the witness is reliable or not, respectively. Then, $P(r) = 0.6$, and $P(\bar{r}) = 0.4$. The question of interest, denoted by variable G, is whether the defendant is guilty (g) or not ($\bar{g}$). If the witness is reliable, then given his statement, the defendant is guilty. If the witness is not reliable, then his claim has no bearing on the question of guilt of the defendant. Thus, we have a multivalued mapping $\Gamma : \{r, \bar{r}\} \to 2^{\{g,\bar{g}\}}$ such that $\Gamma(r) = \{g\}$, and $\Gamma(\bar{r}) = \{g, \bar{g}\}$. In this example, the hint $\mathcal{H} = (\{r, \bar{r}\}, \{g, \bar{g}\}, P, \Gamma)$. The hint $\mathcal{H}$ induces a BPA for G as follows: $\mathrm{m}(\{\mathrm{g}\}) = 0.6, m(\{g, \bar{g}\}) = 0.4$.

Pouly *et al.*'s definition of entropy of hint $\mathcal{H} = (\Omega_1, \Omega_2, P, \Gamma)$ is as follows:

$$H_r(\mathcal{H}) = H_s(P) + H_d(m) \tag{35}$$

where m is the BPA on state space Ω_2 induced by hint $\mathcal{H}$. The expression in Eq. (35) is derived using Shannon's entropy of a joint PMF on the space $\Omega_1 \times \Omega_2$ whose marginal for Ω_1 is P, and an assumption of uniform conditional PMF for $\Gamma(\omega) \subseteq \Omega_2$ given $\omega \in \Omega_1$. This assumption results in a marginal PMF for Ω_2 that is equal to $BetP_m$, where m is the BPA on state space Ω_2 induced by hint $\mathcal{H}$. Dempster's combination rule never enters the picture in the derivation on $H_r(\mathcal{H})$. $H_r(\mathcal{H})$ has nice properties (on the space of hints). $H_r(\mathcal{H})$ in Eq. (35) is on the scale $[0, \log(|\Omega_1|) + \log(|\Omega_2|)]$. For a BPA m defined on the state space Ω_2, it would make sense to use only the marginal of the joint PMF on $\Omega_1 \times \Omega_2$ for Ω_2, which is $BetP_m$. Thus, if one were to adapt Pouly *et al.*'s definition for BPAs, it would coincide with the Jousselme *et al.*'s definition, i.e.,

$$H_r(m) = H_j(m) = H_s(BetP_m) = \sum_{\theta \in \Omega_2} BetP_m(\theta) \log\left(\frac{1}{BetP_m(\theta)}\right) \tag{36}$$

Thus, Pouly *et al.*'s definition of entropy of BPA m would have the same properties as Jousselme *et al.*'s definition.

Table 1 summarizes the properties of the various definitions of entropy for belief functions in the D-S theory.

5 A New Definition of Entropy for D-S Theory

In this section, we propose a new definition of entropy for D-S theory. The new definition of entropy is based on the plausibility transformation of a belief function to a probability function. Therefore, we start this section by describing the plausibility transformation introduced originally in [7].

Table 1: A Summary of Properties of Various Definitions of Entropy of D-S Belief Functions

Definition	*Non-neg.*	*Max. ent.*	*Monoton.*	*Prob. Cons.*	*Additivity*	*Subadd.*	*Cons. w D-S*
Höhle, Eq. (22)	no	no	no	yes	yes	no	yes
Smets, Eq. (23)	no	no	no	no	yes	no	yes
Yager, Eq. (24)	no	no	no	yes	yes	no	yes
Nguyen, Eq. (25)	no	no	no	yes	yes	no	yes
Dubois-Prade, Eq. (26)	no	yes	yes	no	yes	yes	yes
Lamata-Moral, Eq. (27)	yes	no	yes	yes	yes	no	yes
Klir-Ramer, Eq. (28)	yes	no	yes	yes	yes	no	yes
Klir-Parviz, Eq. (29)	yes	no	yes	yes	yes	no	yes
Pal *et al.*, Eq. (30)	yes	no	yes	yes	yes	no	yes
Maeda-Ichihashi, Eq. (31)	yes	yes	yes	yes	yes	yes	no
Harmanec-Klir, Eq. (32)	yes	no	yes	yes	yes	yes	no
Jousselme *et al.*, Eq. (34)	yes	no	yes	yes	yes	no	no
Pouly *et al.*, Eq. (36)	yes	no	yes	yes	yes	no	no

Plausibility Transformation of a BPA to a PMF Suppose m is a BPA for X. What is the PMF of X that best represents m in the D-S theory? An answer to this question is given by Cobb and Shenoy [7], who propose the plausibility transformation of m as follows. First consider the plausibility function Pl_m corresponding to m. Next, construct a PMF for X, denoted by P_{Pl_m}, by the values of Pl_m for singleton subsets suitably normalized, i.e.,

$$P_{Pl_m}(x) = K^{-1} \cdot Pl_m(\{x\}) = K^{-1} \cdot Q_m(\{x\}) \tag{37}$$

for all $x \in \Omega_X$, where K is a normalization constant that ensures P_{Pl_m} is a PMF, i.e., $K = \sum_{x \in \Omega_X} Pl_m(\{x\}) = \sum_{x \in \Omega_X} Q_m(\{x\})$.

