Nonlinear desirability as a linear classification problem

Arianna Casanova[a,*], Alessio Benavoli[b], Marco Zaffalon[a]

[a] *Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland*
[b] *School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland*

Abstract

This paper presents an interpretation as *classification problem* for standard desirability and other instances of nonlinear desirability (*convex coherence* and *positive additive coherence*). In particular, we analyze different sets of rationality axioms and, for each one of them, we show that proving that a subject respects these axioms on the basis of a finite set of *acceptable* and a finite set of *rejectable* gambles can be reformulated as a binary classification problem where the family of classifiers used changes with the axioms considered. Moreover, by borrowing ideas from machine learning, we show the possibility of defining a *feature mapping*, which allows us to reformulate the above nonlinear classification problems as linear ones in higher-dimensional spaces. This allows us to interpret gambles directly as payoffs vectors of *monetary lotteries*, as well as to provide a practical tool to check the rationality of an agent.

Keywords: Imprecise probabilities, Coherence, Convex coherence, Monetary scale, Piecewise separators

1. Introduction

The Bayesian framework is a sound and consistent theory because it is a *logic*. In fact, it can be shown that the rules of (Bayesian) probabilities can be inferred via mathematical duality from a set of *logical axioms* [1, 2, 3, 4, 5, 6]. These axioms can also be interpreted as rationality requirements on the way a subject is willing to accept *gambles*, i.e., bets whose possible outcomes are uncertain gains or losses of a single commodity depending on the result of an experiment. The resulting theory is called *theory of desirable gambles* or *desirability theory*.

Let us consider an experiment with a finite set of possible outcomes, a so-called (finite) *possibility space*, $\Omega \coloneqq \{\omega_1, ..., \omega_n\}$. In this context, gambles g are represented by vectors in $\mathbb{R}^n$. If we denote by $\mathcal{D} \subseteq \mathbb{R}^n$, the set of gambles an agent is disposed to accept, a so called *set of desirable gambles*, rationality axioms usually required on it can be expressed as follows.[1]

D1:	Tautologies	if $g \in \mathbb{R}^n,\ g \geq 0$ then $g \in \mathcal{D}$,
D2:	Falsum	if $g \in \mathbb{R}^n,\ g < 0$ then $g \notin \mathcal{D}$,
D3:	Linearity	if $g, h \in \mathcal{D}$ then $\lambda g + \mu h \in \mathcal{D}$ for any $\lambda, \mu \in \mathbb{R}_+$,
D4:	Closure	if $g \in \mathbb{R}^n,\ g + \epsilon \in \mathcal{D}$ for all $\epsilon \in \mathbb{R}^*_+$, then $g \in \mathcal{D}$,
D5:	Completeness	if $g \in \mathbb{R}^n,\ g \notin \mathcal{D}$ then $-g \in \mathcal{D}$.

The first axiom – D1 – also known as *accepting partial gains* criterion in literature [5], states that an agent should always accept non-negative gambles, because they can only provide gains (or keep the status

* Corresponding author
Email addresses: arianna@idsia.ch (Arianna Casanova), alessio.benavoli@tcd.ie (Alessio Benavoli), zaffalon@idsia.ch (Marco Zaffalon)

[1] In D4, we use ϵ to denote also the constant gamble given by $g(\omega) = \epsilon$ for every $\omega \in \Omega$. Moreover, we indicate with $\mathbb{R}_+$ the positive real numbers (including zero) and with $\mathbb{R}^*_+$ the strictly positive real numbers.

quo) for the agent. Analogously, the second axiom – D2 –also known as *avoiding sure loss* criterion in literature [5], states that an agent should always avoid negative gambles, because they can only provide losses. The third axiom – D3 – states instead that the agent judges rewards of gambles with a linear utility scale. The fourth axiom – D4 – instead, is a closure requirement.[2] A set of gambles $\mathcal{D}$ satisfying axioms D1–D4 is the mathematical dual of a set of *finitely additive probabilities.* D5 restricts this set to have a unique element [5].

When the set of possible outcomes of an experiment Ω is enriched with a set of prizes, these axioms provide not only the foundation for a powerful model of uncertainty but also the one for a theory of rational decision making [7, 8, 9], thus providing a connection between desirability and Von Neumann and Morgestern's [10] and Anscombe and Aumann's [1] axiomatisation of rationality. However, in several situations the above axioms can be restrictive and researchers have relaxed and generalised them in many ways.

For instance, it is not very realistic to assume that an agent can always compare alternatives. This was shown to be equivalent to the axiom D5 that, therefore, is often dropped [5, 6]. In this case we can obtain axiomatisations of rational decision making under *incompleteness* [5, 7, 11, 12, 13, 14].

The closure axiom D4 can also be abandoned. Although D4 makes the connection between desirability and probability theory tighter [15, 16, 17], the extra generality of sets of desirable gambles without D4 makes them useful when dealing with the problem of conditioning on sets of probability zero, or of choosing between two options under zero expectation [5, 18]. The axioms D1 and D2, even though their requirement seems very natural, can also be restrictive. For instance, in case Ω is an infinite possibility space, evalutating the positivity (or negativity) of a gamble can be computationally very demanding. This leads to a notion of computational rationality [19, 20] which restricts D1 and D2. [21] instead, focuses on relaxations of the notion of avoiding sure loss (D2) only. Finally, the axiom D3 (linearity) is also not very realistic, especially when one considers *monetary* gambles, i.e., gambles that return money. Indeed, for example, for a large positive λ difficulties might be encountered at accepting λg for every acceptable gamble g, because of lack of market liquidity at some degree. The linearity axiom has been replaced by convexity in [22, 23] and it is further relaxed in [24].

In this paper we analyze some generalisations of the linearity axiom D3. Specifically, since we work only with finite possibility spaces Ω and, for the computational counterpart of our results, is more convenient to work with closed sets, we assume that D1,D2,D4 hold true and we analyse sets of rationality axioms where D3 is relaxed and D5 is no more present. Given that in practical situations an agent can provide only a finite set $\mathcal{A}$ of acceptable, or even a set $\mathcal{R}$ of *rejectable*,[3] gambles, we focus on furnishing a way to check, each time, if $\mathcal{A}$ and $\mathcal{R}$ are compatible with the set of rationality axioms considered.[4] We show that a way to check it, both when we consider standard sets of axioms and in more general cases, is to solving a binary classification problem. Specifically, after giving some preliminary notions about sets of desirable gambles respecting standard assumptions in Section 2, we show that:

- if we assume rationality axioms D1–D5, this check can be made solving a *binary linear classification* problem (see Section 3);
- if we assume D1–D4 but not *completeness* axiom D5, this check can be made solving a *binary piecewise linear classification task*, i.e., a binary classification problem where the decision boundary is a *piecewise linear function*[5](see Section 4);
- if rationality is expressed through D1, D2, convexity replacing linearity, and D4 [22], this check can be made solving a *binary piecewise affine classification task*, i.e., a generalization of a binary piece-

[2] We will see later on indeed that, assuming other standard axioms, D4 makes $\mathcal{D}$ closed with respect to the usual topology of $\mathbb{R}^n$.

[3] By rejecting a gamble, the agent expresses that they consider accepting that gamble unreasonable.

[4] A framework for modelling and reasoning with uncertainty based on accept and reject statements about gambles has already been developed in [25]. In [25] however, unlike here, also non-finite sets of acceptance/rejection statements are considered and it is always assumed the linearity axiom D3.

[5] Notice that here by piecewise linear functions we mean only those functions composed by $N \geq 1$ functions $\{f_j(\cdot)\}_{j=1}^N$ defined as $f_j(g) := g^\top \beta_j$ on different domains, with $\beta_j \in \mathbb{R}^n$. We do not consider as piecewise linear functions instead, those functions composed by $N \geq 1$ functions $\{f_j(\cdot)\}_{j=1}^N$ defined as $f_j(g) := g^\top \beta_j + \alpha_j$ on different domains, with $\beta_j \in \mathbb{R}^n$, $\alpha_j \in \mathbb{R}$ and $\alpha_k \neq 0$ for at least a $k \in \{1, ..., N\}$ (see Section 4)

wise linear classification problem where the decision boundary is a piecewise *affine*, instead of linear, function (see Section 5);

- if rationality is expressed through D1, D2, 'positive additivity' replacing linearity,[6] and D4, this check can be made solving a *binary piecewise positive affine classification task*, i.e., a binary classification problem where one of the regions identified by the classifier is a union of orthants (see Section 6);
- for more general cases the check can be made using a more general *nonlinear classifier* (see Section 8).

