

Scoring Rules for Belief Functions and Imprecise Probabilities: A Comparison

Esther Anna Corsi¹[0000-0002-4586-6049], Tommaso
Flaminio²[0000-0002-9180-7808], and Hykel Hosni¹[0000-0003-3695-6752]

¹ Department of Philosophy, University of Milan, Via Festa del Perdono 7, 20122
Milano, Italy. {esther.corsi, hykel.hosni}@unimi.it

² Artificial Intelligence Research Institute (IIIA – CSIC), Campus UAB, Bellaterra
08193, Spain. tommaso@iiaa.csic.es

Abstract. This paper investigates de Finetti’s *coherence* as an operational foundation for a wide range of non-additive uncertainty measures and focuses, in particular, on Belief functions and Lower probabilities. In a companion paper, we identify a number of non-limiting circumstances under which Dutch Book criteria for Belief functions and Lower probability are undistinguishable, which is surprising given that Lower probabilities are known to exist which do not satisfy the axioms of Belief functions. The main contribution of this paper consists in putting forward a comparison between a criterion based on the Brier scoring rule for Belief Functions and the scoring rule introduced in 2012 by Seidenfeld, Schervish and Kadane for Imprecise probabilities. Through this comparison we show that scoring rules allow us to distinguish coherence-wise between Belief functions and Imprecise probabilities.

Keywords: Scoring Rules · Belief functions · Lower probabilities · Imprecise probabilities · Coherence.

1 Introduction

In a companion paper [4], we observe that in any finite boolean algebra with at least three atoms there exist books defined over rich-enough sets of events that do not distinguish assignments extendible to Belief functions from those extendible to Lower probabilities. Given that all Belief functions are Lower probabilities, but the converse does not hold, our previous finding is somewhat puzzling. In this work we extend the comparison between those two well-known non-additive measures of uncertainty to the alternative, but logically equivalent (in probability), criterion of coherence based on the Brier scoring rule.

To do so, we introduce a new scoring rule in Definition 6 and (i) show that it characterises assignments that are extendible to Belief functions and (ii) compare it to the scoring rule for imprecise probabilities introduced by Seidenfeld et al in [20]. This latter allows us to make a distinction between coherent Belief functions on the one hand, and coherent Lower probabilities on the other hand.

The paper is organised as follows. In Section 2, we recall the Dutch Book and the proper scoring rule coherence criteria for probability functions. Early generalisations of probabilistic coherence have been proposed in [22, 15]. In Section 3, we give the required background on uncertainty measures, a geometric view on coherence and extendibility, and a brief outline of the key results obtained in [4]. In Section 4, we introduce a scoring rule that characterises assignments extendible to Belief functions. In Section 5, we relate our work to [20], draw some conclusions and outline the research questions opened up by the present investigation.

2 Dutch Books and Proper Scoring Rules

Bruno de Finetti proves in Chapter 3 of [7] the equivalence of two criteria for the coherent assessment of uncertainty. The first is based on the no-existence of a Dutch book and it is defined in terms of a two-player zero-sum game. Suppose that ψ_1, \dots, ψ_n are elements of the set of sentences built recursively from a finite set of propositional variables as usual, which are interpreted (see, e.g. [17]) as the events of interest to a bookmaker **B**. Suppose further that this interest materialises with the publication of a *book* $\beta: \psi_1 \mapsto \beta_1, \dots, \psi_n \mapsto \beta_n$ where for $i = 1, \dots, n$, $\beta_i \in [0, 1]$. A gambler **G** then chooses real-valued stakes $\sigma_1, \dots, \sigma_n$ and for $i = 1, \dots, n$, pays $\sigma_i \beta_i$ to **B**. **G** will then receive back $\sigma_i v(\psi_i)$, where $v(\psi_i) = 1$, if ψ_i is true, and $v(\psi_i) = 0$ otherwise. Thus, **G**'s payoff is $\sum_{i=1}^n \sigma_i (v(\psi_i) - \beta_i)$ and **B**'s payoff is $\sum_{i=1}^n \sigma_i (\beta_i - v(\psi_i))$. The book published by **B** is *coherent* if there is no choice of (possibly negative) stakes which **G** can make, exposing **B** to a sure loss. More precisely, for every $\sigma_1, \dots, \sigma_n \in \mathbb{R}$ there is a valuation v such that,

$$\sum_{i=1}^n \sigma_i (\beta_i - v(\psi_i)) \geq 0. \quad (1)$$

The second criterion is framed in terms of the individual decision of a forecaster **F**, who is asked to assess a value β_i to each event ψ_i (a forecast). **F** knows that they will suffer a penalty L_i (L stands for *loss*) proportional to the square of the euclidean distance between the realized value of ψ_i and the chosen value β_i . It is assumed that the forecaster's objective is to minimize their loss. A notable example of a loss function just introduced is the Brier scoring rule

$$L_i(\psi_i, \beta_i) = (\|v(\psi_i) - \beta_i\|_2)^2. \quad (2)$$

De Finetti shows that minimising the expectation of loss under the Brier rule is equivalent to avoiding sure loss in the Dutch Book setting. However he argues that it is preferable as it neutralises some potential shortcomings arising from the strategic aspects of the betting game. Even if they are formally defined in the same way, in the context of the Dutch Book, any assignment on a set of events is referred to in what follows as a *book*, while in the framework of scoring rules, it is referred to as a *forecast*.

