A tribute to Klir's research on entropy for belief functions

Radim Jiroušek **Václav Kratochvíl**
Institute of Information Theory and Automation,
Czech Academy of Sciences, Prague, Czech Republic
[radim/velorex]@utia.cas.cz

Abstract

This is a short paper to remind us that although Prof. George Klir passed away eight years ago, his ideas are still alive. He was one of the first who had the idea to extend information theory beyond probability theory and to see the potential of fuzzy sets and belief functions. Some of the problems he raised have not yet been satisfactorily solved. In this paper, we take a fresh look at the problem of what properties entropy should have if it is to be considered a measure of information. Based on these requirements, we consider 24 definitions of entropy for belief functions and study how well they satisfy the proposed conditions.

1 Introduction

We met Prof. Klir in the early 1990s when he was lecturing at the Institute of Information Theory and Automation of the Czech Academy of Sciences. He was talking about the open problem of defining entropy in the context of the theory of belief functions. At that time he had about five to seven definitions and was trying to answer several questions. He noticed that belief functions were burdened with different kinds of ambiguity. He recognized three types. The first was nonspecificity, which was related to the size of the focal elements; the larger the focal elements, the less specific the basic assignment. The second type of ambiguity measured internal conflict or dissonance in the belief function, and the third type was confusion. Thus, one question was which types of uncertainty were measured by the entropy functions under study. The main question, however, was what were the most important properties of the concept of entropy. Of course, the answer to this question depends on what we want to measure with this quantity. Is it a measure of uncertainty? If so, how should the measure deal with its different types? Or should entropy measure the information content of a random variable? Although much time has passed since then, no one has satisfactorily answered most of these questions. In this paper, however, we will show that if we restrict the problem to finding an entropy that allows us to define some notions of information theory (in the case of this paper, the measure of mutual information) in the same way as Shannon did, we can deduce which properties are most important. We will also see how Prof. Klir's definitions meet these requirements.

2 Belief functions notation

We expect the reader to be familiar with the theory of belief functions (Shafer, 1976), so we will just briefly recall the main notions of this theory and introduce the notation used in this paper.

Thus, Ω denotes a finite *frame of discernment*. A *basic probability assignment* (BPA) is a function $m : 2^\Omega \rightarrow [0,1]$ such that $\sum_{a \subseteq \Omega} m(a) = 1$, $m(\emptyset) = 0$. We will also consider a two-dimensional case when $\Omega = \Omega_X \times \Omega_Y$. For such $\omega \in \Omega$, its projections (coordinates) will be denoted by $\omega^{\downarrow X}$ and $\omega^{\downarrow Y}$; i.e., $\omega = (\omega^{\downarrow X}, \omega^{\downarrow Y})$. The same symbol will also be used for pro projection of subsets $a \subseteq \Omega$, $a^{\downarrow X} = \{\omega^{\downarrow X} : \omega \in a\}$, and marginalization. $m^{\downarrow X} : 2^{\Omega_X} \rightarrow [0,1]$ is a marginal BPA defined defined on Ω_X by

$$m^{\downarrow X}(a) = \sum_{b \subseteq \Omega : b = a^{\downarrow X}} m(b)$$

for all $a \subseteq \Omega_X$. Analogous symbols will also be used for projections and marginalization concerning the other coordinate Y.

A subset $a \subseteq \Omega$ is said to be a *focal element* of m if $m(a) > 0$. A BPA with only one focal element is said to be *deterministic*, denoted ι_a, where $\iota_a(a) = 1$. Since ι_Ω represents total ignorance, it is called *vacuous*. *Bayesian* BPAs are those BPAs whose focal elements are only singletons.

We will also use the standard alternative representation of a BPA: the *belief function, plausibility function,* and *commonality function*.

$$Bel_m(a) = \sum_{b \subseteq a} m(b); \qquad Pl_m(a) = \sum_{b \subseteq \Omega : b \cap a \neq \emptyset} m(b); \qquad Q_m(a) = \sum_{b \subseteq \Omega : b \supseteq a} m(b).$$

