Belief Functions: Theory and Applications

4th International Conference, BELIEF 2016
Prague, Czech Republic, September 21–23, 2016
Proceedings

Springer
Lecture Notes in Artificial Intelligence 9861

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel
  University of Alberta, Edmonton, Canada
Yuzuru Tanaka
  Hokkaido University, Sapporo, Japan
Wolfgang Wahlster
  DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann
  DFKI and Saarland University, Saarbrücken, Germany
More information about this series at http://www.springer.com/series/1244
Belief Functions: Theory and Applications

4th International Conference, BELIEF 2016
Prague, Czech Republic, September 21–23, 2016 Proceedings
Entropy of Belief Functions in the
Dempster-Shafer Theory: A New Perspective

Radim Jiroušek\textsuperscript{1,3} and Prakash P. Shenoy\textsuperscript{2(\textsuperscript{E})}

\textsuperscript{1} Faculty of Management, University of Economics,
Jindřichův Hradec, Czech Republic

\textsuperscript{2} School of Business, University of Kansas, Lawrence, KS, USA
pshenoy@ku.edu

\textsuperscript{3} Institute of Information Theory and Automation, Academy of Sciences,
Prague, Czech Republic
radim@utia.cz

Abstract. We propose a new definition of entropy of 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. 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 of probability functions, together with the implicit requirement that any definition should be consistent with 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 semantics of D-S theory, whereas none of the existing definitions do. Our definition does not satisfy the sub-additivity property. Whether there exists a definition that satisfies our five properties plus sub-additivity, and that is consistent with semantics for the D-S theory, remains an open question.

1 Introduction

The main goal of this paper is to provide a new definition of entropy of belief functions in the D-S theory that is consistent with semantics of the D-S theory. By entropy, we mean a real-valued measure of uncertainty in the tradition of Hartley [12] and Shannon [30]. Also, while there are several theories of belief functions (see, e.g., [10,36]), our goal is to define entropy for the D-S theory that uses Dempster’s product-intersection rule [6] as the combination rule.

Hartley’s Entropy. Suppose $X$ is a discrete random variable with a finite state space $\Omega_X$, whose elements are assumed to be mutually exclusive and exhaustive. Suppose this is all we know about $X$, i.e., we do not know the probability mass function (PMF) of $X$. What is a measure of uncertainty? Hartley [12] defines entropy of $\Omega_X$ as follows:

$$U(\Omega_X) = \log_2(|\Omega_X|),$$

(1)
where \( U(\Omega_X) \) denotes a real-valued measure of uncertainty of \( \Omega_X \), with units of bits. First, notice that \( U(\Omega_X) \) does not depend on the labels attached to the states in \( \Omega_X \), only on the number of states in \( \Omega_X \). Second, Rényi [26] shows that Hartley’s definition in Eq. (1) is characterized by the following three properties.

1. (Additivity) Suppose \( X \) and \( Y \) are random variables with finite state spaces \( \Omega_X \) and \( \Omega_Y \), respectively. The joint state space of \( (X, Y) \) is \( \Omega_X \times \Omega_Y \). Then, \( U(\Omega_X \times \Omega_Y) = U(\Omega_X) + U(\Omega_Y) \).
2. (Monotonicity) If \( |\Omega_{X_1}| > |\Omega_{X_2}| \), then \( U(\Omega_{X_1}) > U(\Omega_{X_2}) \).
3. (Units) If \( |\Omega_X| = 2 \), then \( U(\Omega_X) = 1 \) bit.

**Shannon’s Entropy.** Now suppose we learn of a probability mass function \( P_X \) of \( X \). What is the information content of \( P_X \)? Or alternatively, we can ask: What is the uncertainty in \( P_X \)? Shannon [30] provides an answer to the second question as follows:

\[
H_s(P_X) = \sum_{x \in \Omega_X} P_X(x) \log_2 \left( \frac{1}{P_X(x)} \right),
\]

where \( H_s(P_X) \) is called Shannon’s measure of entropy (uncertainty) in PMF \( P_X \). Shannon’s entropy is characterized (up to a constant) by the following two properties [30].