Definition 1 (Proper Scoring Rule Criterion). Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a set of events and, for $i = 1, \dots, n$, let $\beta_i \in [0, 1]$ be the value the forecaster \mathbf{F} assigns to each ψ_i . The forecast $\beta: \psi_1 \mapsto \beta_1, \dots, \psi_n \mapsto \beta_n$ is coherent if there is no distinct forecast β' whose Brier score uniformly dominates β , i.e. if $L(\Psi, \beta) = \sum_i (\|v(\psi_i) - \beta_i\|_2)^2$, that is to say there is no β' s.t. $L(\Psi, \beta') < L(\Psi, \beta)$ for every valuation v .

De Finetti in [7], shows the equivalence between the Dutch Book and the Proper Scoring Rule criteria for defining (probabilistic) coherence. Therefore the following proposition holds.

Proposition 1. Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a set of events and let $\beta: \psi_i \mapsto \beta_i$ for $i = 1, \dots, n$ a forecast over Ψ . The forecast β is coherent iff it extends to a probability measure over the algebra of events.

Example 1. Let us consider the set of events $\Psi = \{\alpha, \neg\alpha\}$ and the forecast $\beta: \alpha \mapsto 0.8, \neg\alpha \mapsto 0.6$ on them. The penalty that \mathbf{F} will suffer depends on the realizations of the events. If $v(\alpha) = 1$ (and $v(\neg\alpha) = 0$), then $L(\Psi, \beta) = (1 - 0.8)^2 + (0 - 0.6)^2 = 0.4$. If $v(\alpha) = 0$ (and $v(\neg\alpha) = 1$), then $L(\Psi, \beta) = (0 - 0.8)^2 + (1 - 0.6)^2 = 0.8$. In the geometric interpretation of Fig. 1, the penalty relative to the first realization of the event α is the square of the euclidean distance between $(1, 0)$ and the point $\beta = (0.8, 0.6)$, while the penalty relative to the second one is the square of the euclidean distance between β and $(0, 1)$. Let us consider the projection of β onto the simplex of probabilities. This projection (also referred to as *de Finetti's projection*) identifies the point $\beta' = (0.6, 0.4)$. Since β' is closer to each endpoint of the simplex of probabilities, its Brier scoring rule is smaller than the Brier scoring rule relative to β , i.e. β' dominates β and β is *incoherent*.

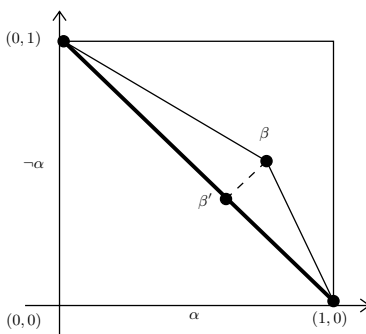


Fig. 1. Example of an incoherent forecast.

In geometrical terms, β is coherent if and only if it cannot be moved in such a way to reduce the distance from the set of all possible points. This is a

characterization of a convex hull. Note that this is not to say that any point inside a convex hull shares the same distance from all its vertices, but that no matter how we move the point, it will necessarily get closer to some points of the convex hull and farther from some others. In particular, this happens if we consider the vertices of the convex hull. Instead, if we consider a point lying outside the convex hull, then we can always move it in such a way as to reduce its distance from all the point of the convex hull.

Through Definition 1, we are not minimizing the sum of the squared euclidean distance between β and the vertices of the convex hull of probabilities (property hold by the *centroid*), but we minimize every single distance between β and all the points of the convex hull. As stated in [7], (footnote 18), if we move the point β to another position β^* , its distance from a generic point P increases or decreases depending on whether P is on the same side as β or β^* with respect to the hyperplane that bisects the segment $\beta\beta^*$ orthogonally. Two cases must be considered then:

1. If β is not in the convex hull \mathcal{P}_Ψ (the convex hull of the probabilities over Ψ), there exists a hyperplane separating it from \mathcal{P}_Ψ . Moving β to β^* , its orthogonal projection into such hyperplane, diminishes its distance from all points $P \in \mathcal{P}_\Psi$.
2. If β belongs to the convex hull \mathcal{P}_Ψ , then to whatever point β^* we move β , it always follows that for some point $P \in \mathcal{P}_\Psi$, the distance increases. If we construct the bisecting orthogonal hyperplane relative to $\beta\beta^*$ and P is on the same side as β^* , the point β would be distinguished from the convex hull of \mathcal{P}_Ψ , but this is contrary to the hypothesis.

3 Background on uncertainty measures and their geometric interpretation

We shall assume the reader to be familiar with basic notions and results of (finitely additive) probability theory. In particular, since we will only consider measures on finite hence atomic boolean algebras, we shall often identify a probability measure P on an algebra \mathbf{A} with the distribution p obtained by restricting P on the atoms of \mathbf{A} . As for the other uncertainty measures we will deal with in the following sections, it is convenient to recall some basic definitions and results from [9, 1, 14, 16].

Boolean algebras are understood as described in the signature $\{\wedge, \vee, \neg, \perp, \top\}$ and their elements are denoted by lower-case Greek letters with possible subscripts. In particular, the atoms of an algebra will be indicated as $\alpha_1, \alpha_2, \dots$.