Example 5. *Consider a BPA m for X as described in Example 4 as follows: $m(\{x_1\}) = 0.30$, $m(\{x_2\}) = 0.01$, $m(\{x_2, \ldots, x_{70}\}) = 0.69$. Then, Pl_m for singleton subsets is as follows: $Pl_m(\{x_1\}) = 0.30$, $Pl_m(\{x_2\}) = 0.70$, $Pl_m(\{x_3\}) = \cdots = Pl_m(\{x_{70}\}) = 0.69$. The plausibility transformation of m is as follows: $P_{Pl_m}(x_1) = 0.3/49.72 \approx 0.0063$, and $P_{Pl_m}(x_2) = 0.7/49.72 \approx 0.0.0146$, and $P_{Pl_m}(x_3) = \cdots = P_{Pl_m}(x_{70}) \approx 0.0144$. Notice that P_{Pl_m} is quite different from $BetP_m$. In $BetP_m$, x_1 is 15 times more likely than x_2. In P_{Pl_m}, x_2 is 2.33 times more likely than x_1.*

[7] argues that of the many methods for transforming belief functions to PMFs, the plausibility transformation is one that is consistent with Dempster's rule of combination in the sense that if we have BPAs $m_1, \ldots, m_k$ for X, then $P_{Pl_{m_1 \oplus \ldots \oplus m_k}} = P_{Pl_{m_1}} \otimes \ldots \otimes P_{Pl_{m_k}}$, where $\otimes$ denotes Bayes combination rule (pointwise multiplication followed by normalization [53]). It can be shown that the plausibility transformation is the *only* transformation method that has this property, which follows from the fact that Dempster's rule of combination is pointwise multiplication of commonality functions followed by normalization (Eq. (15)).

Example 6. *Consider the BPA m for X in Example 4. Notice that as per P_{Pl_m}, x_2 is 2.33 times more likely than x_1. Now suppose we get a distinct piece of evidence that is identical to m, so that our total evidence is $m \oplus m$. If we compute $m \oplus m$ and $P_{Pl_{m \oplus m}}$, then as per $P_{Pl_{m \oplus m}}$, x_2 is 2.33^2 more likely than x_1. This is a direct consequence of the consistency of the plausibility transformation with Dempster's combination rule.*

A New Definition of Entropy of a BPA To explain the basic idea behind the following definition consider a simple example with an urn containing n balls of up to two colors: white (w), and black (b). Suppose we draw a ball at random from the urn and X denotes its color. What is the entropy of the BPA for X in the situation where we know that there is at least one ball of each color in the urn? The simplest case is when $n = 2$. In this case the entropy is exactly the same as in tossing a fair coin: $\log(2) = 1$. Naturally, the greater n is, the greater uncertainty in the model. As there is no information preferring one color to another one, the only probabilistic description of the model is a uniform PMF. In D-S theory, the BPA describing this situation is $m(\{w\}) = m(\{b\}) = \frac{1}{n}$, and $m(\{w, b\}) = \frac{n-2}{n}$. Therefore, the entropy function for this BPA must be greater than or equal to Shannon's entropy of a uniform PMF with two states ($\log(2) = 1$), and increasing with increasing n. This is why the following definition of entropy of a BPA m consists of two components. The first component is Shannon's entropy of a PMF that corresponds to m, and the second component includes entropy associated with non-singleton focal sets of m.

Suppose m is a BPA for X. The entropy of m is defined as follows:

$$H(m) = H_s(P_{Pl_m}) + H_d(m) = \sum_{x\in\Omega_X} P_{Pl_m}(x)\log\left(\frac{1}{P_{Pl_m}(x)}\right) + \sum_{\mathsf{a}\in 2^{\Omega_X}} m(\mathsf{a})\log(|\mathsf{a}|). \tag{38}$$

Like some of the definitions in the literature, the first component in Eq. (38) is designed to measure conflict in m, and the second component is designed to measure non-specificity in m. Both components are on the scale $[0, \log(|\Omega_X|)]$, and therefore, $H(m)$ is on the scale $[0, 2\log(|\Omega_X|)]$.

Theorem 1. *The entropy $H(m)$ for BPA m for X defined in Eq. (38) satisfies the non-negativity, maximum entropy, monotonicity, probability consistency, and additivity properties.*

Proof. We know that $H_s(P_{Pl_m}) \geq 0$, and $H_d(m) \geq 0$. Thus, $H(m) \geq 0$. For $H(m) = 0$ to hold, both $H_s(P_{Pl_m}) = 0$, and $H_d(m) = 0$ must be satisfied. $H_s(P_{Pl_m}) = 0$ if and only if there exists $x \in \Omega_X$ such that $P_{Pl_m}(x) = 1$, which occurs if and only if $m(\{x\}) = 1$. $H_d(m) = 0$ if and only if m is Bayesian. Thus, $H(m)$ satisfies the non-negativity property.

Let n denote $|\Omega_X|$. Then $P_{Pl_{\iota_X}}(x) = \frac{1}{n}$ for all $x \in \Omega_X$, and therefore $H_s(P_{Pl_{\iota_X}}) = \log(n)$, which is the maximum of all PMFs defined on Ω_X. Also $H_d(\iota_X) = \log(n)$, which is the maximum of Dubois-Prade's entropy over all BPAs m for X. Thus, $H(m)$ satisfies the maximum entropy property.

