# Generalized Information Theory Based on the Theory of Hints

Marc Pouly

Interdisciplinary Centre for Security, Reliability and Trust,
University of Luxembourg

**Abstract.** The *aggregate uncertainty* is the only known functional for Dempster-Shafer theory that generalizes the Shannon and Hartley measures and satisfies all classical requirements for uncertainty measures, including subadditivity. Although being posed several times in the literature, it is still an open problem whether the aggregate uncertainty is unique under these properties. This paper derives an uncertainty measure based on the theory of hints and shows its equivalence to the *pignistic entropy*. It does not satisfy subadditivity, but the viewpoint of hints uncovers a weaker version of subadditivity. On the other hand, the pignistic entropy has some crucial advantages over the aggregate uncertainty. i.e. explicitness of the formula and sensitivity to changes in evidence. We observe that neither of the two measures captures the full uncertainty of hints and propose an extension of the pignistic entropy called *hints entropy* that satisfies all axiomatic requirements, including subadditivity, while preserving the above advantages over the aggregate uncertainty.

**Keywords:** Generalized Information Theory, Theory of Hints, Dempster-Shafer Theory, Pignistic Entropy, Hints Entropy.

## 1 Introduction

Generalizing the Shannon entropy from probability theory to the various notions of imprecise probabilities, in particular to the Dempster-Shafer theory of evidence [2,19], has been a long discussed issue in the literature. The challenge is to come up with a functional that satisfies all properties one would expect from an uncertainty measure and also reduces to the well-established Shannon and Hartley measures on special cases. Since the early 1980s, various functionals have been proposed for this task, but they were all shown under closer inspection to violate some of the essential properties of uncertainty. In most cases, this was the property of *subadditivity*. We refer to [10] for a historical survey of these unsuccessful attempts. Ultimately, several groups of researchers independently proposed a functional called *aggregate uncertainty* [6,1,15] that satisfies all basic properties and also generalizes the Shannon and Hartley measures. To this day and to the best knowledge of the author, no other functional with the same properties was found. It is further a well-known fact in information theory that the Hartley and Shannon measures are both unique under specific sets of properties.

W. Liu (Ed.): ECSQARU 2011, LNAI 6717, pp. 299–313, 2011.

This naturally raises the question whether the aggregate uncertainty is also fully characterized. Although being posed several times in the literature, for example in [5,10], uniqueness of the aggregate uncertainty remains an open problem. Instead, it was only shown that the aggregate uncertainty satisfies an additional property called *monotone dispensability* [4], from which it follows that the measure is smallest among all functionals that satisfy the classical requirements and monotone dispensability, provided that such functionals exist of course.

A promising candidate for disproving uniqueness of the aggregate uncertainty was the *pignistic entropy* proposed in [9]. This functional operates on mass functions and was claimed to satisfy all classical properties, but a technical mistake in the proof of subadditivity was later found by [11]. Nonetheless, the pignistic entropy has some advantages over the aggregate uncertainty. The practical utility of the latter has often been criticised, because it is defined in terms of the solution to a nonlinear optimisation problem. Although an algorithm for this problem exists [7], it nevertheless prevents us from calculating even very simple examples by hand. The pignistic entropy is expressed by an explicit formula and therefore does not suffer from such a defect. Also, it was observed that the aggregate uncertainty is highly insensitive to changes in evidence [10], which lead to the study of composed functionals to overcome this practical shortcoming [21]. Again, there is no indication for a similar weakness of the pignistic entropy.

This paper analyses the pignistic entropy from the perspective of the *theory of hints* [13,16,14,17], which is a particular approach to Dempster-Shafer theory. Hints are defined as multi-valued mappings between a probability space and the usual frame of discernment. The relationship to mass functions is then established by an equivalence relation between hints that only considers the information with respect to the frame of discernment. Compared to mass functions, hints therefore express more fine-grained information. It will be shown in this paper that the pignistic entropy derives very naturally from this basic model of a hint. Moreover, the additional structure of hints enables a more differentiated view on the lack of subadditivity of the pignistic entropy, and it turns out that the defective argument in [9] can be corrected to prove a weaker version of subadditivity for hints. Also, we show that the pignistic entropy does not satisfy monotone dispensability, but it will be argued that from the viewpoint of hints it has little justification as an axiomatic requirement for uncertainty measures. If we consider the aggregate uncertainty and the pignistic entropy as uncertainty measures in the theory of hints, we notice that both functionals quantify uncertainty with respect to the frame of discernment, ignoring the information on the probability space. We therefore extend the pignistic entropy to take the complete uncertainty of hints into consideration. This leads to a new uncertainty measure called *hints entropy* that also generalizes the Shannon and Hartley measures and further satisfies all classical requirements, including the strong version of subadditivity, while preserving the advantages of the pignistic entropy over the aggregate uncertainty.

The outline of this paper is as follows: Section 2 introduces the theory of hints and establishes the connection to mass functions. Section 3 derives the pignistic entropy, whose properties are analysed in Section 4. Based on the observation that the pignistic entropy does not capture the full uncertainty, this measure is extended to the hints entropy in Section 5. We verify all classical properties for the hints entropy and contrast it with the aggregate uncertainty in Section 6.

## 2 The Theory of Hints

Let $r$ be a countable set of variables. Each variable $X \in r$ has a finite set $\Theta_X$ of possible values. A configuration $\theta$ over a finite set of variables $s \subseteq r$ associates a value from $\Theta_X$ with each variable $X \in s$. We write $\Theta_s$ for the set of all possible configurations over $s$. A hint $\mathcal{H}$ with domain $s$ refers to a question whose true but unknown answer is contained in the set $\Theta_s$ called *frame of discernment*. We further assume a finite set $\Omega$ of possible *interpretations*. Each interpretation restricts the possible answers within $\Theta_s$. If $\omega \in \Omega$ is the correct interpretation, then the correct answer must belong to some non-empty subset $\Gamma(\omega) \subseteq \Theta_s$, where $\Gamma$ is a multi-valued mapping from interpretations $\Omega$ to the powerset $\mathcal{P}(\Theta_s)$. The set $\Gamma(\omega)$ is called the *focal set* of the interpretation $\omega \in \Omega$. However, not all interpretations are equally likely. We therefore assume a probability distribution $p$ that assigns a probability $p(\omega) > 0$ to each interpretation $\omega \in \Omega$. A hint $\mathcal{H}$ with domain $d(\mathcal{H}) = s$ is thus defined as a quadruple $\mathcal{H} = (\Theta_s, \Omega, p, \Gamma)$. Subsequently, we simply write $\Theta$ for the frame of discernment of a hint, if the domain of the latter is not significant. Also, we refer to $(\Omega, p)$ as the probability space with $\sigma$-algebra $\mathcal{P}(\Omega)$, over which the hint $\mathcal{H}$ is defined.