By borrowing ideas from machine learning then, we show that we can define a *feature mapping* that allows us to reformulate the above nonlinear classification problems as linear ones in higher-dimensional spaces.

We also extend some notions related to the probabilistic interpretation of sets of desirable gambles in the new contexts considered in Section 7.

2. Preliminaries

Here we summarize necessary notation and basic definitions about sets of desirable gambles assuming standard rationality axioms. For additional comments see, for example, [5, 26].

Let $\Omega = \{\omega_1, ..., \omega_n\}$ denote the finite set of the possible outcomes of an experiment. A gamble is a bounded function $g : \Omega \to \mathbb{R}$, which is usually interpreted as an uncertain reward in a linear utility scale. In this paper however, given that we want to relax the linearity axiom, we choose to interpret a gamble as a payoff vector of a monetary lottery on the experiment whose outcomes are summarized in Ω. Hence, if an agent is disposed to accept a gamble g, they are disposed to obtain the monetary value $g(\omega)$ if ω is the true unknown outcome of the experiment. Given a gamble g, we represent it as a column vector in $\mathbb{R}^n$ and we use g_i to indicate its component $g(\omega_i)$ for every $i \in \{1, ..., n\}$. A similar notation is used also for other generic vectors in $\mathbb{R}^m$ with $m \in \mathbb{N}^*$.

Let us indicate with $\mathcal{L}$ the set of all gambles defined on Ω. Then, $\mathcal{L} = \mathbb{R}^n$. Given $g, h \in \mathcal{L}$, let $g \geq h$ mean that $g_i \geq h_i$ for every $i \in \{1, ..., n\}$, let $g \gneq h$ mean that $g \geq h$ and $g_k > h_k$ for some $k \in \{1, ..., n\}$ and finally let $g > h$ mean that $g_i > h_i$ for every $i \in \{1, ..., n\}$. Let us indicate also with T the set of all non-negative gambles $T := \{g \in \mathcal{L} : g \geq 0\}$ and with F the set of negative ones $F := \{g \in \mathcal{L} : g < 0\}$.

Let us consider again the rationality axioms D1–D5 listed before, often assumed in literature for a set of desirable gambles $\mathcal{D} \subseteq \mathcal{L}$. If a set of desirable gambles $\mathcal{D}$ satisfies axioms D1–D4, we call it *coherent set of gambles*.[7] If it satisfies also D5, we call it *maximal coherent set of gambles*.

In particular, axioms D1–D4 lead to the concept of *natural extension* for a set of gambles.

Definition 1 (Natural extension). Given a set $\mathcal{D} \subseteq \mathcal{L}$, we call $\overline{\mathcal{E}(\mathcal{D})} := \overline{\mathsf{posi}(\mathcal{D} \cup T)}$ its *natural extension*, where, given $\mathcal{D}' \subseteq \mathcal{L}$:

$$\mathsf{posi}(\mathcal{D}') := \left\{ \sum_{j=1}^{r} \lambda_j f_j : f_j \in \mathcal{D}',\ \lambda_j \in \mathbb{R}_+, r \geq 1 \right\}, \tag{1}$$

and where $\overline{\mathcal{D}'}$ represents the closure of $\mathcal{D}'$ under the supremum norm topology [5, Section 3.7.2–3.7.4] or under the usual topology of $\mathbb{R}^n$ [5, Appendix D].

Given a set $\mathcal{D} \subseteq \mathcal{L}$, its natural extension $\overline{\mathcal{E}(\mathcal{D})}$ represents the set of all and only the gambles that an agent, rational in the sense of D1–D4, should regard as acceptable once they state that those in $\mathcal{D}$ are. Hence, when $\overline{\mathcal{E}(\mathcal{D})}$ is coherent - i.e., when it satisfies D2 - $\overline{\mathcal{E}(\mathcal{D})}$ is the smallest coherent set of gambles containing $\mathcal{D}$, see [5].

We do not want to introduce here a formal treatment of the logical view of desirability, for which we refer the reader to, for instance, [27, 28]. Nevertheless, it is possible to notice that the natural extension of

[6] With 'positive additivity' we mean that $\mathcal{D}$ is closed with respect to the sum with a non-negative gamble, see the next Section 6.

[7] In literature a set $\mathcal{D}$ satisfying axioms D1–D4 is usually called *coherent set of almost desirable gambles*, see for example [26]. Since we do not need to emphasize it, we drop the term *almost* to simplify the notation.

a set $\mathcal{D}$ can also be seen as a form of *deductive closure* for $\mathcal{D}$, here understood as the outcome of a *closure operator*, in this case $\overline{\mathcal{E}(\cdot)}$, applied to $\mathcal{D}$.[8] Moreover, it is also possible to observe that the set T plays the role of *tautologies* in logic. Indeed, it is formed by the gambles that should always be acceptable for any agent respecting the rationality axioms D1–D4. Analogously, any element of F represents the *falsum.* They are, in fact, elements that added to $\mathcal{D}$ satisfying D1–D4, making it to collapse on the set of all gambles $\mathcal{L}$. For this reason, we chose to call these sets T (tautologies) and F (falsum) respectively, and we denoted with the same name also the two associated rationality axioms D1 and D2.

When $\mathcal{D}$ is finite, $\overline{\mathcal{E}(\mathcal{D})}$ is said to be *finitely generated.* We have, in particular, the following definition [5, Chapter 4].

Definition 2 (Finitely generated coherent set). A coherent set of gambles $\mathcal{D} \subseteq \mathcal{L}$ is called *finitely generated (coherent set of gambles)* if and only if $\mathcal{D} = \overline{\mathcal{E}(\mathcal{A})} := \overline{\mathsf{posi}(\mathcal{A} \cup T)}$, where $\mathcal{A}$ is a finite set of gambles.

A coherent set of gambles can also be equivalently expressed in probabilistic terms. To see this, let us start by introducing the concept of *lower prevision.* We use the term *lower prevision* to indicate every real-valued functional $\underline{P}(\cdot)$ defined on $\mathcal{L}$.[9] An important class of lower previsions are the *coherent* ones.

Definition 3 (Coherent lower prevision). A lower prevision $\underline{P}(\cdot)$ is called *coherent* if and only if it satisfies the following properties. For every $g, h \in \mathcal{L}$:

L1. $\underline{P}(g) \geq \min_{\omega\in\Omega} g(\omega)$;

L2. $\underline{P}(\lambda g) = \lambda \underline{P}(g)$, for every $\lambda \in \mathbb{R}^*_+$;

L3. $\underline{P}(g + h) \geq \underline{P}(g) + \underline{P}(h)$.

There is a one-to-one correspondence between coherent sets of gambles and coherent lower previsions. Indeed, given a coherent lower prevision $\underline{P}(\cdot)$, the set:

$$\mathcal{D}_{\underline{P}} := \{g \in \mathcal{L} : \underline{P}(g) \geq 0\} \tag{2}$$

satisfies axioms D1–D4. Vice versa, given a coherent set of gambles $\mathcal{D}$, the lower prevision defined as:

$$(\forall g \in \mathcal{L})\ \underline{P}_{\mathcal{D}}(g) := \sup\{c \in \mathbb{R} : g - c \in \mathcal{D}\} \tag{3}$$

is coherent. Moreover, both procedures commute, see [5]. So, given a coherent set $\mathcal{D}$, the equivalent coherent lower prevision $\underline{P}_{\mathcal{D}}(\cdot)$ represents for every g the supremum buying price the agent, whose gambles acceptance is characterized by $\mathcal{D}$, is disposed to set for g.

If a coherent lower prevision is in particular *self-conjugate* it becomes a linear functional called *linear prevision.*

Definition 4 (Linear prevision). Consider a coherent lower prevision $\underline{P}(\cdot)$. If $\underline{P}(g) = -\underline{P}(-g)$ for some $g \in \mathcal{L}$, we then call the common value the *prevision* of g and we denote it by $P(g)$. If this happens for all $g \in \mathcal{L}$, we then call the functional $P(\cdot)$ a *linear prevision.*

It is possible to show that, given a linear prevision $P(\cdot)$, the set $\mathcal{D}_P$ obtained from $P(\cdot)$ through Eq.(2) is maximal coherent. Vice versa, given a maximal coherent set $\mathcal{D}$, the coherent lower prevision $\underline{P}_{\mathcal{D}}(\cdot)$ obtained from $\mathcal{D}$ through Eq. (3) is linear, see [5].