$H(\iota_X) = 2\log(|\Omega_X|)$. Thus, since it is monotonic in $|\Omega_X|$, $H(m)$ satisfies the monotonicity property.

If m is Bayesian, then $P_{Pl_m}(x) = m(\{x\})$ for all $x \in \Omega_X$, and $H_d(m) = 0$. Thus, $H(m)$ satisfies the probability consistency property.

Suppose m_X is a BPA for X, and m_Y is a BPA for Y. Then, as it is shown in [7], $P_{Pl_{m_X\oplus m_Y}} = P_{Pl_{m_X}} \otimes P_{Pl_{m_Y}}$, and the normalization constant in the case of PMFs for disjoint arguments is 1. Thus, $H_s(P_{Pl_{m_X\oplus m_Y}}) = H_s(P_{Pl_{m_X}}) + H_s(P_{Pl_{m_Y}})$. Also, it is proved in [12], that $H_d(m_X \oplus m_Y) = H_d(m_X) + H_d(m_Y)$. Thus, $H(m)$ satisfies the additivity property. □

In the next section, we provide an example that shows that our definition does not satisfy the subadditivity property.

Property 5 was stated in terms of BPAs m_X for X and m_Y for Y. Suppose we have a set of variables, say v, and $r, s \subseteq v$. This property could have been stated more generally in terms of BPAs m_1 for r and m_2 for s where $r \cap s = \emptyset$. In this case still $H(m_1 \oplus m_2) = H(m_1) + H(m_2)$ because both components of the new definition (i.e., H_s and H_d) satisfy the more general property. However, if $r \cap s \neq \emptyset$, then generally $H(m_1 \oplus m_2)$ may be different from $H(m_1) + H(m_2)$. This is because neither the first component of the new definition, nor the Dubois-Prade component, satisfy the stronger property. An example illustrating this is described next. Thus, our new definition does not satisfy the compound property in Eq. (21).

Example 7. *Consider BPA* m_1 *for binary-valued variable* X *as follows:* $m_1(\{x\}) = 0.1$, $m_1(\{\bar{x}\}) = 0.2$, $m_1(\Omega_X) = 0.7$, *and BPA* m_2 *for* $\{X, Y\}$ *as follows:* $m_2(\{(x, y), (\bar{x}, y)\}) = 0.08$, $m_2(\{(x, y), (\bar{x}, \bar{y})\}) = 0.72$, $m_2(\{(x, \bar{y}), (\bar{x}, y)\}) = 0.02$, $m_2(\{(x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.18$. *Assuming these two BPAs represent distinct pieces of evidence, we can combine them with Dempster's rule obtaining* $m = m_1 \oplus m_2$ *for* $\{X, Y\}$ *as follows:* $m(\{(x, y)\}) = 0.08$, $m(\{(x, \bar{y})\}) = 0.02$, $m(\{(\bar{x}, y)\}) = 0.02$, $m(\{(\bar{x}, \bar{y})\}) = 0.18$, $m(\{(x, y), (\bar{x}, y)\}) = 0.056$, $m(\{(x, y), (\bar{x}, \bar{y})\}) = 0.504$, $m(\{(x, \bar{y}), (\bar{x}, y)\}) = 0.014$, $m(\{(x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.126$.

Now, the PMF $P_{Pl_{m_1}}$ *of* X *obtained using the plausibility transformation of* m_1 *is as follows:* $P_{Pl_{m_1}}(x) = 0.47$, *and* $P_{Pl_{m_1}}(\bar{x}) = 0.53$, *and its Shannon's entropy is* $H_s(P_{Pl_{m_1}}) = 0.998$. $H_d(m_1) = 0.7$. *Thus,* $H(m_1) = 1.698$.

The PMF $P_{Pl_{m_2}}$ *of* $\{X, Y\}$ *obtained using the plausibility transformation is* $P_{Pl_{m_2}}(x, y) = 0.4$, $P_{Pl_{m_2}}(x, \bar{y}) = 0.1$, $P_{Pl_{m_2}}(\bar{x}, y) = 0.05$, $P_{Pl_{m_2}}(\bar{x}, \bar{y}) = 0.45$, *and its Shannon's entropy is* $H_s(P_{Pl_{m_2}}) = 1.595$. $H_d(m_2) = 1$. *Thus,* $H(m_2) = 2.595$.

The joint PMF of $\{X, Y\}$ *obtained using the plausibility transformation is as follows:* $P_{Pl_m}(x, y) = 0.38$, $P_{Pl_m}(x, \bar{y}) = 0.09$, $P_{Pl_m}(\bar{x}, y) = 0.05$, $P_{Pl_m}(\bar{x}, \bar{y}) = 0.48$, *and its Shannon's entropy is* $H(P_{Pl_m}) = 1.586$. *Also, Dubois-Prade's entropy of* m *is* $H_d(m) = 0.7$. *Thus,* $H(m) = 2.286$.

Notice that $H(m) = 2.286 \neq H(m_1) + H(m_2) = 1.698 + 2.595 = 4.293$, $H(Pl_m) = 1.586 \neq H(P_{Pl_{m_1}}) + H(P_{Pl_{m_2}}) = 0.998 + 1.595 = 2.593$, *and* $H_d(m) = 0.7 \neq H_d(m_1) + H_d(m_2) = 0.7 + 1 = 1.7$.