From a hint $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$ we derive a mapping $m : \mathcal{P}(\Theta) \to [0,1]$ by

$$m(A) = \sum_{\omega \in \Omega : \Gamma(\omega) = A} p(\omega) \tag{1}$$

for all $A \subseteq \Theta$. Since $m(\emptyset) = 0$ and $\sum_{A \subseteq \Theta_s} m(A) = 1$ this mapping defines a *mass function* or *basic probability assignment* [19]. The *support* of $m$ is defined as $supp(m) = \{A \subseteq \Theta : m(A) > 0\}$. Observe that several hints can produce the same mass function due to the sum in (1). Given two hints $\mathcal{H}_1$ and $\mathcal{H}_2$, we write $\mathcal{H}_1 \equiv \mathcal{H}_2$ if, and only if, they induce the same mass function. This defines an equivalence relation on the universe of hints $\Phi$ with equivalence classes $[\mathcal{H}] = \{\tilde{\mathcal{H}} \in \Phi : \mathcal{H} \equiv \tilde{\mathcal{H}}\}$. Note also that equivalent hints have the same domain. Given a mass function $m : \mathcal{P}(\Theta) \to [0,1]$, we can always find a canonical hint $\mathcal{H}_c$ in the equivalence class of all hints that induce $m$. If $supp(m) = \{A_1, \ldots, A_n\}$ the canonical hint associated with $m$ is $\mathcal{H}_c = (\Theta, \Omega, p, \Gamma)$ with $\Omega = \{\omega_1, \ldots, \omega_n\}$, $\Gamma(\omega_i) = A_i$ and $p(\omega_i) = m(A_i)$.

We distinguish some important classes of hints. A hint $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$ expresses vacuous information about $\Theta$, if for all $\omega \in \Omega$ we have $\Gamma(\omega) = \Theta$. This represents total ignorance with respect to the question that is represented by the possible answers in $\Theta$. The induced mass function is $m(\Theta) = 1$ and $m(A) = 0$ for all $A \subset \Theta$. If all focal sets are singletons, then the hint is called *precise*. In this case $\Gamma$ represents a random variable and there is full contradiction between interpretations pointing to different singletons. Hence, all discrete random variables can be seen as precise hints. If the focal sets of a hint are disjoint, i.e. if $\omega_1 \neq \omega_2$ implies $\Gamma(\omega_1) \cap \Gamma(\omega_2) = \emptyset$ for all $\omega_1, \omega_2 \in \Omega$, the hint is called *Bayesian*.

Consider two hints $\mathcal{H}_1 = (\Theta, \Omega_1, p_1, \Gamma_1)$ and $\mathcal{H}_2 = (\Theta, \Omega_2, p_2, \Gamma_2)$ on the same frame of discernment $\Theta$. The combined hint $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is defined for interpretations $(\omega_1, \omega_2)$ with $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. Then, $\Gamma(\omega_1, \omega_2) = \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$ is the set of possible answers from $\Theta$, if both interpretations $\omega_1$ and $\omega_2$ hold. However, certain interpretations $(\omega_1, \omega_2)$ may be contradictory, i.e. $\Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) = \emptyset$. We therefore define the combined interpretation set $\Omega$ as the set of all contradiction-free pairs of interpretations, $\Omega = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2 \text{ and } \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) \neq \emptyset\}$. A general rule for computing the probabilities of interpretations in $\Omega$ can be given, if the probability distributions $p_1$ and $p_2$ of the two hints are independent. We then also say that the two hints are *independent*. The probability that two interpretations from independent hints are contradictory is

$$K = \sum_{(\omega_1,\omega_2) \in \Omega_1 \times \Omega_2 : \Gamma(\omega_1,\omega_2)=\emptyset} p(\omega_1) \cdot p(\omega_2).$$

We condition the probability on contradiction-free interpretations and obtain

$$p(\omega_1, \omega_2) = \frac{p_1(\omega_1) \cdot p_2(\omega_2)}{1 - K}$$

for $(\omega_1, \omega_2) \in \Omega$. If the two hints are fully contradictory, i.e. if all intersections of focal sets are empty, then $K = 1$ and the combination is undefined, see [13]. The hint $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, obtained from combining two independent hints $\mathcal{H}_1$ and $\mathcal{H}_2$, is given by $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$. This procedure is called *Dempster's rule of combination* [2]. Further, the combination rule is generalized to dependent hints by replacing the product probability on $\Omega$ by a probability measure that reflects the stochastic dependencies between the interpretations of the two hints. As an important special case of dependent hints, we assume two non-contradictory hints $\mathcal{H}_1 = (\Theta, \Omega, p, \Gamma_1)$ and $\mathcal{H}_2 = (\Theta, \Omega, p, \Gamma_2)$ over the same probability space. Their combination simply yields $\mathcal{H}_1 \otimes \mathcal{H}_2 = (\Theta, \Omega, p, \Gamma_1 \cap \Gamma_2)$. Finally, it is shown in [13] that hints $\mathcal{H}_1$ and $\mathcal{H}_2$ on different domains $d(\mathcal{H}_1) \neq d(\mathcal{H}_2)$ can always be brought to the common domain $d(\mathcal{H}_1) \cup d(\mathcal{H}_2)$ by an operation called *vacuous extension*. This enables the application of the above procedure for combination.

Suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are two independent hints with domains $d(\mathcal{H}_1) = s$ and $d(\mathcal{H}_2) = t$ inducing the mass functions $m_1$ and $m_2$, respectively. We can obtain the mass function of the combined hint $\mathcal{H}_1 \otimes \mathcal{H}_2$ by combining the two mass functions $m_1$ and $m_2$. Following [13] and using the notation of natural join, this combination rule is defined as $m_1 \otimes m_2(\emptyset) = 0$ and

$$m_1 \otimes m_2(A) = \frac{1}{1 - K} \sum_{B \subseteq \Theta_s, C \subseteq \Theta_t : B \bowtie C = A} m_1(B) \cdot m_2(C) \qquad (2)$$

for all other sets $\emptyset \subset A \subseteq \Theta_{s \cup t}$ with

$$K = \sum_{B \subseteq \Theta_s, C \subseteq \Theta_t : B \bowtie C = \emptyset} m_1(B) \cdot m_2(C).$$

Moreover, the combination rule (2) for mass functions becomes particularly simple, if the mass functions are derived from independent hints with disjoint domains $s \cap t = \emptyset$. Such hints and their associated mass functions are subsequently called *non-interactive*. The proof of the following lemma is given in [10].