Every coherent lower prevision is the lower envelope of a set of dominating linear previsions [5]:

$$(\forall g \in \mathcal{L})\ \underline{P}(g) = \min\{P(g) : P(\cdot) \in \mathcal{M}(\underline{P})\} \tag{4}$$

[8] For the definition of *closure operator* used here, see [29].

[9] It can be defined also on different classes of gambles. We consider here a simplified version of this notion. For more details, see for example [5].

where $\mathcal{M}(\underline{P})$ is defined as:

$$\mathcal{M}(\underline{P}) \coloneqq \{P(\cdot) \text{ linear prevision } : (\forall g \in \mathcal{L})\ \underline{P}(g) \leq P(g)\}. \tag{5}$$

It turns out that, in particular, $\mathcal{M}(\underline{P})$ for a coherent lower prevision $\underline{P}(\cdot)$ is non-empty, closed[10] and convex. Vice versa, every non-empty, closed and convex set of linear previsions $\mathcal{M}$ has as lower envelope a coherent lower prevision $\underline{P}(\cdot)$ such that $\mathcal{M} = \mathcal{M}(\underline{P})$ [5, Theorem 3.6.1]. Hence, coherent lower previsions are in a one-to-one correspondence with non-empty, closed and convex sets of linear previsions. Moreover, the same results can be obtained working only with the set of extreme points of $\mathcal{M}(\underline{P})$, denoted as $\text{ext}(\mathcal{M}(\underline{P}))$ [5, Theorem 3.6.2]. Therefore, to define a coherent lower prevision $\underline{P}(\cdot)$, it suffices to specify $\text{ext}(\mathcal{M}(\underline{P}))$:

$$(\forall g \in \mathcal{L})\ \underline{P}(g) = \min\{P(g) : P(\cdot) \in \text{ext}(\mathcal{M}(\underline{P}))\}.$$

From these considerations, it follows that a coherent set of gambles can be directly characterized in terms of the extreme points of $\mathcal{M}(\underline{P}_{\mathcal{D}})$ [5]:

$$\mathcal{D} = \{g \in \mathcal{L} : P(g) \geq 0, \forall P(\cdot) \in \text{ext}(\mathcal{M}(\underline{P}_{\mathcal{D}}))\}. \tag{6}$$

In particular, if $\mathcal{D}$ is finitely generated, $\text{ext}(\mathcal{M}(\underline{P}_{\mathcal{D}}))$ is finite and vice versa, see the following proof of Proposition 3 in Appendix or [5, Chapter 4]. If a set $\mathcal{D}$ is instead maximal coherent, it corresponds to a unique linear prevision, as previously noticed:

$$\mathcal{D} = \{g \in \mathcal{L} : P_{\mathcal{D}}(g) \geq 0\}. \tag{7}$$

Any linear prevision corresponds essentially to an expectation operator taken with respect to a finitely additive probability which, in turn, can be obtained by making the restriction of the linear prevision to indicators of events. Therefore, it is possible to regard $\mathcal{M}(\underline{P})$ of a coherent lower prevision $\underline{P}(\cdot)$ also as a set of probabilities (called *credal set*). Hence, coherent lower previsions are in a one-to-one correspondence also with sets of finitely additive probabilities.

Coherent sets of gambles can be equivalently represented in terms of preference relations between gambles. Indeed, starting from a set of desirable gambles $\mathcal{D}$, it is possible to define a preference relation over gambles as follows:

$$\forall g, h \in \mathcal{L},\ g \succsim_{\mathcal{D}} h \iff g - h \in \mathcal{D}. \tag{8}$$

If $\mathcal{D}$ is coherent, $\succsim_{\mathcal{D}}$ is called *coherent* and it is characterized by the following properties:

R1. $-1 \not\succsim_{\mathcal{D}} 0$ [avoiding sure loss];

R2. if $g \geq h$ then $g \succsim_{\mathcal{D}} h$, for any $g, h \in \mathcal{L}$ [monotonicity];

R3. if $g \succsim_{\mathcal{D}} h$ and $\lambda \in \mathbb{R}^*_+$ then $\lambda g \succsim_{\mathcal{D}} \lambda h$, for any $g, h \in \mathcal{L}$ [positive homogeneity];

R4. if $g \succsim_{\mathcal{D}} h$ and $h \succsim_{\mathcal{D}} l$ then $g \succsim_{\mathcal{D}} l$, for any $g, h, l \in \mathcal{L}$ [transitivity];

R5. if $g + \delta \succsim_{\mathcal{D}} h$ for every $\delta \in \mathbb{R}^*_+$ then $g \succsim_{\mathcal{D}} h$, for any $g, h \in \mathcal{L}$ [continuity];

R6. $g \succsim_{\mathcal{D}} h$ if and only if $g - h \succsim_{\mathcal{D}} 0$, for any $g, h \in \mathcal{L}$ [cancellation].

If it is also maximal coherent, $\succsim_{\mathcal{D}}$ satisfies in addition the following property:

$$(\forall g, h \in \mathcal{L})\ g \succsim_{\mathcal{D}} h \text{ or } h \succsim_{\mathcal{D}} g \text{ [completeness]}.$$

Vice versa, if a preference relation over gambles $\succsim$ satisfies properties R1-R6, the set of gambles $\mathcal{D}$ constructed as:

$$\mathcal{D}_{\succsim} \coloneqq \{f \in \mathcal{L} : (\exists g, h \in \mathcal{L})\ f = g - h, g \succsim h\} \tag{9}$$

[10] Under the *weak* topology*, which is the smallest topology such that all the evaluation functionals given by $g(P) \coloneqq P(g)$, where $g \in \mathcal{L}$, are continuous, see [5].

is coherent. If $\succsim$ satisfies also completeness, $\mathcal{D}_{\succsim}$ constructed from $\succsim$ through Eq.(9) is maximal coherent.

In the following sections we consider different sets of rationality axioms that a set $\mathcal{D} \subseteq \mathcal{L}$ could satisfy (both standard and other ones). In particular, we show that establishing if an agent respects these axioms on the basis of a finite set $\mathcal{A}$ of acceptable gambles and a finite set $\mathcal{R}$ of rejectable gambles is always equivalent to solve a binary classification problem where the family of classifiers involved changes with the set of axioms considered. In the case of standard rationality axioms, this further interpretation of coherence follows directly from the existing link between a coherent set and a set of linear previsions, see for example the summary provided in the actual section or [5]. In Section 3 and Section 4 therefore, we start by reformulating these standard cases setting the basic framework used then to analyze different definitions of coherence given in Section 5 and Section 6. In particular, in Section 4, we introduce a procedure to reformulate all the nonlinear classification problems emerging in our analysis as linear ones in higher-dimensional spaces.

We warn the reader that, in the rest of the paper, the assumption of finite sets of acceptance/rejection statements is fundamental. We limit ourselves to this case, in fact, to be closer to standard classification tasks considered in machine learning assuming finite training sets. It is important to notice however that in our cases, unlike the classical set-up, the classification of the training set $(\mathcal{A}, \mathcal{R})$ is always accompanied by the classification of the two non-finite sets T and F. In this work therefore, we extend the standard framework of classification problems considering also infinite logical constraints sets T and F to be respected. To the best knowledge of the authors, this type of classification tasks has only been studied before in [30] for support vector machines.

3. Standard maximal coherence

Suppose to have a modeller who considers an agent *rational* if they respect axioms D1–D5 (so in particular D3), which is, if they have a set of desirable gambles satisfying D1–D5. As seen in Section 2, the beliefs of an agent rational in this sense are expressed with a single probability mass function. Equivalently, the agent is disposed to accept/reject gambles on the basis of a linear prevision, i.e., an expected value operator, see Eq.(7).

Suppose that, however, the only information about the agent the modeller has are two finite sets of gambles $\mathcal{A}$ and $\mathcal{R}$ that the agent is respectively willing to accept and reject. The modeller will then consider the agent rational on the basis of these two finite sets if and only if there exists a set of gambles $\mathcal{D}$ that is maximal coherent and such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. Notice that it is also possible that the agent is *not* rational in the sense of D1–D5 and nevertheless they are judged rational by the modeller. The idea, however, is that by querying the agent on additional gambles, i.e., by augmenting the dimension of the sets $\mathcal{A}$ and/or $\mathcal{R}$, it shall be possible to disentangle the situation, see Example 1. Similar considerations can be made for the other rationality definitions analysed later on.