A central concept in Dempster-Shafer's theory is Dempster's combination rule (Shafer, 1976), which combines information from two distinct sources: BPAs m_1 and m_2. The combined BPA $m_1 \oplus m_2$ is computed (for each subset $c \subseteq \Omega$) as follows:

$$(m_1 \oplus m_2)(c) = (1-K)^{-1} \sum_{a \subseteq \Omega} \sum_{b \subseteq \Omega : a \cap b = c} m_1(a) \cdot m_2(b),$$

where $K = \sum_{a \subseteq \Omega} \sum_{b \subseteq \Omega : a \cap b = \emptyset} m_1(a) \cdot m_2(b)$ is usually interpreted as the amount of conflict between m_1 and m_2 (if $K = 1$, then the combination is undefined). This rule is also used when creating a joint BPA for independent variables, i.e., for BPAs m_X and m_Y defined on Ω_X and Ω_Y, respectively. In this case, the focal elements of the combination are only those $c \subseteq \Omega$, for which $c = c^{\downarrow X} \times c^{\downarrow Y}$, and Dempster's rule simplifies to a simple product $(m_X \oplus m_Y)(c) = m_X(c^{\downarrow X}) \cdot m_Y(c^{\downarrow Y})$.

Recall that each BPA m is associated with a set of probability distributions π defined on Ω. A *credal set*, is a set of probability distributions π

$$\mathcal{P}_m = \left\{\pi \text{ defined on } \Omega : \left(\forall a \subseteq \Omega : \pi(a) \geq Bel(a)\right)\right\}.$$

We will consider special probability distributions representing BPA m in specific situations. We will consider so-called *pignistic transform*, *plausibility transform*, and *maximum entropy transform* defined (respectively)

$$\pi_m(x) = \sum_{a \subseteq \Omega : x \in a} \frac{m(a)}{|a|},$$

$$\lambda_m(x) = \frac{Pl_m(x)}{\sum_{y \in \Omega} Pl_m(y)},$$

$$\mu_m = \arg \max_{\pi \in \mathcal{P}_m} \{\mathcal{H}(\pi)\},$$

where $\mathcal{H}(\pi) = -\sum_{x \in \Omega} \pi(x) \log_2 \pi(x)$ is Shannon entropy of a probability distribution π.

3 Requirements on entropy functions

Belief function theory was designed as a generalization of probability theory to surpass the imperfections of the latter. Therefore, most authors introducing the entropy within the theory of belief functions see it as a generalization of Shannon entropy. This is why they require it to equal Shannon entropy for all Bayesian BPAs. This is also why we accept that the belief function's entropy should have the following property.

Probability consistency property. We say that a function H that assigns a real value to each BPA is *consistent with Shannon entropy* if, for all Bayesian BPAs m, the value $H(m)$ is equal to the Shannon entropy of the corresponding probability function, i.e., $H(m) = -\sum_{x \in \Omega} m(\{x\} \log_2 m(\{x\})$.

The properties studied in this paper are deduced from the requirement that the belief function entropy makes it possible to define mutual information in the same way as in probability theory. Recall that in probabilistic information theory

$$\mathcal{MI}_\pi(X;Y) = \mathcal{H}(\pi(X)) - \mathcal{H}(\pi(X|Y)) = \mathcal{H}(\pi(X)) + \mathcal{H}(\pi(Y)) - \mathcal{H}(\pi(X,Y)).$$

Since mutual information is always non-negative and equals zero if and only if the variables are independent, we get that in the ideal case, the belief function entropy should have the following property:

Strict subadditivity property. We say H is *strictly subadditive*, if

$$H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y}) \quad (1)$$

for all BPA m defined on Ω, with the equality in (1) if and only if $m = m^{\downarrow X} \oplus m^{\downarrow Y}$.

Thus, for the strictly subadditive function H, we could apply Shannon's idea of defining *mutual information* by the formula

$$MI_H(m[X;Y]) = H(m^{\downarrow X}) + H(m^{\downarrow Y}) - H(m). \tag{2}$$

Then MI_H would be symmetric, always non-negative, and equal to 0 if and only if X and Y are independent under BPA m, i.e., if $m = m^{\downarrow X} \oplus m^{\downarrow Y}$. However, we admit that we do not know such a strictly subadditive function. None of the entropies listed in Table 1 is strictly subadditive; none can be used as a basis for introducing information theory in the framework of belief functions. For this, one should find another strictly subadditive function or another way to introduce information-theoretic notions that do not follow the Shannon idea.