1. (Monotonicity) If \( P_X \) is an equally likely PMF, then \( H_s(P_X) \) is a monotonically increasing function of \( |\Omega_X| \).
2. (Compound distributions) If a PMF \( P_{X,Y} \) is factored into two PMFs \( P_{X,Y}(x, y) = P_X(x) P_{Y|x}(y) \), then \( H_s(P_{X,Y}) = H_s(P_X) + \sum_{x \in \Omega_X} P_X(x) H_s(P_{Y|x}) \).

The uncertainty prior to learning \( P_X \) was \( U(\Omega_X) \). After learning \( P_X \), it is now \( H_s(P_X) \). Thus, if \( I(P_X) \) denotes information content of \( P_X \), then we have the equality

\[
I(P_X) + H_s(P_X) = U(\Omega_X).
\]

The maximum value of \( H_s(P_X) \) (over the space of all PMFs for \( X \)) is \( \log_2(|\Omega_X|) \), which is attained by the uniform PMF for \( X \), \( P_X(x) = 1/|\Omega_X| \) for all \( x \in \Omega_X \). Thus, \( I(P_X) \geq 0 \), with equality if and only if \( P_X \) is the uniform PMF. At the other extreme, \( H_s(P_X) \geq 0 \), with equality if and only if there exists \( x \in \Omega_X \) such that \( P_X(x) = 1 \). Such a PMF has no uncertainty, and therefore, it must have maximum information. Thus \( I(P_X) \leq U(\Omega_X) \), with equality if and only if there exists \( x \in \Omega_X \) such that \( P_X(x) = 1 \).

**Entropy for the D-S Theory.** In the case of D-S theory of belief functions, if \( m \) is a basic probability assignment (BPA) for \( X \), let \( H(m) \) denote the entropy of BPA \( m \). First, the D-S theory is a generalization of probability theory. The equiprobable PMF is represented by a Bayesian uniform basic probability assignment (BPA) \( m_u \) for \( X \) such that \( m_u(\{x\}) = 1/|\Omega_X| \) for all \( x \in \Omega_X \). So to be consistent with probability theory, we should have \( H(m_u) = \log_2(|\Omega_X|) \). However, such a BPA, \( m_u \) for \( X \), does not have the maximum uncertainty. The
vacuous BPA $\iota_X$ for $X$ such that $\iota_X(\Omega_X) = 1$ has more uncertainty than the equiprobable Bayesian $m_u$. As we cannot imagine a BPA for $X$ that has more uncertainty than $\iota_X$, we assume that $H(\iota_X)$ has the maximum uncertainty. Klir [16] and others argue that a measure of uncertainty can capture a measure of conflict as well as a measure of non-specificity. Assuming that each of these two measures is scaled so that they are each measured on a scale $[0, \log_2(|\Omega_X|)]$, then, $H(\iota_X) = 2\log_2(|\Omega_X|)$. Like in probability theory, we can define a measure of information content of BPA $m$ for $X$ so that the following holds:

$$I(m) + H(m) = 2\log_2(|\Omega_X|),$$

where $I(m)$ denotes the information content of BPA $m$ for $X$. Thus, for the vacuous BPA $\iota_X$ for $X$, we have $I(\iota_X) = 0$, whereas for the Bayesian uniform BPA $m_u$ for $X$, we have $I(m_u) = \log_2(|\Omega_X|)$.

In this paper, we are interested in defining a measure of entropy (uncertainty) of BPAs $m$ for $X$ in the D-S theory of belief functions on the scale $0 \leq H(m) \leq 2\log_2(|\Omega_X|)$, so that $H(m) \leq 2\log_2(|\Omega_X|)$ with equality if and only if $m = \iota_X$, and $H(m) \geq 0$, with equality if and only if $m$ is such that $m(\{x\}) = 1$ for some $x \in \Omega_X$. Also, we require a monotonicity property, a probability consistency property, an additivity property, and a requirement that any definition should be based on semantics consistent with D-S theory. These are discussed in detail in Sect. 2.