To face the problem of establishing the existence of such a maximal coherent set, we re-interpret the equivalence between maximal coherent sets of gambles and linear previsions/probability mass functions, as a binary linear classification problem. Indeed, if we consider two finite sets $\mathcal{A}$ and $\mathcal{R}$, there exists a maximal coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$ if and only if there exists a linear prevision $P(\cdot)$ such that $P(g) \geq 0$ for every $g \in \mathcal{A}$ and $P(g) < 0$ for every $g \in \mathcal{R}$, see Eq.(7). Every linear prevision, however, is an expected value operator, i.e., a linear operator $P(g) = g^\top \beta$ for every $g \in \mathbb{R}^n$, where components of $\beta \in \mathbb{R}^n$ form a probability mass function on Ω: $\beta \gneq 0$ and $\sum_{i=1}^n \beta_i = 1$. Thus, the agent is rational if and only if there exists a binary linear classifier classifying gambles on the basis of the sign of a linear prevision, which labels $\mathcal{A}$ as T, in such a way to guarantee that $P(g) = g^\top \beta \geq 0$ for every $g \in \mathcal{A}$, and $\mathcal{R}$ as F. More formally, let us consider the class of binary linear classifiers $LC(\cdot)$ defined as:

$$LC(g) := \begin{cases} 1 & \text{if } g^\top \beta \geq 0, \\ -1 & \text{otherwise}, \end{cases} \tag{10}$$

for every $g \in \mathcal{L}$, with $\beta \in \mathbb{R}^n$.[11]

[11] We assume without loss of generality that the binary classifier has as classes 1 and -1.

Definition 5 (Linear separability). A pair of sets of gambles (A, B) is *linearly separable* if and only if there exists a binary linear classifier $LC(\cdot)$ of type (10) such that $LC(A) = 1$ and $LC(B) = -1$.[12] We indicate the set of these classifiers with $\mathrm{LC}(A, B)$.[13]

Notice that if (A, B) is *nonlinearly separable*, $\mathrm{LC}(A, B)$ is empty. The same consideration is valid for the analogous separability concepts defined later on.

The following result is valid.

Proposition 1. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a maximal coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is linearly separable.*

The proof follows the previous reasoning. If there is a maximal coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, then the linear prevision $P_{\mathcal{D}}(\cdot)$ defined from $\mathcal{D}$ through Eq.(3) is such that $P_{\mathcal{D}}(g) \geq 0$ for every $g \in \mathcal{D}$ and $P_{\mathcal{D}}(g) < 0$ for every $g \notin \mathcal{D}$, see Eq.(7). Therefore, the binary linear classifier $LC(\cdot)$ of type (10) classifying gambles g on the basis of the sign of $P_{\mathcal{D}}(g)$, is a classifier $LC(\cdot) \in \mathrm{LC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$. Vice versa, if the pair $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is linearly separable, all the linear classifiers $LC(\cdot) \in \mathrm{LC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$ belong also to $\mathrm{LC}(\mathcal{A}, \mathcal{R})$ and are characterized by a parameter $\beta \gneq 0$. Indeed, consider a linear classifier $LC(\cdot) \in \mathrm{LC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$ characterized by a parameter $\beta \in \mathbb{R}^n$. Clearly, $LC(\cdot) \in \mathrm{LC}(\mathcal{A}, \mathcal{R})$. Moreover, if $\beta = 0$ then $LC(F) = 1$ that is a contradiction. Hence, $\beta \neq 0$. If β has at least a component $\beta_l < 0$ with $l \in \{1, ..., n\}$, it classifies the gamble $t \in T$, defined as $t_l = 1$ and $t_i = 0$ for every $i \in \{1, ..., n\}$ with $i \neq l$, as -1 and this is again a contradiction. Therefore, for every binary linear classifier $LC(\cdot)$ in $\mathrm{LC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$ characterized by the coefficient β, it is possible to construct a linear prevision $P(\cdot)$ defined as $P(g) := g^\top \beta'$ for every $g \in \mathcal{L}$, where β' is the normalized version of β. Hence, $\mathcal{D} := \{g \in \mathcal{L} : LC(g) = 1\} = \{g \in \mathcal{L} : P(g) \geq 0\}$ is a maximal coherent set of gambles that contains $\mathcal{A}$ and has empty intersection with $\mathcal{R}$.

From the latter proof, it is also possible to deduce the following result.

Proposition 2. *Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$. Every classifier $LC(\cdot) \in LC(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is a classifier $LC(\cdot) \in LC(\mathcal{A}, \mathcal{R})$ with coefficient $\beta \gneq 0$, and vice versa.*

From Proposition 1 and Proposition 2 it also follows the following corollary.

Corollary 1. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a maximal coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is linearly separable and there exists a classifier $LC(\cdot) \in LC(\mathcal{A}, \mathcal{R})$ with coefficient $\beta \gneq 0$.*

Thus, checking if an agent is rational in the sense of D1–D5 on the basis of a finite set of acceptable gambles $\mathcal{A}$ and a finite set of rejectable gambles $\mathcal{R}$, boils down to linearly separate only the two finite sets $\mathcal{A}$ and $\mathcal{R}$.

Notice that, as also follows from the proof of Proposition 1, if a binary linear classifier of type (10) has coefficient $\beta \gneq 0$, i.e., satisfying the constraint provided by Proposition 2 and Corollary 1, it is always possible to normalize it and interpret its components as the values of a probability mass function on Ω. Clearly, the converse is also true: starting from any probability mass function on Ω, it is always possible to interpret its values as the components of the coefficient of a binary linear classifier satisfying the constraint of Proposition 2 and Corollary 1. From now on therefore, we assume without loss of generality that if the coefficient β of a binary linear classifier satisfies $\beta \gneq 0$, it is normalized. These considerations lead also to another one-to-one correspondence between regions classified as 1 by those classifiers and maximal coherent sets $\mathcal{D}$, see Section 2.

The following diagram summarizes the implications highlighted above. In particular, we reported in black the implications known in literature (Section 2) and in green the ones established by us.

[12] With a little abuse of notation, with $LC(\mathcal{K}) = c$ for $\mathcal{K} \subseteq \mathcal{L}$, $c \in \{-1, 1\}$, we mean $LC(g) = c$, for all $g \in \mathcal{K}$. We will use the same notation also for the other types of binary classifiers considered later on.

[13] In this definition, we assume the classification constraints to hold only for non-empty sets. In particular, if $A = B = \emptyset$ we assume the pair (A, B) to be linearly separable and $\mathrm{LC}(A, B)$ to be the whole set of binary linear classifiers of type (10).

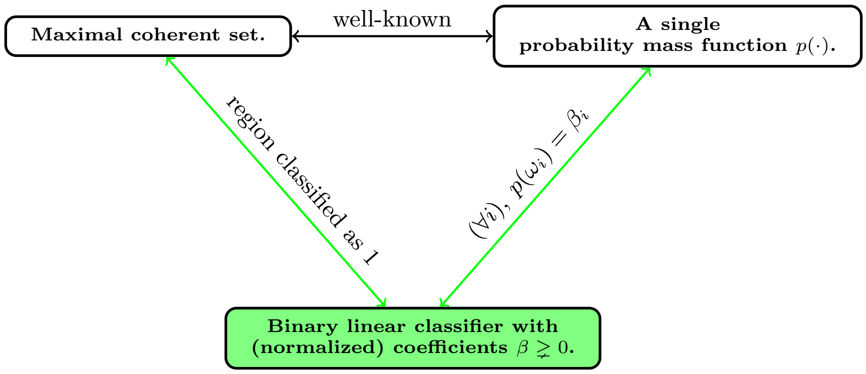

Figure 1: Diagram showing equivalent models for representing maximal coherent sets of gambles.

Given these considerations, Corollary 1 can be rewritten in the following way

Corollary 2. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a maximal coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if equivalently:*

- *there exists a binary linear classifier $LC(\cdot) \in LC(\mathcal{A}, \mathcal{R})$ with (normalized) coefficient $\beta \gneq 0$;*
- *there exists a probability mass function on Ω, $p(\cdot)$, such that $E_p(g) \geq 0$ for every $g \in \mathcal{A}$ and $E_p(g) < 0$ for every $g \in \mathcal{R}$.*

However, given two finite sets $\mathcal{A}$ and $\mathcal{R}$, there can be more than one classifier $LC(\cdot) \in \mathrm{LC}(\mathcal{A}, \mathcal{R})$ with (normalized) coefficient $\beta \gneq 0$. How can we learn a unique classifier of this type? An idea can be to learn the one that leads to make the minimal assumptions on the agent's dispositions to accept gambles, i.e., the one that identifies the minimal acceptance region $\{g \in \mathcal{L} : LC(g) = 1\}$. However, all the regions $\{g \in \mathcal{L} : LC(g) = 1\}$ are minimal, indeed they all are maximal coherent sets of gambles and no maximal coherent set is greater than another one with respect to inclusion.[14].