**Lemma 1.** *It holds that*

$$m_1 \otimes m_2(A) = \begin{cases} m_1(B) \cdot m_2(C) & \text{if } B \times C = A \\ 0 & \text{otherwise,} \end{cases}$$

*if, and only if, $m_1$ and $m_2$ are non-interactive mass functions.*

## 3 The Pignistic Entropy

We now derive an entropy notion for hints based on Shannon's entropy. Given a hint $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$, we know that under the interpretation $\omega \in \Omega$, the true answer belongs to the set $\Gamma(\omega) \subseteq \Theta$, but there is no more precise information about which element in $\Gamma(\omega)$. Hence, under the interpretation $\omega \in \Omega$ the remaining uncertainty about the true answer is most naturally expressed by Hartley's measure, i.e. $H(\mathcal{H}|\omega) = \log |\Gamma(\omega)|$[1]. Likewise, we can see Hartley's measure as a special case of the Shannon entropy applied to a uniform probability distribution. The above claim is therefore equivalent to assuming $p(\theta|\omega) = 1/|\Gamma(\omega)|$ for all $\theta \in \Gamma(\omega)$ and evaluating Shannon's entropy. Hence, the joint probability $p(\theta, \omega)$ for all $\omega \in \Omega$ and $\theta \in \Theta$ is $p(\theta, \omega) = p(\omega)/|\Gamma(\omega)|$ if $\theta \in \Gamma(\omega)$ and $p(\theta, \omega) = 0$ otherwise. From this we derive the marginal distribution

$$p(\theta) = \sum_{\omega \in \Omega : \theta \in \Gamma(\omega)} p(\theta, \omega) = \sum_{\omega \in \Omega : \theta \in \Gamma(\omega)} \frac{p(\omega)}{|\Gamma(\omega)|}, \tag{3}$$

which is called the *pignistic probability distribution* [3,20] associated with $\mathcal{H}$. Hence, we define the entropy of a hint $\mathcal{H}$ with respect to the elements of its frame of discernment $\Theta$ as Shannon's entropy applied to the pignistic distribution.

**Definition 1.** *Let $\mathcal{H}$ be a hint with pignistic probabilities $p(\theta)$ for all $\theta \in \Theta$ as defined in (3). The* pignistic entropy *$H_{\mathcal{H}}(\Theta)$ of $\mathcal{H}$ is defined as*

$$H_{\mathcal{H}}(\Theta) = - \sum_{\theta \in \Theta} p(\theta) \log p(\theta). \tag{4}$$

The next lemma expresses pignistic distributions in terms of mass functions.

[1] We always take logarithms to the base 2.

**Lemma 2.** *If the hints $\mathcal{H}$ induces the mass function $m$, we have for all $\theta \in \Theta$*

$$p(\theta) = \sum_{A \subseteq \Theta : \theta \in A} \frac{m(A)}{|A|}. \quad (5)$$

*Proof.* It follows from equation (1) that

$$\sum_{A \subseteq \Theta : \theta \in A} \frac{m(A)}{|A|} = \sum_{A \subseteq \Theta : \theta \in A} \sum_{\omega \in \Omega : \Gamma(\omega) = A} \frac{p(\omega)}{|A|} = \sum_{\omega \in \Omega : \theta \in \Gamma(\omega)} \frac{p(\omega)}{|\Gamma(\omega)|} = p(\theta).$$

□

In the previous section we called two hints equivalent if they induce the same mass function. It follows from Lemma 2 that equivalent hints share the same pignistic probability distribution and therefore have the same pignistic entropy, i.e. $\mathcal{H}_1 \equiv \mathcal{H}_2$ implies $H_{\mathcal{H}_1}(\Theta) = H_{\mathcal{H}_2}(\Theta)$. This property allows us to apply the pignistic entropy to mass function. We define $H_m(\Theta)$ of a mass function $m$ by the pignistic entropy $H_{\mathcal{H}_c}(\Theta)$ of its canonical hint $\mathcal{H}_c$. It follows that the pignistic entropy coincides with the *ambiguity measure* given in [9].

A meaningful measure for uncertainty in Dempster-Shafer theory must satisfy some basic properties. Such a set of requirements is given in [4,10]. We will next discuss these properties for the pignistic entropy of Definition 1. Since this functional corresponds to the Shannon measure applied to a particular probability distribution, we may directly transfer some properties of the Shannon measure to the pignistic entropy. First, we observe that the pignistic distribution of a mass function $m : \mathcal{P}(\Theta) \to [0,1]$ has exactly $|\Theta|$ values. From the corresponding *range* property of the Shannon measure follows that $0 \leq H_m(\Theta) \leq \log|\Theta|$. Likewise, we know that the Shannon measure is *continuous* in its arguments. According to Lemma 2, each value of the pignistic distribution is given as a finite sum of values from the mass function. Since the limit distributes over finite sums we conclude that the pignistic entropy is also continuous in the values of the mass function. Next, the pignistic entropy must reproduce the Shannon measure on probabilistic evidence. Indeed, a mass function $m$ defines a probability distribution if all focal sets are singletons. The pignistic probability distribution then satisfies $p(\theta) = m(\{\theta\})$ for all $\theta \in \Theta$ which proves the following lemma.

**Lemma 3 (Probability Consistency).** *If a mass function $m$ defines a probability distribution, then the pignistic entropy is equal to Shannon's entropy*

$$H_m(\Theta) = -\sum_{\theta \in \Theta} m(\{\theta\}) \log m(\{\theta\}).$$

This has been observed in [9]. Moreover, it is shown there that the pignistic entropy reproduces Hartley's measure if all mass is given to a single focal set.

**Lemma 4 (Set Consistency).** *If a mass function $m$ has a single focal set $A \subseteq \Theta$, the pignistic entropy is equal to Hartley's measure $H_m(\Theta) = \log|A|$.*

*Proof.* If $m$ has a single focal set $A \subseteq \Theta$, we obtain the canonical hint $\Omega = \{\omega\}$, $\Gamma(\omega) = A$ and $p(\omega) = 1$. This induces the pignistic distribution $p(\theta) = 1/|A|$, if $\theta \in A$, and $p(\theta) = 0$ otherwise. Then the statement follows immediately. □

If a mass function $m$ has a single focal set with exactly two elements, then $H_m(\Theta) = 1$ due to Lemma 4. This shows that the functional is *normalized*. Also, the pignistic entropy is *extensible*, i.e. if a new element $\theta$ is added to the frame of discernment and no mass is given to this element, its pignistic probability is $p(\theta) = 0$ and the pignistic entropy does not change. It is further proved in [9] that the pignistic entropy is additive only for non-interactive mass functions.