6 Additional Properties of $H(m)$

In this section, we will describe some additional properties of $H(m)$ defined in Eq. (38).

Entropy as an expected value One interpretation of Shannon entropy in probability theory is that it equals the expected value of an information received when learning one symbol $x \in \Omega_X$, i.e.,

$$H_s(P_X) = \sum_{x \in \Omega_X} P_X(x) I(x), \tag{39}$$

where

$$I(x) = \log_2 \left(\frac{1}{P_X(x)} \right)$$

expresses the information received when learning that state $x \in \Omega_X$ has occurred. Notice that the amount of this information is *not* the property of state x, but that of its probability.

In the case where our knowledge is encoded by a BPA m (instead of a PMF), we can decompose the information in m into two parts. The first part is the PMF P_{Pl_m}, and the second part (not captured by the first part) is $\log(|\mathsf{a}|)$, which happens with probability $m(\mathsf{a})$. Consider the vacuous BPA function ι_X for X, where $\Omega_X = \{x, \bar{x}\}$. We can decompose the uncertainty in ι_X into the uncertainty in the PMF $P_{Pl_{\iota_X}}$ (which is given by $P_{Pl_{\iota_X}}(x) = 1/2$, and $P_{Pl_{\iota_X}}(\bar{x}) = 1/2$). But this doesn't capture the entire uncertainty in ι_X. We also have to include the uncertainty $\log(|\Omega_X|)$. The expected value of the first part is Shannon's entropy $H(P_{Pl_{\iota_X}}) = 1$, and the expected value of the second is $\iota_X(\Omega_X)\,\log(|\Omega_X|) = 1$.

Thus, we can interpret $H(m)$ as an expected value, but with respect to two different sources of uncertainty. The first part is expected value of information $I(x)$ with respect to PMF P_{Pl_m}, and the second part is expected value of information necessary to eliminate the uncertainty emerging from the size of Ω_X, i.e., $\log(|\mathsf{a}|)$, with respect to "distribution" m, i.e., $\sum_{\mathsf{a}\in 2^{\Omega_X}} m(\mathsf{a})\log(|\mathsf{a}|)$. The second part corresponds to the measure of uncertainty suggested by Richard Hartley in 1928 [17], about which Rényi showed that it is the only one satisfying additivity and monotonicity properties (for a precise formulation of this property see [39]).

Subadditivity Property As shown in Example 8 below, our definition does not satisfy the subadditivity property in Eq. (20).

Example 8. *Consider a two-dimensional BPA m for binary-valued variables $\{X, Y\}$ with five focal elements:*

$$m(\{(x,y)\}) = m(\{(x,\bar{y})\}) = 0.1,\ m(\{(\bar{x},y)\}) = m(\{(\bar{x},\bar{y})\}) = 0.3,\ \text{and}\quad m(\Omega_{\{X,Y\}}) = 0.2.$$

The joint PMF of $\{X, Y\}$ using the plausibility transformation is as follows: $P_{Pl_m}((x,y)) = 0.1875$, $P_{Pl_m}((x,\bar{y})) = 0.1875$, $P_{Pl_m}((\bar{x},y)) = 0.3125$, $P_{Pl_m}((\bar{x},\bar{y})) = 0.3125$. Its Shannon's entropy is $H_s(P_{Pl_m}) = 1.9544$. The Dubois-Prade's entropy of m is $H_d(m) = 0.4$. Thus, $H(m) = 2.3544$.

The marginal BPA $m^{\downarrow X}$ is as follows: $m^{\downarrow X}(\{x\}) = 0.2$, $m^{\downarrow X}(\{\bar{x}\}) = 0.6$, and $m^{\downarrow X}(\Omega_X) = 0.2$. The PMF $P_{Pl_{m^{\downarrow X}}}$ of X obtained using the plausibility transformation of $m^{\downarrow X}$ is as follows: $P_{Pl_{m^{\downarrow X}}}(x) = 0.333$, and $P_{Pl_{m^{\downarrow X}}}(\bar{x}) = 0.667$, and its Shannon's entropy is $H_s(P_{Pl_{m^{\downarrow X}}}) = 0.9183$.

Similarly, the marginal BPA $m^{\downarrow Y}$ is as follows: $m^{\downarrow Y}(\{y\}) = 0.4$, $m^{\downarrow Y}(\{\bar{y}\}) = 0.4$, and $m^{\downarrow Y}(\Omega_Y) = 0.2$. The PMF $P_{Pl_{m^{\downarrow Y}}}$ of Y is as follows: $P_{Pl_{m^{\downarrow Y}}}(y) = P_{Pl_{m^{\downarrow Y}}}(\bar{y}) = 0.5$, and therefore its Shannon's entropy is $H_s(P_{Pl_{m^{\downarrow Y}}}) = 1$.

Thus, $H_s(P_{Pl_m}) = 1.9544 > H_s(P_{Pl_{m^{\downarrow X}}}) + H_s(P_{Pl_{m^{\downarrow Y}}}) = 0.9183 + 1 = 1.9182$. Dubois-Prade's entropies are as follows: $H_d(m^{\downarrow X}) = H_d(m^{\downarrow Y}) = 0.2$. Thus, $H_d(m) = 0.4 = H_d(m^{\downarrow X}) + H_d(m^{\downarrow Y}) = 0.2 + 0.2 = 0.4$. Therefore, $H(m) = 2.3544 > H(m^{\downarrow X}) + H(m^{\downarrow Y}) = (0.9183 + 0.2) + (1 + 0.2) = 1.1183 + 1.2 = 2.3183$.