We need therefore other criteria to learn a classifier. Another one could be to get, if it exists, the binary linear classifier leading to the probability distribution with the minimal Kullback–Leibler divergence from the uniform one:[15]

$$LC(\cdot) \in \mathrm{LC}(\mathcal{A}, \mathcal{R}) \text{ with (normalized) } \beta \gneq 0, \text{ minimizing: } \sum_{i \in \{1,..,n\}} \beta_i \log_2(n\beta_i),$$

but other criteria could be used.

This learning process can lead to a binary linear classifier whose coefficient corresponds to a probability distribution different from the real beliefs of the agent, as the following example shows. However, the idea here is again that augmenting the dimension of the finite sets of acceptable and/or rejectable gambles declared by the agent can solve the problem. Similar considerations can be made for the other classification problems considered in the next sections.

The following example illustrates the results obtained in this section considering a specific numerical case and reports a numerical optimization procedure to solve the classification problem and find the coefficient of the classifer.

Example 1. Consider a coin tossing experiment whose outcomes are h, Heads, and t, Tails. A gamble g in this case has two components $g(h) = g_1$ and $g(t) = g_2$. We can think that outcomes of gambles correspond

[14]Let us consider two maximal coherent sets $\mathcal{D}, \mathcal{D}' \subseteq \mathcal{L}$. If $\mathcal{D} \neq \mathcal{D}'$, then $P_{\mathcal{D}}(\cdot) \neq P_{\mathcal{D}'}(\cdot)$, hence there exists $g \in \mathcal{L}$ such that $P_{\mathcal{D}}(g) < 0 < P_{\mathcal{D}'}(g)$. Therefore, $g \in \mathcal{D}' \setminus \mathcal{D}$ and $\mathcal{D}' \not\subseteq \mathcal{D}$. Analogously, we can prove that $\mathcal{D} \not\subseteq \mathcal{D}'$

[15]This approach leads to a unique solution. Indeed, the discrete probability distribution that minimizes the Kullback-Leibler divergence with respect to the uniform one corresponds to the distribution that maximizes the Shannon's entropy. The latter then is unique if we consider linear constraints, see for example [31].

to amounts in thousand of dollars. Therefore, if an agent is disposed to accept a gamble g, they commit themselves to receive/pay g_1 thousands of dollars if the result of the experiment is Heads and g_2 thousands of dollars if Tails.

In this framework, let us consider Alice (an agent) who is disposed to accept/reject gambles on the basis of the linear prevision $P_1(\cdot)$, defined as $P_1(g) = E_{\{2/3,1/3\}}(g)$,[16] for every $g \in \mathcal{L}$ (this can be interpreted as she believes that the probability of the event Heads is 2/3). That is, she is characterized by the following maximal coherent set of gambles:

$$\mathcal{D}_1 := \{g \in \mathcal{L} : P_1(g) \geq 0\} := \{g \in \mathcal{L} : E_{\{2/3,1/3\}}(g) \geq 0\}.$$

Then, for example, she is disposed to accept gambles in $\mathcal{A}_1 := \{[-1,2]^\top, [2,-1]^\top, [1,-1]^\top, [4,-2]^\top, [1,-0.5]^\top\}$, and she is disposed to reject gambles in $\mathcal{R}_1 := \{[-3,2]^\top\}$.

In Figure 2 it is possible to find a graphical representation of the gambles contained in $\mathcal{A}_1$ and $\mathcal{R}_1$ in the two-dimensional Cartesian coordinate system representing gambles' values respectively on the alternatives h and t. Gambles in $\mathcal{A}_1$ are denoted with (blue) points, gambles in $\mathcal{R}_1$ are instead denoted with (red) triangles.

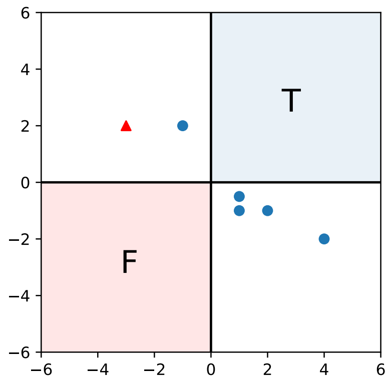

Figure 2: Gambles contained in $\mathcal{A}_1$ and in $\mathcal{R}_1$.

Suppose now that the only information a modeller, Bob, who judges an agent rational if they respect D1–D5, has on Alice is represented by the two sets $\mathcal{A}_1$, $\mathcal{R}_1$ just defined. Then, Bob believes that Alice is rational because, for example, the binary linear classifier $LC_1(\cdot)$ of type (10) characterized by the coefficient $\beta^{(1)} := \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$, is such that $LC_1(\cdot) \in \mathrm{LC}(\mathcal{A}_1, \mathcal{R}_1)$. Notice that this is the linear classifier in $\mathrm{LC}(\mathcal{A}_1, \mathcal{R}_1)$ with (normalized) coefficient corresponding to the probability distribution closest to the uniform one with respect to the Kullback–Leibler divergence (the divergence between the two distributions is indeed 0).

[16] We indicate with $E_{\{p,1-p\}}(g)$, the expected value of a gamble g calculated with respect to the probability mass function that assigns probability p to Heads and $(1-p)$ to Tails.

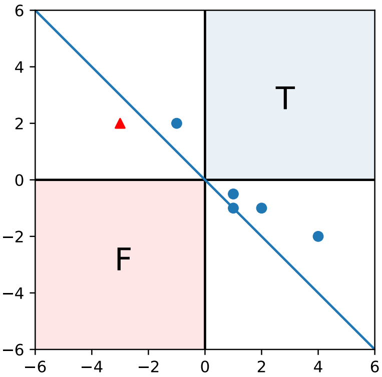

Figure 3: Graphical representation of the decision boundary of the linear classifier $LC_1(\cdot)$.

However, the region $\{g \in \mathcal{L} : LC_1(g) = 1\}$ *does not* correspond to Alice's set of desirable gambles. The idea, however, is that by querying Alice on additional gambles, that is by augmenting the dimension of the sets $\mathcal{A}_1$ and/or $\mathcal{R}_1$, it shall be possible to find a binary linear classifier $LC(\cdot) \in \mathrm{LC}(\mathcal{A}_1 \cup T, \mathcal{R}_1 \cup F)$ whose region classified as 1 corresponds to Alice's set of desirable gambles.

For example, suppose that Alice adds the gamble $[1, -2]^\top$ to $\mathcal{A}_1$ forming a new set of acceptable gambles $\mathcal{A}_1' := \mathcal{A}_1 \cup \{[1, -2]^\top\}$. Then, the binary linear classifier $LC_2(\cdot) \in \mathrm{LC}(\mathcal{A}_1' \cup T, \mathcal{R}_1 \cup F)$ corresponding to the probability distribution closest to the uniform one with respect to the Kullback-Leibler divergence can be found numerically by solving the convex optimization problem

$$\begin{cases} \max_{\{\beta \in \mathbb{R}^2\}} (-\sum_{i=1}^2 \beta_i \log_2 \beta_i) \\ s.t. \\ -g^\top \beta \le 0 \quad \forall g \in \mathcal{A}'_1, \\ \;\; g^\top \beta \le -\epsilon \;\; \forall g \in \mathcal{R}_1, \\ \displaystyle\sum_{i=1}^{2} \beta_i = 1, \\ \beta_1, \beta_2 \ge 0, \end{cases}$$

where ϵ is a small positive number. The solution is $\beta^{(2)} := \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$, which is also the coefficient of the unique (with normalized coefficient) binary linear classifier in $\mathrm{LC}(\mathcal{A}_1' \cup T, \mathcal{R}_1 \cup F)$. The decision boundary of this latter classifier is shown in Figure 4. In this case, however, the region $\{g \in \mathcal{L} : LC_2(g) = 1\}$ *corresponds* to Alice's set of desirable gambles and the components of $\beta^{(2)}$ coincide with the values of her probability mass function on Ω.