Definition 2 (Belief function). *A Belief function B on an algebra \mathbf{A} is a $[0, 1]$ -valued map satisfying:*

$$(B1) \quad B(\top) = 1, \quad B(\perp) = 0;$$

$$(B2) \quad B\left(\bigvee_{i=1}^n \psi_i\right) \geq \sum_{i=1}^n \sum_{\{J \subseteq \{1, \dots, n\} : |J|=i\}} (-1)^{i+1} B\left(\bigwedge_{j \in J} \psi_j\right)$$

for $n = 1, 2, 3, \dots$

Belief functions on boolean algebras can be characterised in terms of *mass functions* as follows. Let \mathbf{A} be any finite boolean algebra with atoms $\alpha_1, \dots, \alpha_t$. A mass function is a map m that assigns to each subset X of atoms, a real number such that $m(\emptyset) = 0$ and $\sum_X m(X) = 1$. Given a mass function m , the map

$$B(\psi) = \sum_{X \subseteq \{\alpha_i \mid \alpha_i \leq \psi\}} m(X)$$

is a Belief function and every Belief function on \mathbf{A} can be defined in this way.

Definition 3 (Lower probability). A Lower probability \underline{P} on an algebra \mathbf{A} is a monotone $[0, 1]$ -valued map satisfying:

- (L1) $\underline{P}(\top) = 1, \underline{P}(\perp) = 0$;
- (L2) For all natural numbers n, m, k and all ψ_1, \dots, ψ_n , if $\{\{\psi_1, \dots, \psi_n\}\}$ is an (m, k) -cover of $(\varphi, \top)^3$, then $k + m\underline{P}(\varphi) \geq \sum_{i=1}^n \underline{P}(\psi_i)$.

Although the definition above does not make clear why those measures are called *Lower probabilities*, [1, Theorem 1] characterises them as follows: Let $\underline{P}: \mathbf{A} \rightarrow [0, 1]$ be a Lower probability and denote with $\mathcal{M}(\underline{P})$ the following set:

$$\mathcal{M}(\underline{P}) = \{P: \mathbf{A} \rightarrow [0, 1] \mid P \text{ is a probability function and } \forall \psi \in A, \underline{P}(\psi) \leq P(\psi)\}.$$

Then, for all $\psi \in A$,

$$\underline{P}(\psi) = \min\{P(\psi) \mid P \in \mathcal{M}(\underline{P})\}.$$

Lower probabilities are more general than Belief functions. The following result characterises the Lower probabilities that are Belief functions.

Remark 1. A Lower probability \underline{P} on an algebra \mathbf{A} is a Belief function iff \underline{P} satisfies (B2), namely

$$\underline{P}\left(\bigvee_{i=1}^n \psi_i\right) \geq \sum_{i=1}^n \sum_{\{J \subseteq \{1, \dots, n\} : |J|=i\}} (-1)^{i+1} \underline{P}\left(\bigwedge_{j \in J} \psi_j\right) \quad (3)$$

for all $n = 1, 2, \dots$

The geometric approach we consider is similar to Paris's [17] and is related to [5, 6].

³ An element φ of a boolean algebra \mathbf{A} is said to be covered m times by a multiset $\{\{\psi_1, \dots, \psi_n\}\}$ of elements of A if every homomorphisms of \mathbf{A} to $\{0, 1\}$ that maps φ to 1, also maps to 1 at least m propositions from ψ_1, \dots, ψ_n as well. An (m, k) -cover of (φ, \top) is a multiset $\{\{\psi_1, \dots, \psi_n\}\}$ that covers \top k times and covers φ $n+k$ times.

Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a finite set of events (i.e., elements of a finite boolean algebra \mathbf{A}). Let us denote by $\mathbb{V} = \{v_1, \dots, v_t\}$ the finite set of all possible homomorphisms of \mathbf{A} to the boolean chain on the two-element set $\{0, 1\}$. For every $j = 1, \dots, t$, call \mathbf{e}_j the binary vector

$$\mathbf{e}_j = (v_j(\psi_1), \dots, v_j(\psi_n)) \in \{0, 1\}^n. \quad (4)$$

Given this basic construction, we can characterise in geometric terms the extendability problem for assignments on Ψ to finitely additive probability measures and Belief functions. The additional notions we need are the euclidean convex hull $\overline{\text{co}}(X)$ of a subset $X \subseteq \mathbb{R}^t$ (which reduces to $\text{co}(X)$ in case X is finite) and the less common *tropical* convex hull $\text{co}_{\wedge,+}(X)$ of X (see [8]).

Definition 4 (Tropical Hull). *Let $\mathbf{x}_1, \dots, \mathbf{x}_t \in [0, 1]^n$. The tropical hull of the \mathbf{x}_j 's is the subset $\text{co}_{\wedge,+}(\mathbf{x}_1, \dots, \mathbf{x}_t)$ of all points \mathbf{y} of $[0, 1]^n$ for which there exist parameters $\lambda_1, \dots, \lambda_t \in [0, 1]$ such that $\bigwedge_{j=1}^t \lambda_j = 0$ and*

$$\mathbf{y} = \bigwedge_{j=1}^t \lambda_j + \mathbf{x}_j.$$

The symbol \wedge stands for the minimum and $+$ for the ordinary addition in the tropical semiring $(\mathbb{R}, \wedge, +)$. Given $\lambda \in [0, 1]$ and $\mathbf{x} \in [0, 1]^n$, $\lambda + \mathbf{x} = (\lambda + x_1, \dots, \lambda + x_n)$ and the \wedge operator is defined component-wise.