**Literature Review.** There is a rich literature on information theoretic measures for the D-S theory of belief functions. Some, e.g., [13,23,32,38], define the information content of BPA $m$ so that $I(\iota_X) = 0$. Some define entropy on the scale $[0, \log_2(|\Omega_X|)]$ so that they define entropy only as a measure of conflict (e.g., [35]), or only as a measure of non-specificity [8]. Some, e.g., [11,15,17,18,21,24,25,37] define entropy as a measure of conflict and non-specificity, but on a scale $[0, \log_2(|\Omega_X|)]$, so that $H(m_u) = H(\iota_X) = \log_2(|\Omega_X|)$. Some, e.g., [1,3,22] define entropy as a measure of conflict and non-specificity on the scale $[0, 2\log_2(|\Omega_X|)]$, but they do so using semantics of belief functions (credal sets of PMFs) that are inconsistent with Dempster’s rule of combination [10,29]. Our definition is the only one that defines entropy as a measure of conflict and non-specificity, on the scale $[0, 2\log_2(|\Omega_X|)]$, using semantics of belief functions that are consistent with Dempster’s combination rule.

## 2 Desired Properties of Entropy of BPAs in the D-S Theory

First, we explain our informal requirement that any definition of entropy for D-S theory should be consistent with the semantics of this theory. Next, we propose five formal properties that a definition of entropy of BPAs in the D-S theory should satisfy. Finally, we compare these properties with those proposed by Klir and Wierman [19] for the same purposes.
Consistency with D-S Theory Semantics Requirement. First, let us stress once more that we are concerned in this paper only with the D-S belief functions theory that includes Dempster’s combination rule as the operation for aggregating knowledge. There are theories of belief functions that use other combination rules. Let $2^\Omega_X$ denote the set of all non-empty subsets of $\Omega_X$. 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 $a \in 2^\Omega_X$ we have:

$$Bel_m(a) = \sum_{b \in 2^\Omega_X : b \subseteq a} m(b) = \min_{P \in \mathcal{P}_m} \sum_{x \in a} P(x). \quad (5)$$

$\mathcal{P}_m$ is referred to as a credal set corresponding to $m$ in the imprecise probability literature (see, e.g., [36]). For such a theory of belief functions, Fagin and Halpern [9] propose a combination rule 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. (5) [10,28]. There are, of course, semantics that are consistent with D-S theory, such as multivalued mappings [6], random codes [28], transferable beliefs [34], and hints [20].

Example 1. 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.

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 $X$ [5]. However, only one of these ways, called the plausibility transform [4], 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 [31], and $\oplus$ is Dempster’s combination rule in D-S theory [4]. [7,15,25] define entropy of $m$ as the Shannon’s entropy of the pignistic transform of $m$. The pignistic transform of $m$ is not compatible with Dempster’s combination rule [4], and therefore, this definition is not consistent with D-S theory semantics. Thus, as per our consistency with D-S theory semantics requirement, any method for defining entropy of $m$ in the D-S theory by first transforming $m$ to a corresponding PMF should use the plausibility transform method. Notice that we are not claiming that a definition of entropy for D-S theory must use the plausibility transform method, only that if one takes the path of first transforming a BPA $m$ to an equivalent PMF and then using Shannon’s entropy of the PMF as the definition of entropy of $m$, then to be compatible with D-S theory semantics, the transformation method used must be the plausibility transform method.
Example 2. Consider a situation where we have vacuous prior knowledge of X with \( \Omega_X = \{x_1, \ldots, x_{70}\} \) and we receive evidence represented as BPA \( m \) for X as follows: \( m(\{x_1\}) = 0.30, m(\{x_2\}) = 0.01, \) and \( m(\{x_2, \ldots, x_{70}\}) = 0.69. \) The pignistic transform of \( m \) [33], denoted by \( \text{Bet}_m \), is as follows: \( \text{Bet}_m(x_1) = 0.30, \text{Bet}_m(x_2) = 0.02, \) and \( \text{Bet}_m(x_3) = \ldots = \text{Bet}_m(x_{70}) = 0.01. \) Thus, as per the pignistic transform, BPA \( m \) is interpreted as evidence where \( x_1 \) is 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 \( \text{Bet}_m \)), \( x_1 \) was 15 times more likely than \( x_2, \) then now that we have evidence \( m \oplus m, \) \( x_1 \) should be \( 15^2 = 225 \) times more likely than \( x_2. \) But \( \text{Bet}_m(x_1) \approx 0.156 \) and \( \text{Bet}_m(x_2) \approx 0.036. \) So according to \( \text{Bet}_m \oplus m, \) \( x_1 \) is only 4.33 more likely than \( x_2. \) Thus, \( \text{Bet}_m \) is not consistent with Dempster’s combination rule.