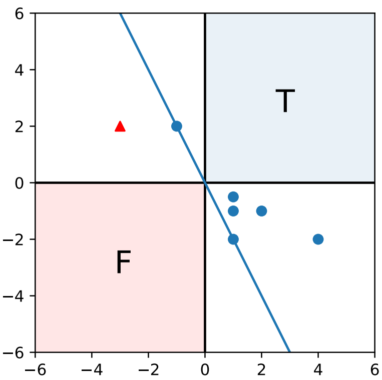

Figure 4: Graphical representation of the decision boundary of the linear classifier $LC_2(\cdot)$.

It is also possible that Bob thinks an agent is rational in the sense of D1–D5 when, actually, they are not.

Consider in this regard another agent Valerie, who is initally disposed to accept $\mathcal{A}_1$ and reject $\mathcal{R}_1$, as well as Alice, and hence is judged rational by Bob. Assume however that she, at a later time, declares to be disposed to reject also the gambles $[-2,4]^\top, [2,-2]^\top$ obtaining a new set of rejectable gambles $\mathcal{R}_1' \coloneqq \mathcal{R}_1 \cup \{[-2,4]^\top, [2,-2]^\top\}$. This leads to a *nonlinearly separable* pair of finite sets of acceptable and rejectable gambles $(\mathcal{A}_1, \mathcal{R}_1')$. The latter sets of gambles are represented in the following Figure 5. As usual, gambles in $\mathcal{A}_1$ are denoted with (blue) points and gambles in $\mathcal{R}_1'$ are denoted with (red) triangles. Now Bob is sure that she is not rational according to D1–D5.

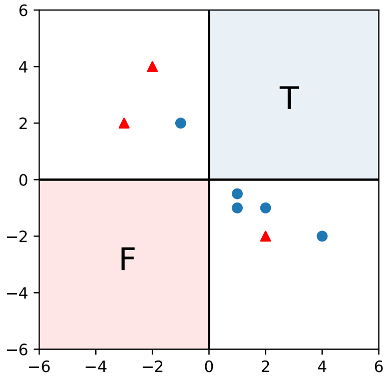

Figure 5: Example of a nonlinearly separable pair of finite sets of gambles.

4. Standard coherence

Suppose now that the modeller considers as *rational* an agent who respects D1–D4 (so, again, D3). As seen in Section 2, the beliefs of an agent rational in this sense are expressed by a set of probability mass functions. Equivalently, the agent is disposed to accept/reject gambles on the basis of a coherent lower prevision or a set of linear previsions, see Eq.(2) and Eq.(6).

Suppose however that, as before, the only information the modeller possesses on the agent consist of a finite set of gambles $\mathcal{A}$, the agent is willing to accept, and a finite set of gambles $\mathcal{R}$, they are willing to reject. The agent will then be considered rational by the modeller if and only if there exists a set of gambles $\mathcal{D}$ that satisfies D1–D4 such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. Establishing the existence of such a set is a problem already analyzed in [25]. While, however, they explicitly analyze the properties that $\mathcal{A}$ and $\mathcal{R}$, which in their case can also be *non-finite*, must satisfy so that such a coherent set exists, we redefine the latter problem as a classification task. Thus, by solving the latter we can solve the former. To do so, we re-interpret the duality between coherent sets of gambles and convex and closed sets of linear previsions/probability mass functions as a binary classification task. This time, however, the resulting classification problem is *nonlinear*. Indeed, if we consider two finite sets $\mathcal{A}$ and $\mathcal{R}$, there exists a coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$ if and only if there is a non-empty closed and convex set of linear previsions $\{P_j(\cdot)\}_{j \in J}$ such that $P_j(g) \geq 0$ for every $g \in \mathcal{A}$ and every $j \in J$, and for every $g \in \mathcal{R}$ there exists $k \in J$ such that $P_k(g) < 0$, see again Eq.(6). We will see later on that, since $\mathcal{A}$ is finite or empty, we can consider only finite sets J. The problem therefore is equivalent to solving a *binary piecewise linear* classification task.

Definition 6 (Binary piecewise linear classifier). We use the term *binary piecewise linear classifier* to denote a classifier $PLC(\cdot)$ defined as follows:

$$PLC(g) \coloneqq \begin{cases} c_1 & \text{if } g^\top \beta_j \geq 0 \text{ for all } j \in \{1, ...N\}, \\ c_2 & \text{otherwise,} \end{cases} \tag{11}$$

for every $g \in \mathcal{L}$, with labels c_1, c_2 and $\beta_j \in \mathbb{R}^n$ for all j, $N \geq 1$.

Without loss of generality, in what follows we limit ourselves to binary piecewise linear classifiers with $c_1 = 1$ and $c_2 = -1$.

Definition 7 (Piecewise linear separability). A pair of sets of gambles (A, B) is *piecewise linearly separable* if and only if there exists a binary piecewise linear classifier $PLC(\cdot)$ such that $PLC(A) = 1$ and $PLC(B) = -1$. We indicate the set of these classifiers with $\mathrm{PLC}(A, B)$.

Now we can state the main result of this section. All the proofs are in Appendix.

Proposition 3. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is piecewise linearly separable.*

Similarly to before, it is then possible to transform the problem of classifying the two non-finite sets of gambles T and F into a set of constraints on the coefficients of the classifier, as the following Proposition 4 shows. Hence, establishing the rationality of an agent in the sense of D1–D4 on the basis of two finite sets of gambles $\mathcal{A}$ and $\mathcal{R}$, boils down to classifying only these two finite sets.

Proposition 4. *Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$. Every classifier $PLC(\cdot) \in PLC(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is a classifier $PLC(\cdot) \in PLC(\mathcal{A}, \mathcal{R})$ with coefficients $\{\beta_j\}_{j=1}^N$ such that $\beta_j \gneq 0$ for every $j \in \{1, ..., N\}$, and vice versa.*

Corollary 3. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is piecewise linearly separable and there exists a classifier $PLC(\cdot) \in PLC(\mathcal{A}, \mathcal{R})$ with coefficients $\{\beta_j\}_{j=1}^N$ such that $\beta_j \gneq 0$ for every $j \in \{1, ..., N\}$.*

Notice that, as also follows from the proof of Proposition 3, if a binary piecewise linear classifier has coefficients $\{\beta_j\}_{j=1}^N$ satisfying the constraints provided by Proposition 4 and Corollary 3 ($\forall j,\ \beta_j \gneq 0$), it is always possible to normalize each β_j and interpret its (normalized) components as the values of a probability mass function on Ω. Clearly, the converse is also true: starting from any finite set of probability mass functions on Ω, it is always possible to interpret the values of each one of them as the components of a coefficient of a binary piecewise linear classifier satisfying the constraints of Proposition 4 and Corollary 3. From now on therefore, we assume without loss of generality that if coefficients $\{\beta_j\}_{j=1}^N$ of a binary piecewise linear classifier satisfy $\beta_j \gneq 0$ for every j, they are normalized. These considerations lead also to a one-to-one correspondence between regions classified as 1 by those classifiers and finitely generated coherent sets $\mathcal{D}$ (through the extreme points of their credal sets), see Section 2. *Non* finitely generated coherent sets of gambles instead *cannot*, in general, be represented through a binary piecewise linear classifier.[17]

The following diagram summarizes the implications highlighted above. In particular, we reported in black the implications known in literature and in green the ones established by us.

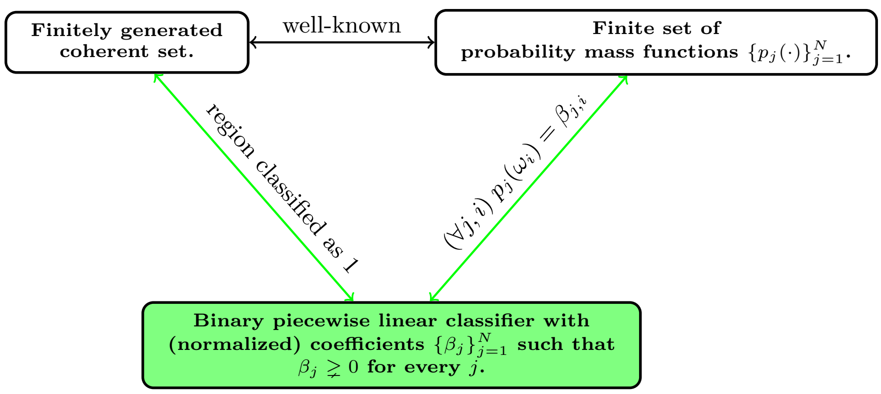

Figure 6: Diagram showing equivalent models for representing finitely generated coherent sets of gambles.