For $\mathbf{e}_1, \dots, \mathbf{e}_t$ being defined as above from the formulas ψ_i 's in Ψ , let us consider the following sets:

1. $\mathcal{P}_\Psi = \text{co}(\mathbf{e}_1, \dots, \mathbf{e}_t)$;
2. $\mathcal{B}_\Psi = \overline{\text{co}}(\text{co}_{\wedge,+}(\mathbf{e}_1, \dots, \mathbf{e}_t))$, where, in this case, being $\text{co}_{\wedge,+}(\mathbf{e}_1, \dots, \mathbf{e}_t)$ usually uncountable, $\overline{\text{co}}$ denotes the topological closure of the Euclidean convex hull co .

Theorem 1 ([7, 10, 12]). *Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a finite set of events and let $\beta : \Psi \rightarrow [0, 1]$ be an assignment. Then,*

1. β extends to a probability measure iff $(\beta(\psi_1), \dots, \beta(\psi_n)) \in \mathcal{P}_\Psi$;
2. β extends to a Belief function iff $(\beta(\psi_1), \dots, \beta(\psi_n)) \in \mathcal{B}_\Psi$.

In general, \mathcal{P}_Ψ is strictly included in \mathcal{B}_Ψ (i.e., $\mathcal{P}_\Psi \subset \mathcal{B}_\Psi$) and this is expected because Belief functions are strictly more general than probabilities measures. In [4], we ask whether avoiding sure loss is sufficient to distinguish Lower probabilities from Belief functions. In other words, we ask whether the above strict inclusion is matched by a detectable difference in coherence, and find that the answer is negative.

We denote with \mathcal{L}_Ψ the set of all assignments β on Ψ that extend to a Lower probability \underline{P} . The sets of events that do not distinguish Belief functions from Lower probabilities are referred to as *adequate*. The main result of our previous work reads as follows.

Theorem 2. *For every algebra \mathbf{A} with at least three atoms there exists an adequate subset Ψ of A , i.e., Ψ is such that $\mathcal{B}_\Psi = \mathcal{L}_\Psi$.*

We now introduce a scoring rule that characterises assignments extendible to Belief functions. In addition, we investigate *adequate* sets of events from the perspective of this rule.

4 Coherence for Belief functions

Let us begin by recalling Jaffray’s extension of de Finetti’s Dutch Book to Belief functions [15]. The key idea of his framework is that if an event ψ occurs, then every non-contradictory event which logically follows from ψ , also occurs.

Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a set of events and $\beta: \psi_1 \mapsto \beta_1, \dots, \psi_n \mapsto \beta_n$ a book published by the bookmaker \mathbf{B} . If the gambler \mathbf{G} places real stakes $\sigma_1, \dots, \sigma_n$ on ψ_1, \dots, ψ_n at the betting odds written in β , then \mathbf{G} pays \mathbf{B} for each ψ_i the amount $\sigma_i \beta_i$ and gains the amount $\sigma_i C_\psi(\psi_i)$. The function $C_\psi(\psi_i)$ is defined as follows, where \models_{cl} denotes the consequence relation of classical propositional logic:

$$C_\psi(\psi_i) = \begin{cases} 1 & \text{if } \models_{cl} \psi \rightarrow \psi_i, \\ 0 & \text{otherwise.} \end{cases}$$

The total balance for \mathbf{B} is

$$\sum_{i=1}^n \sigma_i (\beta_i - C_\psi(\psi_i)).$$

Definition 5 (Coherence Under Partially Resolved Uncertainty). *Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a set of events and β a book over Ψ . The book β is coherent under partially resolved uncertainty if there is no choice of stakes which \mathbf{G} can make, exposing \mathbf{B} to a sure loss, i.e. it is not the case that, for every fixed non-contradictory event ψ , $\sum_{i=1}^n \sigma_i (\beta_i - C_\psi(\psi_i)) < 0$.*

This notion of coherence characterises Dempster-Shafer Belief functions [21] in the sense that a book β on Ψ is coherent under partially resolved uncertainty if and only if it extends to a Belief function on the algebra of events, [19].

A criterion based on a Brier-like scoring rule can be put forward for Belief functions as follows. From Definition 1 above it follows that a forecast β is *dominated* by another forecast β' if the squared euclidean distance between β and the vertices of \mathcal{P}_Ψ is greater than the squared euclidean distance between β' and the same points. Thus, we can move β to reduce the distance from all points of \mathcal{P}_Ψ , vertices included. Hence, we can extend the definition of *dominating forecast* to Belief functions. To do this we consider the vertices of the polytope \mathcal{B}_Ψ , rather than those of \mathcal{P}_Ψ . This distance can be recovered from the C_ψ function used by Jaffray recalled above:

$$L(\psi_i, \beta_i) = (\|C_\psi(\psi_i) - \beta_i\|_2)^2. \quad (5)$$

Remark 2. As pointed out by two anonymous referees, the forthcoming [18] provides a characterisation of scoring rules for Belief functions yielding a result which is essentially equivalent to Theorem 3 below. Whilst we take this as welcome news to the effect that scoring-rule based coherence is of interest beyond the well-known probabilistic case, the present work differs substantially in its motivating question from [18]. In particular, whilst this latter reviews several candidate notions for ‘coherent’ Belief functions, our present concern lies in comparing Belief functions and Lower probabilities against the natural formulation provided by Definition 6. Indeed, as we point out in Subsection 5.2, the notion captured by this Definition allows us to tell *coherence-wise* Belief functions from Lower probability, a distinction not permitted by Dutch-Book coherence [4].