Entropy of $m \oplus m$ It is well known that Dempster's rule of combination $\oplus$ is, in general, not idempotent, i.e., in general $m \oplus m \neq m$. It is easy to confirm that vacuous, Bayesian uniform, and Bayesian deterministic BPAs are idempotent. Notice that these types of BPAs express extreme uncertainty. For Bayesian deterministic BPA there is no uncertainty, and therefore the measure of uncertainty is 0. The Bayesian uniform and vacuous BPAs express maximal uncertainties: probabilistic and absolute, respectively. Therefore, it is quite natural that for these specific BPAs,

$$H(m \oplus m) = H(m).$$

However, a natural question arises: What is the value of $H(m \oplus m)$ in the general case? *Repetitio est mater studiorum.* Learning the same knowledge twice should contribute to our cognizance more than just learning it only once. Therefore we should not be surprised that the following assertion holds true.

Theorem 2. *For a Bayesian BPA m for X,*

$$H(m \oplus m) \leq H(m), \tag{40}$$

with equality if and only if m is idempotent with respect to Dempster's rule of combination, i.e., $m \oplus m = m$.

Proof. We will use an obvious property of Shannon entropy H_s: for two PMF P_1 and P_2 for X, such that $P_1(x) = P_2(x)$ for all $x \in \Omega_X \setminus \{y,z\}$, $P_1(y) - P_2(y) = P_2(z) - P_1(z) > 0$, and $P_1(y) \leq P_1(z)$ it holds that

$$H_s(P_1) > H_s(P_2). \tag{41}$$

This property can be proven by the following simple consideration. Denote $\varepsilon = P_1(y) - P_2(y)$, and

$$\begin{aligned} f(\varepsilon) &= H_s(P_1) - H_s(P_2) \\ &= -P_1(y)\log_2(P_1(y)) - P_1(z)\log_2(P_1(z)) \\ &\quad +(P_1(y)-\varepsilon)\log_2(P_1(y)-\varepsilon) + (P_1(z)+\varepsilon)\log_2(P_1(z)+\varepsilon). \end{aligned}$$

Since $f(0) = 0$, and

$$f'(\alpha) = \log_2\left(\frac{P_1(z)+\alpha}{P_1(y)-\alpha}\right) \cdot (\ln(2))^{-1}$$

is non-negative for all $\alpha \in [0,\varepsilon]$, it is clear that $f(\varepsilon) > 0$, and therefore strict inequality (41) holds true.

To prove inequality (40) for Bayesian BPA m, we use the fact that in this case $H(m) = H_s(P_{Pl_m}) = H_s((m(\{x\}))_{x\in\Omega_X})$. Moreover, we can restrict our attention to situations when $m \neq m \oplus m$, because if $m = m \oplus m$, the assertion (Eq. (40)) holds with equality. Notice that in this case

$$(m \oplus m)(\{x\}) = (m(\{x\}))^2 \cdot K^{-1},$$

where $K = \sum_{x\in\Omega_X} (m(\{x\}))^2$, and therefore

$$\begin{array}{ll} (m\oplus m)(\{x\}) < m(\{x\}) & \text{iff } m(\{x\}) < K, \\ (m\oplus m)(\{x\}) > m(\{x\}) & \text{iff } m(\{x\}) > K, \end{array} \tag{42}$$

which means that the values $m(\{y\})$ to be decreased are always less than the values $m(\{z\})$ to be increased,

To finish the proof we will construct a finite sequence of Bayesian BPAs, such that $m = m_0, m_1, m_2, \ldots, m_k = m \oplus m$, and $H(m_i) < H(m_{i-1})$ for all $i = 1, 2, \ldots, k$.

Consider m_i, and denote $\mathsf{a}_i = \{x \in \Omega_X : m_i(\{x\}) \neq (m\oplus m)(\{x\})\}$. Let y be the element of a_i, for which the difference between m_i and $m \oplus m$ is minimal, i.e.,

$$|m_i(\{y\}) - (m\oplus m)(\{y\})| \leq |m_i(\{x\}) - (m\oplus m)(\{x\})| \ \ \forall x \in \mathsf{a}_i. \tag{43}$$

Naturally, there must exist $z \in \mathsf{a}_i$ such that

$$\text{sign}(m_i(\{z\}) - (m\oplus m)(\{z\})) = -\text{sign}(m_i(\{y\}) - (m\oplus m)(\{y\})), \tag{44}$$

and, because of (43), $|m_i(\{z\}) - (m\oplus m)(\{z\})| \geq |m_i(\{y\}) - (m\oplus m)(\{y\})|$. Therefore we can define BPA m_{i+1}:

$$\begin{aligned} m_{i+1}(\{y\}) &= 0, \\ m_{i+1}(\{z\}) &= m_i(\{z\}) + (m_i(\{y\}) - (m\oplus m)(\{y\})), \\ m_{i+1}(\{x\}) &= m_i(\{x\}) \ \text{ for all } x \in \Omega_X \setminus \{y,z\}. \end{aligned}$$

We immediately see that $|\mathsf{a}_{i+1}| > |\mathsf{a}_i|$, and therefore the sequence $m = m_0, m_1, m_2, \ldots, m_k = m \oplus m$ must be finite. We also can see that we did not violate property (41), and therefore each pair m_i and m_{i+1} meets the assumptions of the property presented at the beginning of this proof. Therefore $H_s((m_{i+1})_{x\in\Omega_X}) < H_s((m_i)_{x\in\Omega_X})$, which completes the proof. $\square$

A Bayesian BPA m for X is equivalent to a PMF P_X for X such that $P_X(x) = m(\{x\})$ for all $x \in \Omega_X$. Also, for a Bayesian BPA m, $m \oplus m = P_X \otimes P_X$, and $H(m) = H_s(P_X)$, where P_X is the PMF for X corresponding to m. Therefore, we have the following corollary.