[17] Consider for example the half right circular cone in $\mathbb{R}^3$ of non-strictly negative gambles with axis $x = y = z$, vertex at the origin and semi-vertical angle of $\pi/3$ radians.

Given these considerations, Corollary 3 can be rewritten in the following way.

Corollary 4. *Given a pair of finite sets of gambles* $(\mathcal{A}, \mathcal{R})$, *there exists a coherent set (not necessarily finitely generated)* $\mathcal{D} \subseteq \mathcal{L}$, *such that* $\mathcal{D} \supseteq \mathcal{A}$ *and* $\mathcal{D} \cap \mathcal{R} = \emptyset$, *if and only if equivalently:*

- *there exists a binary piecewise linear classifier* $PLC(\cdot) \in PLC(\mathcal{A}, \mathcal{R})$ *with (normalized) coefficients* $\{\beta_j\}_{j=1}^N$ *such that* $\beta_j \gneq 0$ *for all* j;
- *there exists a finite set of probability mass functions on* Ω, $\{p_j(\cdot)\}_{j=1}^N$, *such that* $E_{p_j}(g) \geq 0$ *for every* $g \in \mathcal{A}$, $j \in \{1, ..., N\}$ *and, for every* $g \in \mathcal{R}$, *there exists* $k \in \{1, ..., N\}$ *such that* $E_{p_k}(g) < 0$;
- *there exists a finitely generated coherent set* $\mathcal{D} \subseteq \mathcal{L}$ *such that* $\mathcal{D} \supseteq \mathcal{A}$ *and* $\mathcal{D} \cap \mathcal{R} = \emptyset$ *(e.g.,* $\mathcal{D} = \overline{\mathcal{E}(\mathcal{A})}$*).*

Since $\overline{\mathcal{E}(\mathcal{A})}$ is the smallest coherent set containing $\mathcal{A}$, a binary piecewise linear classifier $PLC(\cdot)$ (satisfying in particular the usual constraints) such that $\{g \in \mathcal{L} : PLC(g) = 1\} = \overline{\mathcal{E}(\mathcal{A})}$ corresponds to make the minimal set of assumptions about the agent's dispositions to accept gambles.

The following example illustrates the results obtained in this section considering a specific numerical case and reports a numerical optimization procedure to solve the classification problem and find the coefficients of the classifer.

Example 2. Consider again the same coin tossing experiment of Example 1.

Let us consider also another agent, Claire, who is disposed to accept/reject gambles on the basis of the coherent lower prevision $\underline{P}_2(\cdot)$, defined as $\underline{P}_2(g) \coloneqq \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g)\}$,[18] for every $g \in \mathcal{L}$ (this can be interpreted as she believes that the probability of the event Heads lies in the interval $[1/3, 2/3]$). That is, she is characterized by the following coherent set of gambles:

$$\mathcal{D}_2 \coloneqq \{g \in \mathcal{L} : \underline{P}_2(g) \geq 0\} \coloneqq \{g \in \mathcal{L} : \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g)\} \geq 0\}.$$

Then, for example, she is disposed to accept gambles in $\mathcal{A}_2 = \{[-1, 2]^\top, [2, -1]^\top, [4, -2]^\top, [1, -0.5]^\top\}$, and she is disposed to reject gambles in $\mathcal{R}_2 = \{[-3, 2]^\top, [1, -1]^\top\}$. Notice that, unlike Alice, Claire is disposed to reject the gamble $[1, -1]^\top$.

Consider then Bob who judges an agent rational if they respect axioms D1–D4. If Bob receives, as information on Claire, only the sets $\mathcal{A}_2$ and $\mathcal{R}_2$, he believes Claire is rational. Indeed, $(\mathcal{A}_2, \mathcal{R}_2)$ is piecewise linearly separable through, at least, the binary piecewise linear classifier:

$$(\forall g \in \mathcal{L})\ PLC_2(g) \coloneqq \begin{cases} 1 & \text{if } E_{\{1/3,2/3\}}(g) \geq 0, E_{\{2/3,1/3\}}(g) \geq 0, \\ -1 & \text{otherwise,} \end{cases}$$

which identifies, as the region classified as 1, the set $\overline{\mathcal{E}(\mathcal{A}_2)}$, i.e., the minimal set of assumptions on Claire's willingness to accept gambles. In this case, in particular, the latter corresponds to the set of desirable gambles of Claire, indeed $\overline{\mathcal{E}(\mathcal{A}_2)} = \{g \in \mathcal{L} : PLC_2(g) = 1\} = \{g \in \mathcal{L} : E_{\{1/3,2/3\}}(g) \geq 0, E_{\{2/3,1/3\}}(g) \geq 0\} = \{g \in \mathcal{L} : \underline{P}_2(g) \geq 0\}$. Therefore, Bob, can also completely reconstruct her probabilistic beliefs from the coefficients of the classifier $PLC_2(\cdot)$.

Figure 7 below shows gambles in $\mathcal{A}_2$, represented again as (blue) points in the plane (g_1, g_2), gambles in $\mathcal{R}_2$, represented as (red) triangles, and, in blue, the region classified as 1 by the classifier $PLC_2(\cdot)$.

[18] $\underline{P}_2(\cdot)$ is coherent since it can be equivalently expressed as the lower envelope of the weak*-closure of the convex hull of $\{E_{\{1/3,2/3\}}(\cdot), E_{\{2/3,1/3\}}(\cdot)\}$, as observed in Section 2. In particular, the two linear previsions $E_{\{1/3,2/3\}}(\cdot), E_{\{2/3,1/3\}}(\cdot)$ correspond to the extreme points of this latter set.

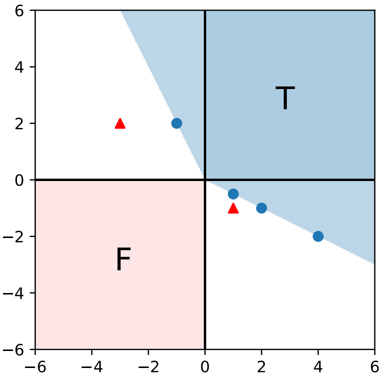

Figure 7: Gambles in $\mathcal{A}_2$, $\mathcal{R}_2$ and the region classified as 1 by $PLC_2(\cdot)$.

The classifier coefficients can be found numerically by solving a sequence of linear programming problems:[19]

$$LP_j = \begin{cases} c_j = \min_{\{\beta_j \in \mathbb{R}^2\}} (f^j)^\top \beta_j \\ s.t. \\ \quad -g^\top \beta_j \leq 0 \quad \forall g \in \mathcal{A}_2, \\ \quad \sum_{i=1}^{2} \beta_{j,i} = 1, \\ \beta_{j,1}, \beta_{j,2} \geq 0, \end{cases}$$

for each $f^j \in \mathcal{A}_2$. Assuming each problem is feasible and denoting the optimal solution of LP_j by $\hat{\beta}_j$, the classifier coefficients are given by all unique $\hat{\beta}_j$ such that $c_j = 0$. The binary piecewise linear classifier having coefficients $\{\hat{\beta}_j\}_j$ is a valid classifier (that is, it separates $\mathcal{A}_2$ and $\mathcal{R}_2$) provided that for each $f^j \in \mathcal{R}_2$ there exists $\hat{\beta}_k$ such that $(f^j)^\top \hat{\beta}_k < 0$.

4.1. Feature Mapping

In this section we will prove that the previous classification problem, which is nonlinear in general, can be reformulated as a linear one in a higher dimensional space. Let $\{\mathcal{B}_j\}_{j=1}^N$ denote a partition of $\mathcal{L} = \mathbb{R}^n$ [32],[20] where for every j:

$$\mathcal{B}_j := \{g \in \mathcal{L} : g^\top \omega_j \leq g^\top \omega_k \text{ for } k = 1, \dots, N, \quad j \neq k\}. \tag{12}$$

The vectors $\omega_j \in \mathbb{R}^n$ with $j \in \{1, ..., N\}$ are parameters defining the partition. We can introduce the feature mapping $\phi : \mathbb{R}^n \to \mathbb{R}^{nN}$, defined as $\phi(g) := [\phi_1(g), \dots, \phi_N(g)]^\top$ for every $g \in \mathcal{L}$, where $\phi_j : \mathbb{R}^n \to \mathbb{R}^n$ is defined in turn as:

$$\phi_j(g) := \mathbb{I}_{\mathcal{B}_j}(g) g, \tag{13}$$

[19] There are other ways to find numerically the classifier coefficients – the one we provided guarantees to find the exact solution in polynomial-time.