Definition 6 (Scoring Rule for Belief functions). *Let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a set of events and, for $i = 1, \dots, n$, let $\beta_i \in [0, 1]$ be the value the forecaster \mathbf{F} assigns to each ψ_i . The forecast $\beta: \psi_1 \mapsto \beta_1, \dots, \psi_n \mapsto \beta_n$ is coherent if there is no distinct forecast β' whose Brier score uniformly dominates β , i.e. if $L(\Psi, \beta) = \sum_i (\|C_\psi(\psi_i) - \beta_i\|_2)^2$, then there is no β' s.t. $L(\Psi, \beta') < L(\Psi, \beta)$ for every non-contradictory event ψ and every possible value of $C_\psi(\psi_i)$.*

As shown in the following Proposition, \mathcal{B}_Ψ can be characterised by using the C_ψ function. We denote by $C_\psi(\Psi)$ the vector $(C_\psi(\psi_1), \dots, C_\psi(\psi_n))$.

Proposition 2. *Let \mathbf{A} be a boolean algebra, $\Psi = \{\psi_1, \dots, \psi_n\}$ be a finite set of events over \mathbf{A} and $\beta: \Psi \rightarrow [0, 1]$ be an assignment over Ψ . Then*

$$\mathcal{B}_\Psi = \text{co}(C_\psi(\Psi) \mid \psi \text{ is a non-contradictory event in } \mathbf{A}).$$

Proof. Since both \mathcal{B}_Ψ and $\text{co}(C_\psi(\Psi) \mid \psi \text{ is a non-contradictory event in } \mathbf{A})$ are convex hulls, to show that they coincide, it is sufficient to show that they share the same vertices. In particular, recall that the vectors \mathbf{e}_j are defined as $\mathbf{e}_j = (v_j(\psi_1), \dots, v_j(\psi_n))$ where v_j is an homomorphisms of \mathbf{A} to the boolean chain on the two-element set $\{0, 1\}$. Thus, \mathbf{e}_j identifies in \mathbb{R}^n the same point of $C_{\alpha_j}(\Psi) = (C_{\alpha_j}(\psi_1), \dots, C_{\alpha_j}(\psi_n))$ where α_j is an atom of \mathbf{A} . The other vertices of \mathcal{B}_Ψ are recovered from the \mathbf{e}_j with $j = 1, \dots, t$ by taking the point-wise minimum of every subset of \mathbf{e}_j 's. Since the entailment relation to which C_ψ refers to is the consequence relation of classical propositional logic, for any propositional formula ψ_i and ψ_j we have $\models_{cl} (\psi_i \vee \psi_j) \rightarrow \psi_i$ iff $\models_{cl} \psi_i \rightarrow \psi_i$ and $\models_{cl} \psi_j \rightarrow \psi_i$. Thus, $C_{\alpha_i \vee \alpha_j}(\psi_i) = 1$ iff $C_{\alpha_i}(\psi_i) = 1$ and $C_{\alpha_j}(\psi_i) = 1$, i.e. $\min\{C_{\alpha_i}(\psi_i), C_{\alpha_j}(\psi_i)\} = 1$. Therefore, the vertices of \mathcal{B}_Ψ obtained as $\mathbf{e}_i \wedge \mathbf{e}_j$ with $i, j \in \{1, \dots, t\}$ identify the same point of $C_{\alpha_i \vee \alpha_j}(\Psi)$ and, in general,

$$\bigwedge_{j \in J \subseteq \{1, \dots, t\}} \mathbf{e}_j = C_{\psi_J}(\Psi),$$

where, for all $J \subseteq \{1, \dots, t\}$, $\psi_J = \bigvee_{j \in J \subseteq \{1, \dots, t\}} \alpha_j$.

Theorem 3. *Let \mathbf{A} be a boolean algebra, and let $\Psi = \{\psi_1, \dots, \psi_n\}$ be a finite set of elements of \mathbf{A} . A forecast β defined over Ψ is coherent if and only if it can be extended to a Belief function on \mathbf{A} .*

Proof. (\Rightarrow) If the forecast β is coherent, then there is no rival forecast β' defined over the same set of events s.t. $L(\Psi, \beta') < L(\Psi, \beta)$ for every non-contradictory event ψ and every possible value of $C_\psi(\psi_i)$. By Proposition 2, the vertices of \mathcal{B}_Ψ coincide with all possible values of C_Ψ . Thus, $\beta \in \mathcal{B}_\Psi$ and by Theorem 1 it can be extended to a Belief function over \mathbf{A} .

(\Leftarrow) If the forecast β can be extended to a Belief function over \mathbf{A} , then by Theorem 1, $\beta \in \mathcal{B}_\Psi$ and by Proposition 2, $\beta \in \text{co}(C_\psi(\psi_i) \mid \psi \text{ is a non-contradictory event and } i = 1, \dots, n)$. Therefore, no β' reduces the (euclidean) distance from all possible point of that convex hull (vertices included), which implies that there is no β' s.t. $L(\Psi, \beta') < L(\Psi, \beta)$ for any possible value of C_ψ .