**Lemma 5 (Additivity).** *Assume two hints $\mathcal{H}_1$ and $\mathcal{H}_2$ with domain $d(\mathcal{H}_1) = s$ and $d(\mathcal{H}_2) = t$. It holds that $H_{\mathcal{H}_1 \otimes \mathcal{H}_2}(\Theta_{s \cup t}) = H_{\mathcal{H}_1}(\Theta_s) + H_{\mathcal{H}_2}(\Theta_t)$ if, and only if, the two hints are non-interactive.*

The last and most profound property for an uncertainty measure in Dempster-Shafer theory is subadditivity. We start the discussion by first proving a weaker version of the definition given in [4]. The proof is based on Gibbs' theorem: If $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$ are two probability distributions, then

$$-\sum_{i=1}^{n} p_i \log p_i \leq -\sum_{i=1}^{n} p_i \log q_i. \tag{6}$$

**Theorem 1 (Weak Subadditivity).** *If $\mathcal{H}_1$ and $\mathcal{H}_2$ are hints with disjoint domains $d(\mathcal{H}_1) \cap d(\mathcal{H}_2) = \emptyset$ it holds that $H_{\mathcal{H}_1 \otimes \mathcal{H}_2}(\Theta_{s \cup t}) \leq H_{\mathcal{H}_1}(\Theta_s) + H_{\mathcal{H}_2}(\Theta_t)$.*

*Proof.* Let $\mathcal{H}_1 = (\Theta_s, \Omega_1, p_1, \Gamma_1)$ and $\mathcal{H}_2 = (\Theta_t, \Omega_2, p_2, \Gamma_2)$ be two hints with disjoint domains $s \cap t = \emptyset$ and $\mathcal{H}_1 \otimes \mathcal{H}_2 = (\Theta_{s \cup t}, \Omega, p, \Gamma)$ with $\Omega = \Omega_1 \times \Omega_2$ their combination as defined in Section 2. Since the domains are disjoint we have $\Gamma(\omega_1, \omega_2) = \Gamma_1(\omega_1) \times \Gamma_2(\omega_2)$ for all $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$. We further write $\tilde{p}_1$ and $\tilde{p}_2$ for the pignistic distributions of $\mathcal{H}_1$ and $\mathcal{H}_2$ and $\tilde{p}$ for the pignistic distribution of $\mathcal{H}_1 \otimes \mathcal{H}_2$. Since $p$ is the joint probability distribution over $\Omega_1 \times \Omega_2$, it has $p_1$ and $p_2$ as marginal distributions. The pignistic probability $\tilde{p}_1$ of $\mathcal{H}_1$ can thus be written for all $\theta_1 \in \Theta_s$ as

$$\begin{aligned}
\tilde{p}_1(\theta_1) &= \sum_{\omega_1 \in \Omega_1 : \theta_1 \in \Gamma_1(\omega_1)} \frac{p_1(\omega_1)}{|\Gamma_1(\omega_1)|} = \sum_{\substack{\omega_1 \in \Omega_1 : \\ \theta_1 \in \Gamma_1(\omega_1)}} \sum_{\omega_2 \in \Omega_2} \frac{p(\omega_1, \omega_2)}{|\Gamma_1(\omega_1)|} \\
&= \sum_{\substack{(\omega_1, \omega_2) \in \Omega : \\ \theta_1 \in \Gamma_1(\omega_1)}} \frac{p(\omega_1, \omega_2)}{|\Gamma_1(\omega_1)|} = \sum_{\substack{(\omega_1, \omega_2) \in \Omega : \\ \theta_1 \in \Gamma_1(\omega_1)}} \frac{p(\omega_1, \omega_2)}{|\Gamma_1(\omega_1)|} \sum_{\substack{\theta_2 \in \Theta_t : \\ \theta_2 \in \Gamma_2(\omega_2)}} \frac{1}{|\Gamma_2(\omega_2)|} \\
&= \sum_{\theta_2 \in \Theta_t} \sum_{\substack{\omega_1 \in \Omega_1 : \\ \theta_1 \in \Gamma_1(\omega_1)}} \sum_{\substack{\omega_2 \in \Omega_2 : \\ \theta_2 \in \Gamma_2(\omega_2)}} \frac{p(\omega_1, \omega_2)}{|\Gamma_1(\omega_1)||\Gamma_2(\omega_2)|} \\
&= \sum_{\theta_2 \in \Theta_t} \sum_{\substack{(\omega_1, \omega_2) \in \Omega : \\ (\theta_1, \theta_2) \in \Gamma(\omega_1, \omega_2)}} \frac{p(\omega_1, \omega_2)}{|\Gamma(\omega_1, \omega_2)|} = \sum_{\theta_2 \in \Theta_t} \tilde{p}(\theta_1, \theta_2).
\end{aligned}$$

A similar argument shows that $\tilde{p}_2$ is the marginal distribution of $\tilde{p}$ for $t$. It then follows from Gibbs' theorem (6) that

$$\begin{aligned} H_{\mathcal{H}_1 \otimes \mathcal{H}_2}(\Theta_{s\cup t}) &= -\sum_{(\theta_1,\theta_2)\in\Theta_{s\cup t}} \tilde{p}(\theta_1,\theta_2)\log \tilde{p}(\theta_1,\theta_2) \\ &\leq -\sum_{(\theta_1,\theta_2)\in\Theta_{s\cup t}} \tilde{p}(\theta_1,\theta_2)\log\Big[\tilde{p}_1(\theta_1)\tilde{p}_2(\theta_2)\Big] \\ &= -\sum_{(\theta_1,\theta_2)\in\Theta_{s\cup t}} \tilde{p}(\theta_1,\theta_2)\log \tilde{p}_1(\theta_1) - \sum_{(\theta_1,\theta_2)\in\Theta_{s\cup t}} \tilde{p}(\theta_1,\theta_2)\log \tilde{p}_2(\theta_2) \\ &= -\sum_{\theta_2\in\Theta_t}\sum_{\theta_1\in\Theta_s} \tilde{p}(\theta_1,\theta_2)\log \tilde{p}_1(\theta_1) - \sum_{\theta_1\in\Theta_s}\sum_{\theta_2\in\Theta_t} \tilde{p}(\theta_1,\theta_2)\log \tilde{p}_2(\theta_2) \\ &= -\sum_{\theta_1\in\Theta_s} \tilde{p}_1(\theta_1)\log \tilde{p}_1(\theta_1) - \sum_{\theta_2\in\Theta_t} \tilde{p}_2(\theta_2)\log \tilde{p}_2(\theta_2) \\ &= H_{\mathcal{H}_1}(\Theta_s) + H_{\mathcal{H}_2}(\Theta_t). \end{aligned}$$

This proof simplifies considerably if both hints are on the same probability space. □

The stronger version of subadditivity [4] is based on the projection operator for hints. Given a hint $\mathcal{H} = (\Theta_s, \Omega, p, \Gamma)$ with domain $d(\mathcal{H}) = s$, the projection of $\mathcal{H}$ to $t \subseteq s$ is defined as $\mathcal{H}^{\downarrow t} = (\Theta_t, \Omega, p, \Gamma^{\downarrow t})$ with $\Gamma^{\downarrow t}(\omega) = \{\theta^{\downarrow t} : \theta \in \Gamma(\omega)\}$.