One reason for developing information theory for belief functions is to transfer the successful model learning algorithms from probability theory to the theory of belief functions. However, for this purpose, one can heuristically use a function that often manifests mutual information's properties. Therefore, we will study weaker properties in the following section. Namely, it is evident that every strictly subadditive function also has the following two properties.

Additivity property. We say H is *additive*, if $H(m_X \oplus m_Y) = H(m_X) + H(m_Y)$ for any pair of one-dimensional BPAs m_X, m_Y defined on Ω_X, Ω_Y, respectively.

Subadditivity property. H is said to be *subadditive*, if $H(m) \leq H(m^{\downarrow X}) + H(m^{\downarrow Y})$ for all BPA m defined on Ω.

The last property we require is based on the intuition that the entropy function H should measure the informativeness of BPAs. Again, using the analogy with probability theory, the more information there is in a BPA, the lower its entropy. For Bayesian BPAs, this requirement is met by the probability consistency property. For others, we assume that BPA m_1 is not less informative than BPA m_2 (assuming both are defined on the same frame of discernment Ω) if $Bel_{m_2} \leq Bel_{m_1}$, which is equivalent to $Pl_{m_1} \leq Pl_{m_2}$, and also to $\mathcal{P}_{m_1} \subseteq \mathcal{P}_{m_2}$. Note that this situation is very general and covers some other specific cases. In a sense, the simplest case is the following. We say that m_1 is a *simple specification* of m_2 if m_1 is created from m_2 by shifting some of its mass (or all of its mass) from some focal element to its subset. More precisely, there exist subsets $a \subset b \subseteq \Omega$ such that $m_1(a) = m_2(a) + \varepsilon$, and $m_1(b) = m_2(b) - \varepsilon$ (all remaining focal elements of m_1 are the copies of the focal elements of m_2). Since we are moving (a part of) the mass from b to its subset, we see directly from the definition of the belief function that[1] $Bel_{m_1} > Bel_{m_2}$. Thus, in this paper, we use the following notion.

[1]Strict inequality $Bel_{m_1} > Bel_{m_2}$ in this paper means that for all $a \subseteq \Omega$, $Bel_{m_1}(a) \geq Bel_{m_2}(a)$, and for at least one a, $Bel_{m_1}(a)$ is strictly greater than $Bel_{m_2}(a)$.

Simple monotonicity property. We say that a function H which assigns a real value to each BPA is *simple monotonic* if, for any simple specification m_1 of m_2, it holds that $H(m_1) < H(m_2)$.

It should be pointed out that several different types of monotonicity of belief functions entropy were studied in the literature (see e.g., monotonicity with respect to the set inclusion (Ramer, 1987), (Abellan and Moral, 1999), and others (Jiroušek and Shenoy, 2018)). Therefore, for the sake of clarity, we will consistently use only the notion of simple monotonicity. It can be shown that it is equivalent to the implication

$$Bel_{m_1} > Bel_{m_2} \Longrightarrow H(m_1) < H(m_2).$$

4 Survey of entropy functions

Without claiming completeness, we consider 24 definitions of entropy-like functions published in the last four decades (see Table 1). This section is devoted to their evaluation with respect to the requirements described in the previous section. As mentioned, none of them is strictly subadditive, so we will consider the introduced weaker properties: their additivity and subadditivity.

From a theoretical point of view, the Maeda-Ichihashi entropy H_I fits our requirements best. It is the only entropy that satisfies all the required properties except strict subadditivity: probability consistency, simple monotonicity, additivity, and subadditivity. A simple but rather singular counterexample has disproved its strict subadditivity. The main drawback of H_I is its computational complexity. Its computation requires the solution of an optimization problem: the search for the maximum entropy transform of the respective BPA. This precludes the application of this entropy not only to machine learning algorithms but also to the computational experiments we have performed to test the behavior of the entropies considered. Note a difference between this entropy and that of Harmanec-Klir H_H. The latter is not simple monotonic since it can take the same value for a BPA as for its simple specification. Nevertheless, H_H is also of extreme computational complexity, and the computation of the Abellán-Moral entropy H_A is even more complex.