Example 2. Let \mathbf{A} be the boolean algebra of 8 elements and 3 atoms $\{\alpha_1, \alpha_2, \alpha_3\}$ and consider the non-trivial set of events $\Psi = \{\psi_1, \psi_2, \psi_3\} \subset A$ where $\psi_1 = \alpha_1 \vee \alpha_2$, $\psi_2 = \alpha_2 \vee \alpha_3$ and $\psi_3 = \alpha_1 \vee \alpha_3$. In Table 2 we compute the $C_\psi(\psi_i)$ function for every non-contradictory event ψ .

ψ	$C_\psi(\psi_1)$	$C_\psi(\psi_2)$	$C_\psi(\psi_3)$
α_1	1	0	1
α_2	1	1	0
α_3	0	1	1
$\alpha_1 \vee \alpha_2$	1	0	0
$\alpha_2 \vee \alpha_3$	0	1	0
$\alpha_1 \vee \alpha_3$	0	0	1
$\alpha_1 \vee \alpha_2 \vee \alpha_3$	0	0	0

Table 1. Values of $C_\psi(\psi_i)$ over the boolean algebra \mathbf{A} and the set of events Ψ

Let us consider the forecasts $\beta_1: \psi_1 \mapsto 3/8, \psi_2 \mapsto 3/8, \psi_3 \mapsto 3/8$ and $\beta_2: \psi_1 \mapsto 7/8, \psi_2 \mapsto 7/8, \psi_3 \mapsto 7/8$. If we compute the convex hull considering the points identified by $C_\psi(\Psi_i)$, we can verify that β_1 belongs to it while β_2 does not. Since the convex hull generated by C_ψ coincides with \mathcal{B}_Ψ , β_1 extends to a Belief function and β_2 does not.

5 Comparing Scoring Rules for Belief functions and Imprecise probabilities

5.1 Scoring Rule for Imprecise probabilities

Seidenfeld et al. introduce in [20] the following scoring rule for imprecise probabilities. Let us consider the event ψ and a Lower and Upper probability forecast (p, q) for the event ψ , that is to say, let us assume that $p, q \in [0, 1]$ are values assigned to ψ by a Lower probability and its dual Upper probability, respectively. The Brier-style IP scoring rule is hence defined as follows.

$$L(\psi, (p, q)) = \begin{cases} (1 - q)^2 & \text{if } v(\psi) = 1, \\ p^2 & \text{if } v(\psi) = 0. \end{cases}$$

Extending the concept of *dominance* between forecasts from the definitions on probabilities and Belief functions, we say that a forecast $\mathcal{F} = \{(p_i, q_i) \mid i = 1, \dots, n\}$ over the events $\Psi = \{\psi_1, \dots, \psi_n\}$ strictly dominates another forecast $\mathcal{F}' = \{(p'_i, q'_i) \mid i = 1, \dots, n\}$ if $L(\Psi, \mathcal{F}) < L(\Psi, \mathcal{F}')$. The penalty for a set of forecasts is the sum of the individual penalty scores.

Since the forecast $\mathcal{F}_0 = \{(p_i = 0, q_i = 1) \mid i = 1, \dots, n\}$ dominates any other forecast, coherence for imprecise probabilities based on scoring rule risks to trivialise if we do not further elaborate it. For this reason, we need to consider an additional restriction on the class of rival forecasts and an index of relative imprecision between forecasts.

This is the reason why, for each forecast $\mathcal{F} = \{(p_i, q_i) \mid i = 1, \dots, n\}$, Seidenfeld et al., construct a *scoring set* $\mathcal{S}_{\mathcal{F}}$ defined as follows:

$$\mathcal{S}_{\mathcal{F}} = \{(q_1, p_2, \dots, p_n), (p_1, q_2, \dots, p_n) \dots, (p_1, \dots, p_{n-1}, q_n)\}.$$

Thus, a forecast \mathcal{F} is *at least as determinate as* a forecast \mathcal{F}' if $\text{co}(\mathcal{S}_{\mathcal{F}})$ is isomorphic under rigid movements of a subset of $\text{co}(\mathcal{S}_{\mathcal{F}'})$.

Finally, the class \mathcal{M} of *rival forecasts* considered is the ε -contamination class.⁴ The scoring criterion for imprecise probabilities is then defined as follows.

Definition 7 (Scoring Rule for Imprecise probabilities). *Let \mathbf{A} be a boolean algebra and let \mathcal{F} be a forecast over the atoms of the algebra. \mathcal{F} is IP-coherent with respect to \mathcal{M} if there is no dominating forecast \mathcal{F}' from \mathcal{M} such that $\mathcal{S}'_{\mathcal{F}}$ is at least as determinate as $\mathcal{S}_{\mathcal{F}}$.*

The following theorem characterises coherent forecasts.

Theorem 4. *Let \mathbf{A} be a boolean algebra and let \mathcal{F} be a forecast over the atoms $\Psi = \{\alpha_1, \dots, \alpha_t\}$ of the algebra \mathbf{A} . Then, $\mathcal{S}_{\mathcal{F}}$ lies entirely within the probability simplex \mathcal{P}_{Ψ} iff \mathcal{F} matches an ε -contamination model and it is IP-coherent with respect to \mathcal{M} .*

It is now worth to point out that, for what concerns the *extendability problem*, an immediate consequence of the theorem above is that a forecast $\mathcal{F} = \{(p_i, q_i) \mid i = 1, \dots, t\}$ on the atoms $\alpha_1, \dots, \alpha_t$ is IP-coherent with respect to \mathcal{M} if and only if for all $i = 1, \dots, t$, $p_i = \min\{P(\alpha_i) \mid P \in \text{co}(\mathcal{S}_{\mathcal{F}})\}$ and $q_i = \max\{P(\alpha_i) \mid P \in \text{co}(\mathcal{S}_{\mathcal{F}})\}$ where $\Psi = \{\alpha_1, \dots, \alpha_t\}$. In addition, as we will see in the following Section, it can be that a forecast \mathcal{F} is not coherent as for Definition 7 and still being extendable to an Upper and Lower probability.