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 the D-S Theory.** The following list of desired properties of entropy \( H(m_X) \), where \( m_X \) is a BPA for \( X \), is motivated by the properties of Shannon’s entropy of PMFs [30].

Let \( X \) and \( Y \) denote random variables with state spaces \( \Omega_X \) and \( \Omega_Y \), respectively. Let \( m_X \) and \( m_Y \) denote distinct BPAs for \( X \) and \( Y \), respectively. Let \( \iota_X \) and \( \iota_Y \) 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 \( \iota_X \) for \( X \) has the most uncertainty among all BPAs for \( X. \) Such a property is advocated in [3].
3. *(Monotonicity)* If \( |\Omega_X| < |\Omega_Y|, \) then \( H(\iota_X) < H(\iota_Y). \) A similar property is used by Shannon to characterize his definition of entropy of PMFs.
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 \( m_X. \)
5. *(Additivity)* \( H(m_X \oplus m_Y) = H(m_X) + H(m_Y). \) This is a weaker form of the compound property of Shannon’s entropy of a PMF.

Klir and Wierman [19] also describe a set of properties that they believe should be satisfied by any meaningful measure of uncertainty based on intuitive grounds. Two of the properties that they suggest, probability consistency and additivity, are also included in the above list. Our maximum entropy property is not in their list. Two of the properties that they require do not make intuitive sense to us.

First, Klir and Wierman require a property they call “set consistency” as follows: \( H(m) = \log_2(|a|) \) whenever \( m \) is deterministic (i.e., it has only one focal
element) with focal set \( a \). This property would require that \( H(\mathcal{I}_X) = \log_2(|\Omega_X|) \). The probability consistency property requires that for the Bayesian uniform BPA \( m_u \), \( H(m_u) = \log_2(|\Omega_X|) \). Thus, these two requirements entail that \( H(\mathcal{I}_X) = H(m_u) = \log_2(|\Omega_X|) \). We disagree, as there is greater uncertainty in \( \mathcal{I}_X \) than in \( m_u \).

Second, Klir and Wierman require a property they call "range" as follows: For any BPA \( m_X \) for \( X \), \( 0 \leq H(m_X) \leq \log_2(|\Omega_X|) \). The probability consistency property requires that \( H(m_u) = \log_2(|\Omega_X|) \). Also including the range property prevents from having \( H(\mathcal{I}_X) > H(m_u) \). So we do not include it in our list.

Finally, Klir and Wierman require a sub-additivity 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,

\[
H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y})
\]

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.