We know of no other entropy that would satisfy all four properties that H_I possesses. The property of probability consistency is not a problem. It is possessed by the vast majority of entropies studied, except H_T, H_D, and H_{GP}. Similarly, most of the studied entropies have the additivity property: H_O, H_D, H_N, H_L, H_R, $H_P, H_B, H_H, H_I, H_J, H_Y, H_\lambda, H_S, H_\pi$. The problems arise with the remaining two properties, simple monotonicity, and subadditivity, which are satisfied by only a few entropies.

Both simple monotonicity and subadditivity properties are simultaneously satisfied only by the Dubois-Prade entropy H_D and the Abellán-Moral entropy H_A

Table 1: Definitions of entropy, chronologically ordered

H_O	Hohle (1982)	$H_O(m) = \sum_{a\subseteq\Omega} m(a)\log\left(\frac{1}{Bel_m(a)}\right)$
H_T	Smets (1983)	$H_T(m) = \sum_{a\subseteq\Omega} \log\left(\frac{1}{Q_m(a)}\right)$
H_D	Dubois, Prade (1987)	$H_D(m) = \sum_{a\subseteq\Omega} m(a)\log(\|a\|)$
H_N	Nguyen (1987)	$H_N(m) = \sum_{a\subseteq\Omega} m(a)\log\left(\frac{1}{m(a)}\right)$
H_L	Lamata, Moral (1988)	$H_L(m) = H_Y(m) + H_D(m)$
H_R	Klir, Ramer (1990)	$H_R(m) = H_D(m) + \sum_{a\subseteq\Omega} m(a)\log\left(\frac{1}{1-\sum_{b\subseteq\Omega} m(b)\frac{\|b\setminus a\|}{\|b\|}}\right)$
H_K	Klir (1991)	$H_K(m) = \sum_{a\subseteq\Omega} Bel_m(a)\log(Pl_m(a))$
H_P	Klir, Parviz (1992)	$H_P(m) = H_D(m) + \sum_{a\subseteq\Omega} m(a)\log\left(\frac{1}{1-\sum_{b\subseteq\Omega} m(b)\frac{\|a\setminus b\|}{\|a\|}}\right)$
H_B	Pal et al. (1992, 1993)	$H_B(m) = H_D(m) + H_N(m)$
H_I	Maeda, Ichihashi (1993)	$H_I(m) = H_H(m) + H_D(m) = \mathcal{H}(\mu_m) + H_D(m)$
H_H	Harmanec, Klir (1994)	$H_H(m) = \max_{\pi\in\mathcal{P}_m}\mathcal{H}(\pi) = \mathcal{H}(\mu_m)$
H_{GP}	George, Pal (1996)	$H_P(m) = \sum_{a\subseteq\Omega} m(a)\sum_{b\subseteq\Omega} m(b)\left(1-\frac{\|a\cap b\|}{\|a\cup b\|}\right)$
H_M	Maluf (1997)	$H_M(m) = -\sum_{a\subseteq\Omega} Pl_m(a)\log(Bel_m(a))$
H_A	Abellán, Moral (1999)[2]	$H_A(m) = H_I(m) + \min_{\pi\in\mathcal{B}_m} KL(\pi;\mu_m)$
H_J	Jousselme et al. (2006)	$H_J(m) = \mathcal{H}(\pi_m)$
H_Y	Yager (2008)	$H_Y(m) = \sum_{a\subseteq\Omega} m(a)\log\left(\frac{1}{Pl_m(a)}\right)$
H_G	Deng (2016)	$H_G(m) = H_N(m) + \sum_{a\subseteq\Omega} m(a)\log(2^{\|a\|}-1)$
H_Z	Zhou et al. (2017)	$H_Z(m) = H_G(m) + \frac{\log(e)}{\|\Omega\|}\sum_{a\subseteq\Omega} m(a)*(1-\|a\|)$
H_λ	Jiroušek, Shenoy (2018)	$H_\lambda(m) = \mathcal{H}(\lambda_m) + H_D(m)$
H_{PD}	Pan, Deng (2018)	$H_{PD}(m) = -\sum_{a\subseteq\Omega}\frac{Bel(a)+Pl(a)}{2}\log\left(\frac{Bel(a)+Pl(a)}{2(2^{\|a\|}-1)}\right)$
H_S	Jiroušek, Shenoy (2020)	$H_S(Q_m) = \sum_{a\subseteq\Omega}(-1)^{\|a\|}Q_m(a)\log(Q_m(a))$
H_Q	Qin et al. (2020)	$H_Q(m) = \sum_{a\subseteq\Omega}\frac{\|a\|}{\|\Omega\|}m(a)\log(\|a\|) + H_N(m)$
H_{YD}	Yan, Deng (2020)	$H_{YD}(m) = -\sum_{a\subseteq\Omega} m(a)log\frac{m(a)+Bel(a)}{2(2^{\|a\|}-1)}e^{\frac{\|a\|-1}{\|\Omega\|}}$
H_π	Jiroušek et al. (2022)	$H_\pi(m) = \mathcal{H}(\pi_m) + H_D(m)$