Corollary 1. *For a PMF P_X for X,*

$$H_s(P_X \otimes P_X) \leq H_s(P_X), \tag{45}$$

with equality if and only if P_X is idempotent with respect to Bayesian combination rule, i.e., $P_X = P_X \otimes P_X$.

One may be tempted to believe that the property in Theorem 2 also holds for all BPAs. But, as shown in Example 9, it is not true.

Example 9. *Consider a BPA m for X, where $\Omega_X = \{x_1, x_2, x_3\}$ as follows: $m(\{x_1\}) = \frac{1}{3}$, $m(\{x_2, x_3\}) = \frac{2}{3}$. The Dubois-Prade entropy $H_d(m) = \frac{2}{3}$. Also, for this BPA m, the PMF P_{Pl_m} is as follows: $P_{Pl_m}(x_1) = \frac{1}{5}$, $P_{Pl_m}(x_2) = P_{Pl_m}(x_3) = \frac{2}{5}$. Thus, $H_s(P_{Pl_m}) = 1.522$, and $H(m) = H_s(P_{Pl_m}) + H_d(m) = 2.189$.*

If we compute $m \oplus m$, we have $(m \oplus m)(\{x_1\}) = \frac{1}{5}$, and $(m \oplus m)(\{x_2, x_3\}) = \frac{4}{5}$. Dubois-Prade entropy $H_d(m \oplus m) = \frac{4}{5}$. Notice that $H_d(m \oplus m) > H_d(m)$. The PMF $P_{Pl_{m \oplus m}}$ is as follows: $P_{Pl_{m \oplus m}}(x_1) = \frac{1}{9}$, $P_{Pl_{m \oplus m}}(x_2) = P_{Pl_{m \oplus m}}(x_3) = \frac{4}{9}$. And, its Shannon entropy $H_s(P_{Pl_{m \oplus m}}) = 1.392$. Notice that $H_s(P_{Pl_{m \oplus m}}) < H_s(P_{Pl_m})$. However, $H(m \oplus m) = H_s(m \oplus m) + H_d(m \oplus m) = 2.192$, which is greater than $H(m) = 2.189$.

Our definition of entropy $H(m)$ has two components. The first one, $H_s(P_{Pl_m})$ can be considered as a measure of conflict (or confusion or dissonance or discord or strife), and the second one, $H_d(m)$ can be considered as a measure of non-specificity. For probability distributions, the second component is zero as all masses are on singleton subsets. The first component does satisfy the intuition behind Theorem 2, the second component doesn't. Thus, while Theorem 2 holds for probability distributions, it is not valid for BPA in the DS theory because of the non-specificity component. When we combine m with itself, probability will migrate from subsets with lower plausibility to subsets with larger plausibility ([7]). If we have a BPA such that a larger subset has higher plausibility, then $H_d(m \oplus m) > H_d(m)$.

7 Summary and Conclusions

Interpreting Shannon's entropy of a PMF of a discrete random variable as the amount of uncertainty in the PMF [45], we propose five desirable properties of entropy of a basic probability assignment in the D-S theory of belief functions. These five properties are motivated by the analogous properties of Shannon's entropy for PMFs, and they are based on our intuition that a vacuous belief function has more uncertainty than a Bayesian belief function. These five properties are different from the five properties proposed by Klir and Wierman [24]. Two of the properties they require, set consistency and range, are inconsistent with some of the properties we propose. Also, one of the properties that they require, subadditivity, is not included in our set as we are unable to formulate a definition of entropy that would simultaneously satisfy the five properties we suggest plus subadditivity. Also, besides the five properties, we also require that any definition should be based on semantics consistent with the D-S theory of belief functions (with Dempster's rule as the combination rule), $H(m)$ should always exist, and $H(m)$ should be a continuous function of m. Thus, a monotonicity-like property suggested by Abellán-Masegosa [3] based on credal set semantics of belief functions that are not compatible with Dempster's rule is not included in our set of requirements.

We review some earlier definitions given by Höhle [18], Smets [47], Yager [55], Nguyen [34], Dubois-Prade [12], Lamata-Moral [27], Klir-Ramer [23], Klir-Parviz [22], Pal *et al.* [36], Maeda-Ichihashi [30], Harmanec-Klir [15], Jousselme *et al.* [19], and Pouly *et al.* [37]. None of these definitions satisfy all five properties listed earlier while satisfying the consistency with D-S theory semantics requirement. Pouly *et al.*'s definition is for the joint space of hints, $\Omega_1 \times \Omega_2$. If one were to adapt Pouly *et al.*'s definition for BPAs, then as the marginal entropy for Ω_2 reduces to the pignistic entropy, their definition for BPAs would coincide with that proposed by Jousselme *et al.*

Smets' definition is motivated by interpreting $H(m)$ as a measure of information contained in m, rather than uncertainty. Höhle's, Yager's, and Nguyen's definitions are motivated by interpreting entropy of a BPA as a measure of only its conflict (or confusion or discord or strife). Dubois-Prade's definition is motivated by interpreting entropy of a BPA as a measure of its non-specificity (or imprecision).