**Definition 2 (Strong Subadditivity).** *An uncertainty measure $H : \Phi \rightarrow \mathbb{R}_{\geq 0}$ for hints satisfies strong subadditivity, if for all $\mathcal{H} \in \Phi$ and $d(\mathcal{H}) = s \cup t$ with $s \cap t = \emptyset$ we have $H(\mathcal{H}) \leq H(\mathcal{H}^{\downarrow s}) + H(\mathcal{H}^{\downarrow t})$.*

Weak subadditivity states that the uncertainty of a combined hint is not larger than the sum of the uncertainties of the two individual hints. This setting corresponds to additivity in Lemma 5 without the additional assumption of independence. In contrast, strong subadditivity assumes a single hint and requires that its uncertainty is not larger than the sum of the uncertainties of its projections to disjoint subdomains. Indeed, from weak subadditivity only follows

$$H_{\mathcal{H}^{\downarrow s} \otimes \mathcal{H}^{\downarrow t}}(\Theta_{s\cup t}) \leq H_{\mathcal{H}^{\downarrow s}}(\Theta_s) + H_{\mathcal{H}^{\downarrow t}}(\Theta_t), \tag{7}$$

but not that $H_{\mathcal{H}}(\Theta_{s\cup t}) \leq H_{\mathcal{H}^{\downarrow s} \otimes \mathcal{H}^{\downarrow t}}(\Theta_{s\cup t})$. Moreover, [11] gave a simple counter-example for strong subadditivity of the pignistic entropy.

**Lemma 6.** *The pignistic entropy does not satisfy strong subadditivity.*

The proof of weak subadditivity for the pignistic entropy exploits that the marginals of the pignistic distribution of $\mathcal{H}_1 \otimes \mathcal{H}_2$ correspond to the pignistic distributions of $\mathcal{H}_1$ and $\mathcal{H}_2$. According to (7) this holds also between $\mathcal{H}^{\downarrow s} \otimes \mathcal{H}^{\downarrow t}$ and its factors, but generally not between $\mathcal{H}$ and its projections $\mathcal{H}^{\downarrow s}$ and $\mathcal{H}^{\downarrow t}$ as shown by the counter-example in [11]. This crucial observation lead to the wrong proof of strong subadditivity for the pignistic entropy in [9].

## 4 Pignistic Entropy and Aggregate Uncertainty

It is well-known in Dempster-Shafer theory that to every mass function we can associate a belief function via the so-called Moebius transform. This mapping from mass functions to belief functions is bijective [19,12]. The *aggregate uncertainty AU* measures the uncertainty of a mass function $m$ in terms of its associated belief function $bel$,

$$AU(bel) = \max_{P_{bel}} \left[ -\sum_{\theta \in \Theta} p(\theta) \log p(\theta) \right]. \quad (8)$$

Here, the maximum is taken over all probability distributions that dominate the belief function $bel$. It is shown in [6,10] that the aggregate uncertainty satisfies the properties of probability and set consistency, range, subadditivity and additivity. The question, whether the aggregate uncertainty is the only mapping from mass functions to real numbers that satisfies these properties, is posed as an open problem in [5,10]. Since equivalent hints induce the same mass function, the aggregate uncertainty can as well be considered as an uncertainty measure for hints with respect to the elements in $\Theta$. It was shown in Section 3 how the viewpoint of hints gives rise to another uncertainty measure in a very natural way. But this measure, called pignistic entropy, suffers from the clear defect that it does not satisfy strong subadditivity. On the other hand, it has been observed in [10] that the aggregate uncertainty is sometimes too insensitive with respect to changes in evidence, which is a severe practical shortcoming.

*Example 1.* Let $\Theta = \{\theta_1, \theta_2\}$ and assume a mass function $m$ defined as $m(\{\theta_1\}) = \alpha$, $m(\{\theta_2\}) = \beta$ and $m(\Theta) = 1 - \alpha - \beta$ with $\alpha, \beta \geq 0$ and $\alpha + \beta \leq 1$. Example 6.14 in [10] gives the associated belief function $bel(\{\theta_1\}) = \alpha$, $bel(\{\theta_2\}) = \beta$ and $bel(\Theta) = 1$ and finally obtains $AU(bel) = 1$ for $0 \leq \alpha, \beta \leq 0.5$. The aggregate uncertainty is therefore insensitive to changes of $\alpha$ and $\beta$ in the interval $[0, 0.5]$.

There is no indication that the pignistic entropy suffers from a similar lack of insensitivity [9,11]. Indeed, the following example shows that we obtain a different pignistic distribution for different values of $\alpha$ and $\beta$ in Example 1.

*Example 2.* The canonical hint $\mathcal{H}_c = (\Theta, \Omega, p, \Gamma)$ for the mass function of Example 1 has $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\Gamma(\omega_1) = \{\theta_1\}$, $\Gamma(\omega_2) = \{\theta_2\}$ and $\Gamma(\omega_3) = \Theta$ with $p(\omega_1) = \alpha$, $p(\omega_2) = \beta$ and $p(\omega_3) = 1 - \alpha - \beta$. We then obtain for the pignistic probability distribution $p(\theta_1) = \alpha + 0.5(1 - \alpha - \beta)$ and $p(\theta_2) = \beta + 0.5(1 - \alpha - \beta)$. This gives a pignistic entropy of $H(m) = H(\mathcal{H}_c) = 1$ if, and only if, $\alpha = \beta$.

To overcome the insensitivity problem of the aggregate uncertainty, another measure $GS$ is proposed in [21,10], defined as the difference between the aggregate uncertainty and the *generalized Hartley measure* [8] $GH$

$$GS(bel) = AU(bel) - GH(m), \text{ where } GH(m) = \sum_{A \subseteq \Omega} m(A) \log |A|. \quad (9)$$

It is shown in [10] that for Bayesian mass functions $m$ we have

$$GS(bel) = -\sum_{A \subseteq \Theta} m(A) \log m(A), \tag{10}$$

and it then follows from equation (9) that

$$AU(bel) = -\sum_{A \subseteq \Theta} m(A) \log \left[ \frac{m(A)}{|A|} \right]. \tag{11}$$

The pignistic entropy and the aggregate uncertainty are clearly different, since they disagree on strong subadditivity. On the other hand, both measures satisfy set and probability consistency and therefore give equal results for mass functions with a single focal set and hints that define a probability distribution. Moreover, the following lemma shows that the same holds for Bayesian hints.