(and the already mentioned Maeda-Ichihashi entropy H_I). However, H_A is dis-

[2] KL denotes the famous Kullback-Leibler divergence of two probability measures defined $KL(\kappa;\pi) = \sum_{x\in\Omega}\kappa(x)\log_2\frac{\kappa(x)}{\pi(x)}$, and $\mathcal{B}_m$ is a borderline of the convex set $\mathcal{P}_m$.

Table 2: Characteristics of entropy

		probability consistency	simple monotonicity	additivity	subadditivity	computational complexity
H_O	Hohle (1982)	■		■		low
H_T	Smets (1983)					high
H_D	Dubois, Prade (1987)		■	■	■	low
H_N	Nguyen (1987)	■		■		low
H_L	Lamata, Moral (1988)	■		■		low
H_R	Klir, Ramer (1990)	■		■		low
H_K	Klir (1991)	■	■			high
H_P	Klir, Parviz (1992)	■		■		low
H_B	Pal et al. (1992, 1993)	■		■		low
H_I	Maeda, Ichihashi (1993)	■	■	■	■	extreme
H_H	Harmanec, Klir (1994)	■		■	■	extreme
H_{GP}	George, Pal (1996)					low
H_M	Maluf (1997)	■				high
H_A	Abellán, Moral (1999)	■	■		■	extreme
H_J	Jousselme et al. (2006)	■		■		low
H_Y	Yager (2008)	■		■		low
H_G	Deng (2016)	■				low
H_Z	Zhou et al. (2017)	■				low
H_λ	Jiroušek, Shenoy (2018)	■	▓	■	░	low
H_{PD}	Pan, Deng (2018)	■				high
H_S	Jiroušek, Shenoy (2020)	■		■		high
H_Q	Qin et al. (2020)	■				low
H_{YD}	Yan, Deng (2020)	■				low
H_π	Jiroušek et al. (2022)	■	▓	■	▓	low

qualified because of its computational complexity, and H_D is zero for all Bayesian BPAs, which disqualifies it as a basis for measuring the dependence of variables in the context of belief functions. Note, however, that it is precisely the H_D entropy whose inclusion makes H_I simple monotonic.

Considering that mutual information is usually used in machine learning only to control heuristic approaches, we can use a criterion that does not have all the

[3] Black stripe - proven property, Dark-grey stripe - in random experiments the property manifested in more than 99 % cases, Light-grey stripe - in random experiments the property manifested in more than 98 % cases.

theoretically required properties for this purpose. Since its computation is repeated many times, we need it to be fast and simple. If we allow its heuristic use, we can use a criterion that sufficiently satisfies the required property. To see if this happens, we randomly generated thousands of BPAs and tested how often the four properties under consideration were satisfied. The results are shown in Table 2, which not only summarizes the properties that were theoretically proven for the entropies considered (black stripes), but also shows that some of the properties were manifested sufficiently often (gray stripes). These results suggest that if we were asked to recommend an entropy for defining mutual information used in efficient machine learning algorithms, we would recommend one based on H_π. Although we know that H_π is neither simply monotonic nor subadditive, the computational experiments showed that the situations in which these negative properties can mislead the model learning process are not frequent.

5 Conclusions

Although we gave a kind of recommendation in the last sentences of the previous paragraph, the paper raised more questions than answers to open questions. The idea behind the Maeda-Ichihashi entropy H_I seems to be fruitful. Although the combination of the Dubois-Prade entropy H_D with neither the pignistic nor the plausibility transform yields an entropy with the required properties, the question remains whether some other transform would serve this purpose as well as the maximum entropy transform. Of course, even the choice of properties may not be optimal. Therefore, we regret that we do not have the opportunity to discuss all such questions with Prof. George J. Klir, who had a deep insight into this problem and the intuition to find answers. If he knew all the results published in the last eight years, he might take a different, more promising path.