⁴ For what concerns the present paper, we do not need to elaborate more on the previously given notions and definitions, and we invite the interested reader to consult [20] for the relative literature on this class of forecasts.

5.2 Distinguishing Belief functions and Imprecise probabilities through Scoring Rules

As illustrated in [4], given a finite boolean algebra \mathbf{A} with at least three atoms we can identify an *adequate* set of events and assignments β_i that do not satisfy equation (3) and still cannot distinguish Belief functions from Lower probabilities. In particular, the main result of [4] is based on the following.

Remark 3. [4, Example 2] Let us consider an algebra \mathbf{A} with atoms $\alpha_1, \dots, \alpha_t$ ($t \geq 3$) and probability distributions $p_1(\alpha_1) = q, p_1(\alpha_2) = 1 - q, p_1(\alpha_i) = 0$ for all $i \neq 1, 2$; $p_2(\alpha_2) = q, p_2(\alpha_3) = 1 - q, p_2(\alpha_i) = 0$ for all $i \neq 2, 3$; $p_3(\alpha_1) = 1 - q, p_3(\alpha_3) = q, p_3(\alpha_i) = 0$ for all $i \neq 1, 3$ where q is any value $1/3 < q \leq 1/2$. Let us consider, as events, the co-atoms of \mathbf{A} : $\psi_1 = \alpha_1 \vee \alpha_2, \psi_2 = \alpha_2 \vee \alpha_3, \psi_3 = \alpha_1 \vee \alpha_3$ and the assignment $\beta : \psi_i \mapsto q$ for every $i = 1, 2, 3$. The Lower probability \underline{P} derived by the probability distributions p_j extends β . Furthermore, it not difficult to show that \underline{P} is not a Belief function as it fails to prove (3) of Remark 1. However, the same assignment β is coherent in the sense of Definition 5 and in fact there is a Belief function of \mathbf{A} that extends it. Thus, for that set of events, it is impossible to distinguish assignments that are coherent in the sense of Belief functions, from those that are coherent in the sense of Lower probabilities.

Now, let us move to analyse what happens if we take into account coherence through proper scoring rules. First, it is easy to see that the assignment displayed in Remark 3 is coherent according to Definition 6. However, if we consider the forecast $\mathcal{F} = \{(0, 1 - q), (0, 1 - q), (0, 1 - q)\}$ relative to the same probability distribution over the atoms of \mathbf{A} and construct the scoring set $\mathcal{S}_{\mathcal{F}} = \{(1 - q, 0, 0), (0, 1 - q, 0), (0, 0, 1 - q)\}$, then it does not lie in $\mathcal{P}_{\Psi'}$ where Ψ' is the set of the three atoms considered. In fact, since $1/3 < q \leq 1/2$, then $1/2 \leq 1 - q < 2/3$ and $\text{co}(\mathcal{S}_{\mathcal{F}}) \not\subseteq \mathcal{P}_{\Psi'} = \text{co}((1, 0, 0), (0, 1, 0), (0, 0, 1))$. Thus, by Theorem 4, the forecast \mathcal{F} is not IP-coherent with respect to the ε -contamination model class even though the assignments $\beta_{\underline{P}}: \alpha_1 \mapsto 0, \alpha_2 \mapsto 0, \alpha_3 \mapsto 0$ and $\beta_{\overline{P}}: \alpha_1 \mapsto 1 - q, \alpha_2 \mapsto 1 - q, \alpha_3 \mapsto 1 - q$ can be extended to Lower and Upper probabilities, respectively.

Let us conclude with a brief discussion on why scoring rule-based coherence seems to be stronger (in the sense that allows to distinguish more) than avoiding sure loss. Indeed, although this argument needs further and deeper insights that will be addressed in our future work, it seems clear that such an asymmetry might be caused by the information that these criteria encode in their formalization. In particular, although scoring rule-based criterion for Belief functions (Definition 6) relies on the sole information carried by the events ψ_1, \dots, ψ_n on which (coherent) forecasts are defined, the same criterion for imprecise probabilities (Definition 7) needs a forecast \mathcal{F} to be defined on the atoms *and on the co-atoms* of \mathbf{A} . In this setting, in fact, a forecast assigns values of Lower and Upper probability to each atoms and hence, in algebraic terms, it needs both atoms and co-atoms of \mathbf{A} to be defined: if $q_i = \overline{P}(\alpha_i) = 1 - \underline{P}(\neg\alpha_i)$ being $\neg\alpha = \bigvee_{j \neq i} \alpha_j$ a co-atom of \mathbf{A} . We believe that the observed asymmetry between the Dutch

Book and scoring rule-based criteria, especially for imprecise probabilities, lies in this additional information that the latter requires.