As first suggested by Lamata and Moral [27], we propose a new definition of entropy for BPA as a combination of Shannon's entropy for an equivalent PMF that captures the conflict measure of entropy, and Dubois-Prade's entropy of a BPA that captures the non-specificity (or Hartley) measure of entropy. The equivalent PMF is that obtained by using the plausibility transformation [7]. We show that this new definition satisfies all five properties we propose. More importantly, our definition is consistent with the semantics for the D-S theory of belief functions.

One could create a definition, e.g., that combines Jousselme *et al.*'s definition (Eq. (34)) with Dubois-Prade's definition (Eq. (26)), i.e., $H(m) = H_j(m) + H_d(m)$, and such a definition would also satisfy all five properties, but as we have argued before, the first component, pignistic entropy, is not consistent with semantics for the D-S theory.

We also describe some additional properties of our definition of entropy of BPA m. In particular, we describe our definition as the sum of an expected value of Shannon's entropy, which is a measure of conflict, and expected value of Hartley's entropy, which is a measure of non-specificity. We demonstrate that our definition does not satisfy the subadditivity property. This is because the first component, $H_s(P_{Pl_m})$, does not satisfy the subadditivity property. Finally, we show that while Shannon's entropy satisfies the inequality $H_s(P_X \otimes P_X) \leq H(P_X)$, our definition of $H(m)$ does not satisfy the corresponding inequality, $H(m \oplus m) \leq H(m)$. This is because the Dubois-Prade component, generalized Hartley entropy, does not satisfy this inequality, i.e., $H_d(m \oplus m)$ may be greater than $H_d(m)$.

An open question is whether there exists a definition of entropy for BPA m in the D-S theory that satisfies the five properties we list in Section 4, the subadditivity property, and most importantly, that is consistent with semantics for the D-S theory. Our definition satisfies the five properties and is consistent with semantics for the D-S theory, but it does not satisfy the subadditivity property.

Another open question is whether there is a class of BPA functions such that for this class, our definition satisfies the subadditivity property.

A more ambitious agenda is to define entropy for D-S belief functions that satisfies the compound property described in Eq. (21), that is analogous to the one that characterizes Shannon's entropy. If we can also show that $H(m_{Y|X}) \leq H(m^{\downarrow Y})$, then such a definition will enable generalizing Kullback-Leibler divergence concept to the D-S theory of belief functions.

Acknowledgments

This work has been supported in part by funds from grant GAČR 15-00215S to the first author, and from the Ronald G. Harper Distinguished Professorship at the University of Kansas to the second author. We are extremely grateful to comments from Thierry Denoeux, Marc Pouly, Anne-Laure Jousselme, and Joaquín Abellán for their comments on earlier drafts of this paper. We are also grateful to Suzanna Emelio for a careful proof-reading of the text.

References

References

[1] J. Abellán. Combining nonspecificity measures in Dempster-Shafer theory of evidence. *International Journal of General Systems*, 40(6):611–622, August 2011.

[2] J. Abellán, G. J. Klir, and S. Moral. Disaggregated total uncertainty measure for credal sets. *International Journal of General Systems*, 35(1):29–44, February 2006.

[3] J. Abellán and A. Masegosa. Requirements for total uncertainty measures in Dempster-Shafer theory of evidence. *International Journal of General Systems*, 37(6):733–747, December 2008.

[4] J. Abellán and S. Moral. Completing a total uncertainty measure in Dempster-Shafer theory. *International Journal of General Systems*, 28(4–5):299–314, 1999.

[5] J. Abellán and S. Moral. An algorithm that computes the upper entropy for order-2 capacities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(2):141–154, 2005.

[6] A. Bronevich and G. J. Klir. Measures of uncertainty for imprecise probabilities: An axiomatic approach. *International Journal of Approximate Reasoning*, 51:365–390, 2010.

[7] B. R. Cobb and P. P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314–340, 2006.

[8] M. Daniel. On transformations of belief functions to probabilities. *International Journal of Intelligent Systems*, 21(3):261–282, 2006.

[9] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[10] J. Dezert, F. Smarandache, and A. Tchamova. On the Blackman's association problem. In *Proceedings of the 6th Annual Conference on Information Fusion*, pages 1349–1356, Cairns, Queensland, Australia, 2003. International Society for Information Fusion.

[11] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M. M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, Amsterdam, 1982.

[12] D. Dubois and H. Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and Systems*, 24(2):161–182, 1987.

[13] R. Fagin and J. Y. Halpern. A new approach to updating beliefs. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 347–374, Amsterdam, 1991. North-Holland.

[14] J. Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54(3):275–317, 1992.

[15] D. Harmanec and G. J. Klir. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *International Journal of General Systems*, 22(4):405–419, 1994.

[16] D. Harmanec, G. Resconi, G. J. Klir, and Y. Pin. On the computation of uncertainty measure in Dempster-Shafer theory. *International Journal of General Systems*, 25(2):153–163, 1996.