Acknowledgement

This study was financially supported by the Czech Science Foundation under Grant No. 19-06569S. A part of this paper previously appeared as Jiroušek and Kratochvíl (2021).

References

Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

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

Joaquin Abellan and Serafin Moral. Completing a total uncertainty measure in the dempster-shafer theory. *International Journal Of General System*, 28(4-5): 299–314, 1999.

Radim Jiroušek and Prakash P Shenoy. A new definition of entropy of belief functions in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 92:49–65, 2018.

Ulrich Hohle. Entropy with respect to plausibility measures. In *Proc. of 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982*, 1982.

Philippe Smets. Information content of an evidence. *International Journal of Man-Machine Studies*, 19(1):33–43, 1983.

Hung T Nguyen. On entropy of random sets and possibility distributions. *The Analysis of Fuzzy Information*, 1:145–156, 1987.

George J Klir. Generalized information theory. *Fuzzy sets and systems*, 40(1): 127–142, 1991.

Nikhil R Pal, James C Bezdek, and Rohan Hemasinha. Uncertainty measures for evidential reasoning i: A review. *International Journal of Approximate Reasoning*, 7(3-4):165–183, 1992.

Nikhil R Pal, James C Bezdek, and Rohan Hemasinha. Uncertainty measures for evidential reasoning ii: A new measure of total uncertainty. *International Journal of Approximate Reasoning*, 8(1):1–16, 1993.

David A. Maluf. Monotonicity of entropy computations in belief functions. *Intelligent Data Analysis*, 1(3):207–213, 1997.

A-L Jousselme, Chunsheng Liu, Dominic Grenier, and Éloi Bossé. Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 36(5):890–903, 2006.

Ronald R Yager. Entropy and specificity in a mathematical theory of evidence. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 291–310. Springer, 2008.

Yong Deng. Deng entropy. *Chaos, Solitons & Fractals*, 91:549–553, 2016.

Deyun Zhou, Yongchuan Tang, and Wen Jiang. A modified belief entropy in dempster-shafer framework. *PloS one*, 12(5):e0176832, 2017.

Miao Qin, Yongchuan Tang, and Junhao Wen. An improved total uncertainty measure in the evidence theory and its application in decision making. *Entropy*, 22(4):487, 2020.

Radim Jiroušek, Václav Kratochvíl, and Prakash P Shenoy. Entropy for evaluation of Dempster-Shafer belief function models. *International Journal of Approximate Reasoning*, 151:164–181, 2022.

Didier Dubois and Henri Prade. Properties of measures of information in evidence and possibility theories. *Fuzzy sets and systems*, 24(2):161–182, 1987.

Maria T Lamata and Serafín Moral. Measures of entropy in the theory of evidence. *International Journal of General System*, 14(4):297–305, 1988.

George J Klir and Arthur Ramer. Uncertainty in the Dempster-Shafer theory: a critical re-examination. *International Journal of General System*, 18(2):155–166, 1990.

George J Klir and Behzad Parviz. A note on the measure of discord. In *Uncertainty in Artificial Intelligence*, pages 138–141. Elsevier, 1992.

Yutaka Maeda and Hidetomo Ichihashi. An uncertainty measure with monotonicity under the random set inclusion. *International Journal of General System*, 21(4): 379–392, 1993.

David Harmanec and George J Klir. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *International Journal of General System*, 22 (4):405–419, 1994.

Thomas George and Nikhil R Pal. Quantification of conflict in dempster-shafer framework: a new approach. *International Journal Of General System*, 24(4): 407–423, 1996.

Lipeng Pan and Yong Deng. A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function. *Entropy*, 20(11):842, 2018.

Radim Jiroušek and Prakash P Shenoy. On properties of a new decomposable entropy of Dempster-Shafer belief functions. *International Journal of Approximate Reasoning*, 119:260–279, 2020.

Hangyu Yan and Yong Deng. An improved belief entropy in evidence theory. *IEEE Access*, 8:57505–57516, 2020.

Radim Jiroušek and Václav Kratochvíl. Approximations of belief functions using compositional models. In *European Conference on Symbolic and Quantitative Approaches with Uncertainty*, pages 354–366. Springer, 2021.