**Lemma 7.** *If $\mathcal{H}$ denotes a Bayesian hint, $m$ its induced mass function and $bel$ the associated belief function, we have $H_{\mathcal{H}}(\Theta) = AU(bel)$.*

*Proof.* Since focal sets are disjoint we have $p(\omega) = m(A)$ if $\Gamma(\omega) = A$. Every $\theta \in \Theta$ is contained in exactly one focal set, which implies that $p(\theta) = p(\omega)/|\Gamma(\omega)|$ for $\omega \in \Omega$ with $\theta \in \Gamma(\omega)$. Moreover, if $\theta_1$ and $\theta_2$ are contained in the same focal set $\Gamma(\omega)$ for some $\omega \in \Omega$ then $p(\theta_1) = p(\theta_2)$ and therefore

$$\begin{aligned} H_{\mathcal{H}}(\Theta) &= -\sum_{\theta \in \Theta} p(\theta) \log p(\theta) = -\sum_{\omega \in \Omega} |\Gamma(\omega)| \cdot \frac{p(\omega)}{|\Gamma(\omega)|} \log \frac{p(\omega)}{|\Gamma(\omega)|} \\ &= -\sum_{\omega \in \Omega} p(\omega) \log \frac{p(\omega)}{|\Gamma(\omega)|} = -\sum_{A \subseteq \Theta} m(A) \log \left[ \frac{m(A)}{|A|} \right] = AU(bel). \end{aligned}$$

□

Another disadvantage of the aggregate uncertainty is that its evaluation requires the solution to a nonlinear optimization problem, see (8). Although an algorithm exists for this task, the lack of a closed form is nonetheless inconvenient. In contrast, the pignistic entropy only requires to compute the pignistic probability distribution, which essentially corresponds to the evaluation of a probability tree.

If we compare the aggregate uncertainty and the pignistic entropy, we observe that both functionals apply the Shannon entropy to a particular probability distribution, i.e. the dominating distribution in case of the aggregate uncertainty and the pignistic distribution in case of the pignistic entropy. Let us consider this commonality in more detail. If $\Theta$ denotes a frame of discernment, we write $P$ for the set of all probability distributions over $\Theta$ and $\Psi$ for the set of all mass functions over $\Theta$. Next, assume a functional $h : \Psi \rightarrow [0, \log |\Theta|]$ that satisfies the property of (range and) probability consistency. Since $P$ is a convex set and the Shannon entropy $S : P \rightarrow [0, \log |\Theta|]$ is continuous, it follows from the generalized intermediate value theorem that for every mass function $m \in \Psi$ there exists a probability distribution $p \in P$ such that $h(m) = S(p)$. On the other hand, a probability distribution $p \in P$ can always be considered as a precise mass function $m_p$ and probability consistency ensures that $h(m_p) = S(p)$. Hence, it follows that for every mass function $m \in \Psi$ there exists a precise mass function $m_p$ such that $h(m) = h(m_p) = S(p)$. This proves that the observed property does not only hold for the aggregate uncertainty and the pignistic entropy but for all functionals $h : \Psi \rightarrow [0, \log|\Theta|]$ that satisfy probability consistency.

**Lemma 8.** *If $h : \Psi \rightarrow [0, \log|\Theta|]$ satisfies probability consistency, it holds for all $m \in \Psi$ that $h(m) = S(\pi(m))$ for some mapping $\pi : \Psi \rightarrow P$.*

Finally, [4] proves that the aggregate uncertainty satisfies a property called *monotone dispensability*, which says that the uncertainty should not decrease after transferring part of a focal set's mass to a superset.

**Lemma 9.** *Let $m : \mathcal{P}(\Theta) \rightarrow [0, 1]$ be a mass function and $A \subseteq \Theta$ a focal set. For a superset $B \supseteq A$ and $0 \leq \alpha \leq 1$ we derive a mass function $m'$ by $m'(A) = \alpha m(A)$, $m'(B) = m(B) + (1 - \alpha)m(A)$ and $m'(C) = m(C)$ for all $C \subseteq \Theta$ with $C \neq A$ and $C \neq B$. If bel and bel′ denote the associated belief functions, we have $AU(bel) \leq AU(bel')$.*

This property is not satisfied by the pignistic entropy:

*Example 3.* Let $\Theta = \{\theta_1, \theta_2\}$ and assume a mass function $m$ defined as $m(\{\theta_1\}) = m(\{\theta_2\}) = 0.5$. We obtain for the pignistic distribution $p(\theta_1) = p(\theta_2) = 0.5$. Next, choose $A = \{\theta_1\}$, $B = \{\theta_1, \theta_2\}$ and $\alpha = 0.5$. We obtain $m'(\{\theta_1\}) = 0.25$, $m'(\{\theta_2\}) = 0.5$ and $m'(\{\theta_1, \theta_2\}) = 0.25$. This gives a pignistic distribution of $p(\theta_1) = 3/8$, $p(\theta_2) = 5/8$ and therefore $H_m(\Theta) > H_{m'}(\Theta)$.

The Shannon entropy is maximal only for uniform distributions. Lemma 2 shows that the pignistic probability distribution strongly depends on the values of the mass function such that every modification of the latter in Example 3 will lead to a different pignistic distribution and therefore to a smaller pignistic entropy. This contradicts the statement of monotone dispensability. Likewise, it is equally simple to come up with an example for $H_m(\Theta) < H_{m'}(\Theta)$. Take for instance $m(\{\theta_1\}) = 0.1$ and $m(\{\theta_2\}) = 0.9$ in Example 3. This non-monotonic behaviour is semantically well-justified. If mass is moved from the focal set $A = \{\theta_1\}$ to a superset $B = \Theta$, it decreases the value for $p(\theta_1)$ and increases the value for $p(\theta_2)$. But the difference $|p(\theta_1) - p(\theta_2)|$ can either increase or decrease, which must affect uncertainty in different ways. Accordingly, the Shannon entropy is small if the difference is large and large if the difference is small. In other words, the absence of this property in case of the pignistic entropy again alludes to the sensitivity of this functional with respect to changes in evidence. For particular applications monotone dispensability may be a desirable property but we refrain from considering it as an axiomatic requirement for uncertainty measures.