The most important property that characterizes Shannon’s definition of entropy is the compound property \( H_s(P_{X,Y}) = H_s(P_X \otimes P_{Y|X}) = H_s(P_X) + H_s(P_{Y|X}) \), where \( H_s(P_{Y|X}) = \sum_{x \in \Omega_X} P_X(x) H_s(P_{Y|x}) \). 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 BPA \( m_{Y|X} \) for \( \{X,Y\} \) such that \( m = m^{\downarrow X} \oplus m_{Y|X} \). This cannot be done for all BPA \( m \) for \( \{X,Y\} \) [31]. But, 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 \( m_{Y|X}^{\downarrow X} = \mathcal{I}_X \), and \( m_X \) and \( m_{Y|X} \) are non-conflicting, i.e., the normalization constant in Dempster’s combination rule is 1. Notice that such a constructive 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:

\[
H(m_X \oplus m_{Y|X}) = H(m_X) + H(m_{Y|X}).
\]

However, we are unable to formulate a definition of \( H(m) \) to satisfy such a compound property. So like the sub-additivity 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 of belief functions in the literature are defined on a log scale, and, thus, they all satisfy the additivity property.

3 A New Definition of Entropy of BPAs in the D-S Theory

In this section, we propose a new definition of entropy of BPAs in the D-S theory. The new definition of entropy is based on the plausibility transform of a belief
function to an equivalent probability function. Therefore, we start this section by describing the plausibility transform introduced originally in [4].

**Plausibility Transform 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 [4], who propose the plausibility transform of \( m \) as follows. First consider the plausibility function \( P_{l_m} \) corresponding to \( m \). Next, construct a PMF for \( X \), denoted by \( P_{P_{l_m}} \), by the values of \( P_{l_m} \) for singleton subsets suitably normalized, i.e.,

\[
P_{P_{l_m}}(x) = K^{-1} \cdot P_{l_m}(\{x\}) = K^{-1} \cdot Q_m(\{x\})
\]

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

[4] argues that of the many methods for transforming belief functions to PMFs, the plausibility transform is one that is consistent with Dempster’s rule of combination in the sense that if we have BPs \( m_1, \ldots, m_k \) for \( X \), then \( P_{P_{l_{m_1}} \otimes \cdots \otimes P_{l_{m_k}}} = P_{P_{l_{m_1}}} \otimes \cdots \otimes P_{P_{l_{m_k}}} \), where \( \otimes \) denotes pointwise multiplication followed by normalization (i.e., Bayesian combination [31]). It can be shown that the plausibility transform is the *only* 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 [27].

**Example 3.** Consider a BPA \( m \) for \( X \) as described in Example 2. Then, \( P_{l_m} \) for singleton subsets is as follows: \( P_{l_m}(\{x_1\}) = 0.30 \), \( P_{l_m}(\{x_2\}) = 0.70 \), \( P_{l_m}(\{x_3\}) = \cdots = P_{l_m}(\{x_70\}) = 0.69 \). The plausibility transform of \( m \) is as follows: \( P_{P_{l_m}}(x_1) = 0.3/49.72 \approx 0.0063 \), and \( P_{P_{l_m}}(x_2) = 0.7/49.72 \approx 0.0146 \), and \( P_{P_{l_m}}(x_3) = \cdots = P_{P_{l_m}}(x_70) \approx 0.0144 \). Notice that \( P_{P_{l_m}} \) is quite different from \( BetP_m \). In \( BetP_m \), \( x_1 \) is 15 times more likely than \( x_2 \). In \( P_{P_{l_m}} \), \( x_2 \) is 2.33 times more likely than \( x_1 \). Now consider the scenario where 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 \( P_{P_{l_{m \oplus m}}} \), then as per \( P_{P_{l_{m \oplus m}}} \), \( x_2 \) is 2.33\(^2\) more likely than \( x_1 \). This is a direct consequence of the consistency of the plausibility transform with Dempster’s combination rule.