[17] R. V. L. Hartley. Transmission of information. *The Bell Systems Technical Journal*, 7(3):535–563, 1928.

[18] U. Höhle. Entropy with respect to plausibility measures. In *Proceedings of the 12th IEEE Symposium on Multiple-Valued Logic*, pages 167–169, 1982.

[19] A.-L. Jousselme, C. Liu, D. Grenier, and E. Bossé. Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 36(5):890–903, 2006.

[20] G. J. Klir. Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like? *Fuzzy Sets and Systems*, 24(2):141–160, 1987.

[21] G. J. Klir and H. W. Lewis, III. Remarks on "Measuring ambiguity in the evidence theory". *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 38(4):995–999, 2008.

[22] G. J. Klir and B. Parviz. A note on the measure of discord. In D. Dubois, M. P. Wellman, B. D'Ambrosio, and P. Smets, editors, *Uncertainty in Artificial Intelligence: Proceedings of the Eighth Conference*, pages 138–141, San Mateo, CA, 1992. Morgan Kaufmann.

[23] G. J. Klir and A. Ramer. Uncertainty in the Dempster-Shafer theory: A critical re-examination. *International Journal of General Systems*, 18(2):155–166, 1990.

[24] G. J. Klir and M. J. Wierman. *Uncertainty-based Information: Elements of Generalized Information Theory*. Springer-Verlag, Berlin, 2nd edition, 1999.

[25] J. Kohlas and P.-A. Monney. *A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*. Springer-Verlag, Berlin, 1995.

[26] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:76–86, 1951.

[27] M. T. Lamata and S. Moral. Measures of entropy in the theory of evidence. *International Journal of General Systems*, 14(4):297–305, 1988.

[28] C. Liu, D. Grenier, A.-L. Jousselme, and E. Bosse. Reducing algorithm complexity for computing an aggregate uncertainty measure. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 37(5):669–679, 2007.

[29] D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, 2003.

[30] Y. Maeda and H. Ichihashi. An uncertainty measure under the random set inclusion. *International Journal of General Systems*, 21(4):379–392, 1993.

[31] Y. Maeda, H. T. Nguyen, and H. Ichihashi. Maximum entropy algorithms for uncertainty measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 1(1):69–93, 1993.

[32] D. A. Maluf. Monotonicity of entropy computations in belief functions. *Intelligent Data Analysis*, 1:297–213, 1997.

[33] A. Meyerowitz, F. Richman, and E. A. Walker. Calculating maximum-entropy probability densities for belief functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(4):377–389, 1994.

[34] H. T. Nguyen. On entropy of random sets and possibility distributions. In J. C. Bezdek, editor, *The Analysis of Fuzzy Information*, pages 145–156. CRC Press, Boca Raton, FL, 1985.

[35] N. R. Pal, J. C. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning I: A review. *International Journal of Approximate Reasoning*, 7(3):165–183, 1992.

[36] N. R. Pal, J. C. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning II: A new measure of total uncertainty. *International Journal of Approximate Reasoning*, 8(1):1–16, 1993.

[37] M. Pouly, J. Kohlas, and P. Y. A. Ryan. Generalized information theory for hints. *International Journal of Approximate Reasoning*, 54(1):228–251, 2013.

[38] A. Ramer. Uniqueness of information measure in the theory of evidence. *Fuzzy Sets and Systems*, 24(2):183–196, 1987.

[39] A. Rényi. *Probability Theory.* North-Holland, 1970.

[40] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, 1976.

[41] G. Shafer. Constructive probability. *Synthese*, 48(1):1–60, 1981.

[42] G. Shafer. Belief functions and parametric models. *Journal of the Royal Statistical Society, Series B*, 44(3):322–352, 1982.

[43] G. Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4(5–6):323–362, September–November 1990.

[44] G. Shafer. Rejoinders to comments on "Perspectives on the theory and practice of belief functions". *International Journal of Approximate Reasoning*, 6(3):445–480, May 1992.

[45] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[46] P. Smets. *Un modele mathematico-statistique simulant le processus du diagnostic medical.* PhD thesis, Free University of Brussels, 1978.

[47] P. Smets. Information content of an evidence. *International Journal of Man-Machine Studies*, 19:33–43, 1983.

[48] P. Smets. Constructing the pignistic probability function in a context of uncertainty. In M. Henrion, R. Shachter, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5*, pages 29–40, Amsterdam, 1990. North-Holland.

[49] P. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. P. Srivastava and T. J. Mock, editors, *Belief Functions in Business Decisions*, pages 316–332. Physica-Verlag, Heidelberg, 2002.

[50] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.

[51] J. Vejnarová. A few remarks on measures of uncertainty in Dempster-Shafer theory. In *Preliminary Proceedings of the 2nd Workshop on Uncertainty Processing in Expert Systems*, 1991.

[52] J. Vejnarová and G. J. Klir. Measure of strife in Dempster-Shafer theory. *International Journal of General Systems*, 22(1):25–42, 1993.

[53] F. Voorbraak. A computationally efficient approximation of Dempster-Shafer theory. *International Journal of Man-Machine Studies*, 30(5):525–536, 1989.

[54] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, New York, NY, 1991.

[55] R. Yager. Entropy and specificity in *A Mathematical Theory of Evidence*. *International Journal of General Systems*, 9(4):249–260, 1983.