## 5 The Hints Entropy

The foregoing section showed that the pignistic entropy satisfies all required properties for an uncertainty measure in Dempster-Shafer theory except strong subadditivity. However, given a hint $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$ the pignistic entropy measures uncertainty with respect to $\Theta$, but the hint $\mathcal{H}$ clearly contains information on both sets $\Omega$ and $\Theta$. We therefore consider the join entropy of $\mathcal{H}$ given as

$$H_{\mathcal{H}}(\Omega, \Theta) = -\sum_{\omega \in \Omega} \sum_{\theta \in \Theta} p(\omega, \theta) \log p(\omega, \theta) = -\sum_{\omega \in \Omega} \sum_{\theta \in \Gamma(\omega)} \frac{p(\omega)}{|\Gamma(\omega)|} \log \frac{p(\omega)}{|\Gamma(\omega)|}$$
$$= -\sum_{\omega \in \Omega} p(\omega) \log p(\omega) + \sum_{\omega \in \Omega} p(\omega) \log |\Gamma(\omega)|.$$

Observe that the second summand, which is also the conditional Shannon entropy of $\Theta$ given $\Omega$, corresponds the generalized Hartley measure of equation (9). Indeed, it follows from equation (1) that

$$GH(\mathcal{H}) = \sum_{\omega \in \Omega} p(\omega) \log |\Gamma(\omega)| = \sum_{A \subseteq \Omega} m(A) \log |A|. \tag{12}$$

**Definition 3.** *The* hints entropy *of $\mathcal{H} = (\Theta, \Omega, p, \Gamma)$ is defined as*

$$H_{\mathcal{H}}(\Omega, \Theta) = H_{\mathcal{H}}(\Omega) + GH(\mathcal{H}). \tag{13}$$

Due to equation (12) the generalized Hartley measure is completely determined by the mass function. The value of the generalized Hartley measure is therefore the same for equivalent hints, i.e. $\mathcal{H}_1 \equiv \mathcal{H}_2$ implies $GH(\mathcal{H}_1) = GH(\mathcal{H}_2)$. However, equivalent hints may differ in the probabilities assigned to interpretations, which thus leads to different values of $H_{\mathcal{H}_1}(\Omega_1)$ and $H_{\mathcal{H}_2}(\Omega_2)$. It thus follows from (13) that equivalent hints do not necessarily share the same hints entropy, naturally because the equivalence relations was defined with respect to the frame of discernment only. This again confirms that hints contain more information than their associated mass functions. We next investigate uncertainty related properties of the hints entropy. First, we observe that the generalized Hartley measure is zero if all focal sets are singletons. This proves the following statement.

**Lemma 10 (Probability Consistency).** *If a hint defines a probability distribution, then the hints entropy is equal to Shannon's entropy $H_{\mathcal{H}}(\Omega, \Theta) = H_{\mathcal{H}}(\Omega)$.*

Conversely, if $\Omega$ is a singleton, we have $H_{\mathcal{H}}(\Omega) = 0$ and $GH(\mathcal{H}) = \log |\Gamma(\omega)|$. This proves the following more restrictive version of set consistency.

**Lemma 11 (Set Consistency).** *If a hint has a single interpretation $\Omega = \{\omega\}$ the hints entropy is equal to Hartley's measure $H_{\mathcal{H}}(\Omega, \Theta) = \log |\Gamma(\omega)|$.*

**Lemma 12 (Additivity).** *Given two hints $\mathcal{H}_1 = (\Theta_s, \Omega_1, p_1, \Gamma_1)$ and $\mathcal{H}_2 = (\Theta_t, \Omega_2, p_2, \Gamma_2)$ it holds that $H_{\mathcal{H}_1 \otimes \mathcal{H}_2}(\Omega, \Theta_s \times \Theta_t) = H_{\mathcal{H}_1}(\Omega_1, \Theta_s) + H_{\mathcal{H}_2}(\Omega_2, \Theta_t)$ if, and only if, the two hints are non-interactive.*

*Proof.* According to [10] the generalized Hartley measure is additive only for non-interactive hints. Likewise, the Shannon entropy is additive only for independent distributions. This proves additivity of the hints entropy. □

**Theorem 2 (Strong Subadditivity).** *For $\mathcal{H} = (\Theta_{s \cup t}, \Omega, p, \Gamma)$ with $s \cap t = \emptyset$ we have $H_{\mathcal{H}}(\Omega, \Theta_{s \cup t}) \leq H_{\mathcal{H}^{\downarrow s}}(\Omega, \Theta_s) + H_{\mathcal{H}^{\downarrow t}}(\Omega, \Theta_t)$.*

*Proof.* According to [10] the generalized Hartley measure is subadditive, hence

$$
\begin{aligned}
H_{\mathcal{H}}(\Omega, \Theta_{s \cup t}) &\leq 2H_{\mathcal{H}}(\Omega) + GH(\mathcal{H}) \leq 2H_{\mathcal{H}}(\Omega) + H_{\mathcal{H}^{\downarrow s}}(\Omega, \Theta_s) + H_{\mathcal{H}^{\downarrow t}}(\Omega, \Theta_t) \\
&= H_{\mathcal{H}^{\downarrow s}}(\Omega, \Theta_s) + H_{\mathcal{H}^{\downarrow t}}(\Omega, \Theta_t).
\end{aligned}
$$

It then follows from the theory of valuation algebras [12] and strong subadditivity that the hints entropy satisfies weak subadditivity as well.

## 6 Hints Entropy and Aggregate Uncertainty

If we compare the aggregate uncertainty, as an uncertainty measure for hints, and the hints entropy of Definition 3, we observe that both measures satisfy the properties of probability and set consistency, additivity and strong subadditivity. However, we refrain from saying that the hints entropy completely disproves the uniqueness claim for the aggregate uncertainty, because the two measures differ in their range property, i.e. the hints entropy measures uncertainty with respect to $\Omega$ and $\Theta$, whereas the aggregate uncertainty only focusses on $\Theta$. This observation leads to a very interesting insight. The pignistic probability distribution was derived in equation (3) as the marginal distribution of the joint probability over $\Omega$ and $\Theta$. It therefore holds that

$$
H_{\mathcal{H}}(\Omega, \Theta) = H_{\mathcal{H}}(\Theta) + H_{\mathcal{H}}(\Omega|\Theta), \tag{14}
$$