**A New Definition of Entropy of a BPA.** Suppose \( m \) is a BPA for \( X \). The entropy of \( m \) is defined as follows:

\[
H(m) = \sum_{x \in \Omega_X} P_{l_m}(x) \log_2 \left( \frac{1}{P_{l_m}(x)} \right) + \sum_{a \in 2^{\Omega_X}} m(a) \log_2(|a|).
\]

The first component is Shannon’s entropy of \( P_{l_m} \), and the second component is generalized Hartley’s entropy of \( m \) [8]. Like some of the definitions in the literature, the first component in Eq. (9) 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 \([0, \log_2(|\Omega_X|)]\), and therefore, \( H(m) \) is on the scale \([0, 2\log_2(|\Omega_X|)]\).
Theorem 1. The entropy $H(m)$ of BPA $m$ for $X$ defined in Eq. (9) satisfies the non-negativity, maximum entropy, monotonicity, probability consistency, and additivity properties. It is also consistent with semantics of the D-S theory.

A proof of this theorem can be found in [14] (that can be downloaded from http://pshenoy.faculty.ku.edu/Papers/WP330.pdf). Finally, we provide an example that shows our definition does not satisfy the sub-additivity property.

Example 4. Consider a BPA $m$ for binary-valued variables $\{X, Y\}$: $m(\{x, y\}) = 0.4, m(\{x, \bar{y}\}) = m(\{\bar{x}, y\}) = 0.3, m(\{x, \bar{x}\}) = 0.2$. It is easy to verify that $H(m) \approx 2.35$. The marginal BPA $m^{1X}$ is as follows: $m^{1X}(\{x\}) = 0.2, m^{1X}(\{\bar{x}\}) = 0.6$, and $m^{1X}(\Omega_X) = 0.2$. It is easy to verify that $H(m^{1X}) \approx 1.12$. Similarly, the marginal BPA $m^{1Y}$ is as follows: $m^{1Y}(\{y\}) = 0.4, m^{1Y}(\{\bar{y}\}) = 0.4$, and $m^{1Y}(\Omega_Y) = 0.2$. It is easy to verify that $H(m^{1Y}) \approx 1.20$. Thus, $H(m) \approx 2.35 > H(m^{1X}) + H(m^{1Y}) \approx 1.12 + 1.20 = 2.32$.

The only definition that satisfies the five properties we state plus the sub-additivity property is that due to Maeda and Ichihashi [22], but this definition is based on credal set semantics of a belief function that is inconsistent with Dempster’s combination rule. Whether there exists a definition that satisfies our five properties plus sub-additivity, and that is based on semantics consistent with the basic tenets of D-S theory, remains an open question.

4 Summary and Conclusions

Interpreting Shannon’s entropy of a PMF of a discrete random variable as the amount of uncertainty in the PMF [30], 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 of PMFs, and they are based on our intuition that a vacuous belief function has more uncertainty than an equiprobable Bayesian belief function. 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 [2], based on credal set semantics of belief functions that are not compatible with Dempster’s rule, is not included in our set of requirements.

It would be ideal if we can state the consistency with D-S theory semantics as a formal requirement, but we are unable to do so. In our opinion, the additivity property for the case of two distinct BPAs for disjoint sets of variables does not fully capture consistency with D-S theory semantics. In any case, the definitions of entropy based on credal sets of probability distributions and pignistic transforms are not consistent with Dempster’s combination rule, and therefore, in our perspective, not appropriate for the D-S theory of evidence.

As first suggested by Lamata and Moral [21], we propose a new definition of entropy of BPA as a combination of Shannon’s entropy of an equivalent PMF
that captures the conflict measure of entropy, and Dubois-Prade’s entropy of a BPA that captures the non-specificity (or generalized Hartley) measure of entropy. The equivalent PMF is that obtained by using the plausibility transform [4]. 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.

An open question is whether there exists a definition of entropy of BPA \( m \) in the D-S theory that satisfies the five properties we list in Sect. 2, the sub-additivity 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 sub-additivity property.

Acknowledgements. This article is a short version of [14], which 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 Thierry Denoeux, Marc Pouly, Anne-Laure Jousselme, Joaquín Abellán, and Mark Wierman for their comments on earlier drafts of [14]. We are grateful to two anonymous reviewers of Belief-2016 conference for their comments. We are also grateful to Suzanna Emelio for a careful proof-reading of the text.

References