6 Conclusion

We investigated particular sets of events that do not distinguish Belief functions from Lower probabilities from the point of view of Dutch Books, but which are distinguishable through scoring rules. We introduced a loss function through which we characterise assignments defined over a generic subset of events and that can be extended to a Belief function. Considering an already existing scoring rule defined for imprecise probabilities, we observed that the same assignments from which we can recover adequate sets of events, through coherence, can distinguish Lower probabilities from Belief functions. As pointed out at the end of the previous section, this result opens up new lines of research. In particular, given the tight connection between the scoring rule criterion and the geometric interpretation of the extendibility criterion, we could first characterise geometrically the sets \mathcal{L}_Ψ and \mathcal{U}_Ψ of the assignments defined over Ψ and that can be extended to Lower and Upper probabilities. Then, following the same process used in Section 4 for the Belief functions, define additional scoring rules through which it is possible to characterise assignments defined over generic sets of events and that can be extended to Upper or Lower probabilities. We could also investigate how to define a scoring rule that characterises assignments extendible to necessity measures.

The comparison reported in this paper are limited to the unconditional case. A general logical framework to investigate conditional probability functions has recently been introduced in [11]. Combining this with the conditional approach to Dutch Book coherence put forward in [3, 2] on the one hand, and the approach based on scoring rules of [13] on the other hand, we aim to extend our comparison to conditional, non-additive measures of uncertainty. With regards to Belief functions the very recent [18] obtains related results from a distinct angle, as briefly noted in Remark 2 above. Further work will elucidate potential mutual relations between the two independent approaches.

Acknowledgments We are grateful to the anonymous referees for their thorough comments on the draft and to the ECSQARU Programme Committee for their very competent handling of the paper. Corsi and Hosni acknowledge funding by the Department of Philosophy “Piero Martinetti” of the University of Milan under the Project “Departments of Excellence 2018-2022” awarded by the Ministry of Education, University and Research (MIUR). Flaminio acknowledges partial support by the Spanish project PID2019-111544GB-C21 and by the Spanish Ramón y Cajal research program RYC-2016-19799. Hosni also acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, grant LA 4093/3-1).

References

1. Anger, B., Lembcke, J.: Infinitely subadditive capacities as upper envelopes of measures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **68**(3), 403–414 (1985)
2. Coletti, G., Petturiti, D., Vantaggi, B.: A Dutch Book coherence condition for conditional completely alternating choquet expectations. *Bollettino dell’Unione Matematica Italiana* **13**(4), 585–593 (2020)
3. Coletti, G., Scozzafava, R.: Probabilistic logic in a coherent setting, vol. 15. Springer Science & Business Media (2002)
4. Corsi, E.A., Flaminio, T., Hosni, H.: When belief functions and lower probabilities are indistinguishable. In: *ISIPTA 2021 - Proceedings of Machine Learning Research* (147). pp. 83–89 (2021)
5. Cuzzolin, F.: The geometry of consonant belief functions: simplicial complexes of necessity measures. *Fuzzy Sets and Systems* **161**(10), 1459–1479 (2012)
6. Cuzzolin, F.: *The geometry of uncertainty*. Springer (2017)
7. De Finetti, B.: *Theory of probability*, vol. 1. John Wiley & Sons (1974)
8. Develin, M., Sturmfels, B.: Tropical convexity. arXiv preprint math/0308254 (2003)
9. Dubois, D., Prade, H.: *Possibility theory: Approach to computerized processing of uncertainty*. Plenum Press, New York (1988)
10. Flaminio, T., Godo, L.: A note on the convex structure of uncertainty measures on MV-algebras. In: *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, pp. 73–81. Springer (2013)
11. Flaminio, T., Godo, L., Hosni, H.: Boolean algebras of conditionals, probability and logic. *Artificial Intelligence* **286**, 103347 (2020)
12. Flaminio, T., Godo, L., Marchioni, E.: Geometrical aspects of possibility measures on finite domain MV-clans. *Soft Computing* **16**(11), 1863–1873 (2012)
13. Gilio, A., Sanfilippo, G.: Coherent conditional probabilities and proper scoring rules. In: *Proc. of ISIPTA*. vol. 11, pp. 189–198. Citeseer (2011)
14. Halpern, J.Y.: *Reasoning about uncertainty*. MIT press (2017)
15. Jaffray, J.Y.: Coherent bets under partially resolving uncertainty and belief functions. *Theory and decision* **26**(2), 99–105 (1989)
16. Miranda, E.: A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning* **48**(2), 628–658 (2008)
17. Paris, J.B.: A note on the Dutch Book method. In: *ISIPTA*. vol. 1, pp. 301–306 (2001)
18. Petturiti, D., Vantaggi, B.: How to assess coherent beliefs: A comparison of different notions of coherence in Dempster-Shafer theory of evidence. arXiv preprint arXiv:2105.10546 (2021)
19. Regoli, G.: Rational comparisons and numerical representations. *Decision Theory and Decision Analysis: Trends and Challenges* pp. 113–126 (1994)
20. Seidenfeld, T., Schervish, M.J., Kadane, J.B.: Forecasting with imprecise probabilities. *International Journal of Approximate Reasoning* **53**(8), 1248–1261 (2012)
21. Shafer, G.: *A mathematical theory of evidence*, vol. 42. Princeton university press (1976)
22. Walley, P.: *Statistical reasoning with imprecise probabilities*. Chapman & Hall (1991)