where $H_{\mathcal{H}}(\Omega|\Theta)$ denotes the conditional entropy of $\Omega$ given $\Theta$. In the hint model, interpretations $\omega \in \Omega$ give information with respect to the elements in $\Theta$ by restricting the set of possible answers to a subset $\Gamma(\omega) \subseteq \Theta$. Conversely, also the elements $\theta \in \Theta$ give information about the correct interpretation in $\Omega$ via the inverse mapping. Given $\theta \in \Theta$, the remaining uncertainty about the elements in $\Omega$ is measured by the conditional entropy $H_{\mathcal{H}}(\Omega|\theta)$ with $H_{\mathcal{H}}(\Omega|\Theta)$ as expected value. The transformation from hints to mass functions looses this information, see equation (1), which also prevents the uncertainty measures for mass functions to take this information into account. This is confirmed by equation (14), showing that the hints and pignistic entropy differ in exactly $H_{\mathcal{H}}(\Omega|\Theta)$. Moreover, we observed that strong subadditivity only holds for the hints entropy $H_{\mathcal{H}}(\Omega, \Theta)$ but not for the pignistic entropy $H_{\mathcal{H}}(\Theta)$. We may therefore conclude that ignoring the additional information brought by the inverse mapping destroys the property of strong subadditivity. Bayesian hints have disjoint focal sets, which intuitively means that the same information is contained in $\Gamma$ and its inverse. Indeed, this is confirmed by the following theorem extending Lemma 7.

**Theorem 3.** *If $\mathcal{H}$ denotes a Bayesian hint and bel the associated belief function we have $H_{\mathcal{H}}(\Omega, \Theta) = H_{\mathcal{H}}(\Theta) = AU(bel)$.*

*Proof.* It remains to show that $H_{\mathcal{H}}(\Omega, \Theta) = AU(bel)$ for Bayesian hints. Since focal sets are disjoint we have $p(\omega) = m(A)$ if $\Gamma(\omega) = A$. Using (11) we obtain

$$H_{\mathcal{H}}(\Omega, \Theta) = H_{\mathcal{H}}(\Omega) + GH(\mathcal{H}) = -\sum_{A \subseteq \Theta} m(A) \log \left[ \frac{m(A)}{|A|} \right] = AU(bel). \qquad \square$$

Finally, an equally simple counter-example as in Example 3 shows that the hints entropy does not satisfy monotone dispensability.

## 7 Conclusion

This paper derives the pignistic entropy for Dempster-Shafer theory based on the theory of hints and proves its equivalence to the known measure for ambiguity. The functional agrees with the Shannon and Hartley measure on corresponding cases and satisfies all classical requirements, which are generally imposed on an uncertainty measure, except subadditivity. But the viewpoint of hints allows us to prove a weaker form of subadditivity. In contrast, the aggregate uncertainty is the only known functional that satisfies all properties including subadditivity, and uniqueness of this measure under the classical properties is stated as an open problem in the literature. Despite the lack of strong subadditivity, the pignistic entropy has some crucial advantages over the aggregate uncertainty, most notably explicitness of the formula and sensitivity with respect to changes in evidence. However, we observed that both uncertainty measures do not capture all information contained in the hint model and therefore extend the pignistic entropy to the hints entropy that takes the total information of a hint into account. This new measure still generalizes the Shannon and Hartley measures and further satisfies all classical requirements, including strong subadditivity, while preserving the advantages of the pignistic entropy over the aggregate uncertainty.

## Acknowledgement

The author would like to thank Juerg Kohlas and Peter Y. A. Ryan for the fruitful discussions on the pignistic and hints entropy and the relevance of monotone dispensability. Early considerations about uncertainty of hints were also made in [18]. Also, I thank Wojtek Jamroga for his valuable comments on subadditivity and the anonymous reviewers for their important suggestions.

## References

1. Chau, C., Lingras, P., Wong, S.: Upper and lower entropies of belief functions using compatible probability functions. In: Komorowski, J., Raś, Z.W. (eds.) ISMIS 1993. LNCS, vol. 689, pp. 306–315. Springer, Heidelberg (1993)
2. Dempster, A.P.: A generalization of bayesian inference. J. Royal Stat. Soc. B 30, 205–247 (1968)

3. Dubois, D., Prade, H.: A note on measures of specificity for fuzzy sets. Int. J. Gen. Systems 10(4), 279–283 (1985)
4. Harmanec, D.: Toward a characterization of uncertainty measure for the dempster-shafer theory. In: UAI 1995: Proc. of the 11th Conference Annual Conference on Uncertainty in Artificial Intelligence, pp. 255–261 (1995)
5. Harmanec, D.: Measure of uncertainty and information. In: Imprecise Probability Project (1999)
6. Harmanec, D., Klir, G.: Measuring total uncertainty in dempster-shafer theory: a novel approach. Int. J. Gen. Systems 22(4), 405–419 (1994)
7. Harmanec, D., Resconi, G., Klir, G.J., Pan, Y.: On the computation of uncertainty measure in the dempster-shafer theory. Int. J. Gen. Systems 25(2), 153 (1996)
8. Higashi, M., Klir, G.J.: Measures of uncertainty and information based on possibility distributions. Int. J. Gen. Systems 9(1), 43–58 (1982)
9. Jousselme, A.-L., Liu, C., Grenier, D., Bossé, E.: Measuring ambiguity in the evidence theory. IEEE Trans. on Systems, Man, and Cybernetics, Part A 36(5), 890–903 (2006)
10. Klir, G.J.: Uncertainty and Information: Foundations of Generalized Information Theory. John Wiley & Sons, Inc., Binghamton University (2005)
11. Klir, G.J., Lewis, H.W.: Remarks on "measuring ambiguity in the evidence theory". IEEE Trans. on Systems, Man, and Cybernetics, Part A 38(4), 995–999 (2008)
12. Kohlas, J.: Information Algebras: Generic Structures for Inference. Springer, Heidelberg (2003)
13. Kohlas, J., Monney, P.-A.: A Mathematical Theory of Hints. An Approach to the Dempster-Shafer Theory of Evidence. LNEMS. Springer, Heidelberg (1995)
14. Kohlas, J., Monney, P.-A.: Statistical Information. Assumption-Based Statistical Inference. Sigma Series in Stochastics, vol. 3. Heldermann (2008)
15. Maeda, Y., Ichihashi, H.: An uncertainty with monotonicity under the random set inclusion. Int. J. Gen. Systems 21(4), 379 (1993)
16. Monney, P.-A.: A Mathematical Theory of Arguments for Statistical Evidence. Contributions to Statistics. Physica-Verlag, Heidelberg (2003)
17. Pouly, M., Kohlas, J.: Generic Inference - A Unifying Theory for Automated Reasoning. John Wiley & Sons, Inc., Chichester (2011)
18. Schneuwly, C.: Information - eine diskussion. Term Paper, University of Fribourg (1999)
19. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
20. Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. 66(2), 191–234 (1994)
21. Smith, R.: Generalized Information Theory: Resolving some old Questions and opening some new ones. PhD thesis, University of Binghamton (2000)