[20] We call it *partition* with a little abuse of notation. Indeed, we guarantee only that $\operatorname{int} \mathcal{B}_j \cap \operatorname{int} \mathcal{B}_k = \emptyset$, for every $j, k \in \{1, \dots, N\}$, $j \neq k$, where $\operatorname{int} \mathcal{B}_j$ represents the interior of $\mathcal{B}_j$ under the usual topology of $\mathbb{R}^n$. Instead, it is guaranteed that $\cup_{j=1}^N \mathcal{B}_j = \mathcal{L}$ (for every g, $\{g^\top \omega_j\}_{j \in N}$ is a finite set of real values so the minimum always exists). We thank a Referee for having suggested us that we can instead obtain a real partition if we combine our definition with a lexicographic order on $\{1, ...N\}$, so that for every j, $\mathcal{B}_j$ becomes the set of gambles for which the minimum is obtained with ω_j but not with ω_k for $k < j$. A similar reasoning can be applied also for the other definitions of partition given in Section 5.1 and Section 6.1. We decided however not to change the definitions of partitions originally introduced in order to not make the possible computational counterparts of our results more complex.

where

$$\mathbb{I}_{\mathcal{B}_j}(g) := \begin{cases} 1 & \text{if } g \in \mathcal{B}_j, \\ 0 & \text{otherwise,} \end{cases}$$

for every $g \in \mathcal{L}$ and $j \in \{1, \ldots, N\}$. [21] Further, we define the following classifier corresponding to a linear classifier in the feature space:

$$LC_\phi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N \phi_j(g)^\top \beta'_j \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{14}$$

for every $g \in \mathcal{L}$, with $\beta'_j \in \mathbb{R}^n$ for all $j = 1, \ldots, N$. In what follows, we consider both $\{\beta'_j\}_{j=1}^N$ and $\{\omega_j\}_{j=1}^N$ as parameters of $LC_\phi(\cdot)$.

Finally, we introduce the following definition to simplify notation.

Definition 8 (Φ-separability). A pair of sets of gambles (A, B) is Φ-*separable* if and only if there exists a classifier $LC_\phi(\cdot)$ of type (14) such that $LC_\phi(A) = 1$ and $LC_\phi(B) = -1$. We indicate the set of these classifiers with $\mathrm{LC}_\Phi(A, B)$.

We can now state the following result.

Proposition 5. *There is a one-to-one correspondence between binary piecewise linear classifiers with coefficients $\{\beta_j\}_{j=1}^N$ and classifiers of type* (14) *with parameters $\{\omega_j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \omega_j = \beta_j$, for every $j \in \{1, .., N\}$. In particular, the classification provided by the two classifiers is the same.*

Classifiers of type (14) satisfying the constraints specified by Proposition 5 are a way to rewrite binary piecewise linear classifiers as linear classifiers in a higher dimensional space. The reasoning is the following. First of all, partitions $\{\mathcal{B}_j\}_{j=1}^N$ allow us to divide $\mathcal{L}$ in subsets where binary piecewise linear classifiers boil down to linear ones. Indeed, consider a binary piecewise linear classifier $PLC(\cdot)$ characterized by parameters $\{\beta_j\}_{j=1}^N$. For every $g \in \mathcal{L}$, its classification through the classifier $PLC(\cdot)$ depends on the value in g of the nonlinear functional $plc : \mathcal{L} \to \mathbb{R}$, defined as $plc(h) := \min\{h^\top \beta_1, ..., h^\top \beta_N\}$, for every $h \in \mathcal{L}$. Consider now the partition $\{\mathcal{B}_j\}_{j=1}^N$ with parameters $\omega_j = \beta_j$ for every $j \in \{1, ..., N\}$. If $g \in \mathcal{B}_j$, $plc(g) = g^\top \beta_j$ so if we consider gambles on the right members of the partition, the nonlinear functional $plc(\cdot)$ collapses on a linear one. The feature mapping ϕ allows us to do that. If we then consider a linear classifier of type (14) with parameters $\{\omega_j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \omega_j = \beta_j$ for every j, we have:

$$\min(g^\top \beta_1, \ldots, g^\top \beta_N) \geq 0 \iff g^\top \beta_j \geq 0 \text{ whenever } g \in \mathcal{B}_j$$
$$\iff 0 \leq \sum_{j=1}^N (\mathbb{I}_{\mathcal{B}_j}(g) g)^\top \beta_j =: \sum_{j=1}^N \phi_j(g)^\top \beta_j.$$

The following corollary, which can be considered the central result of this section, thus holds.

Corollary 5. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is Φ−separable and there exists a classifier $LC_\phi(\cdot) \in LC_\Phi(\mathcal{A}, \mathcal{R})$ with parameters $\{\omega_j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \omega_j \gneq 0$ for every $j \in \{1, ..., N\}$.*

In particular, we have that, if $\mathcal{D}$ is a finitely generated coherent set, we can find $LC_\phi(\cdot)$ of the type specified in Corollary 5 such that $\mathcal{D} = \{g \in \mathcal{L} : LC_\phi(g) = 1\}$ and vice versa.

Analogous considerations will be made for the feature mappings introduced for the other definitions of *coherence* treated in Section 5.1 and in Section 6.1.

[21] Therefore, for every j and every $g \in \mathcal{L}$, multiplying the scalar $\mathbb{I}_{\mathcal{B}_j}(g)$ with the gamble g, we obtain: $\mathbb{I}_{\mathcal{B}_j}(g)g = g$ if $g \in \mathcal{B}_j$ and $\mathbb{I}_{\mathcal{B}_j}(g)g = 0$ otherwise, where 0 is the null vector in $\mathbb{R}^n$. Analogous notation is used for the other feature mappings described in the article.

Notice that this feature mapping allows us to transform an imprecise probabilistic model into a precise one in a higher dimensional space. Indeed, let us consider an agent having a finitely generated coherent set $\mathcal{D}$ of desirable gambles. There exists a binary piecewise linear classifier $PLC(\cdot)$ such that:

$$\mathcal{D} = \{g \in \mathcal{L} : PLC(g) = 1\} = \{g \in \mathcal{L} : g^\top \beta_j \geq 0,\ \forall j = 1, ..., N\},$$

with normalized coefficients $\{\beta_j\}_{j=1}^N$ such that $\beta_j \gneq 0$ for every j. Then the classifier $LC_\phi(\cdot)$ of type (14) associated to $PLC(\cdot)$ is such that:

$$\mathcal{D} = \{g \in \mathcal{L} : LC_\phi(g) = 1\} = \{g \in \mathcal{L} : \sum_{j=1}^N \phi_j(g)^\top \beta_j \geq 0\},$$

where $\phi(\cdot)$ is constructed from $\{\mathcal{B}_j\}_{j=1}^N$ having parameters $\omega_j = \beta_j$ for every j. This means that gambles g are equivalently evaluated through the expected utility operator: $(\forall g \in \mathcal{L})\ U(g) \coloneqq \sum_{j=1}^N \sum_{i=1}^n u_{j,i}(g)\tilde{\beta}_{j,i}$, with $u_{j,i}(g) = \phi_{j,i}(g) = id(\phi_{j,i}(g))$, where $id(\cdot)$ is the identity function, and $\tilde{\beta}_{j,i} = \frac{\beta_{j,i}}{\sum_{l,k}\beta_{l,k}}$ for every i, j. Thus, the imprecision of the probabilistic model is somewhat 'incorporated' into the nonlinear feature mapping, leading to a precise probabilistic model on the feature space.

Example 3. Consider again the framework of Example 2. Consider also the classifier $LC_\phi^2(\cdot)$ of type (14) with parameters $\{\omega_j, \beta'_j\}_{j=1}^2$ such that $\beta'_j = \omega_j = \beta_j$ for every j, where $\{\beta_j\}_{j=1}^2$ are the parameters of $PLC_2(\cdot)$ introduced in Example 2: $\beta_1 = [\frac{1}{3}, \frac{2}{3}]^\top$, $\beta_2 = [\frac{2}{3}, \frac{1}{3}]^\top$.

It corresponds to $PLC_2(\cdot)$. It therefore classifies $\mathcal{A}_2$ as 1 and $\mathcal{R}_2$ as -1. Moreover, $\{g \in \mathcal{L} : LC_\phi^2(g) = 1\} = \overline{\mathcal{E}(\mathcal{A}_2)}$.[22]

5. Convex coherence

As previously noticed, when one considers monetary gambles, the linearity assumption D3 is not very realistic. In this section we consider therefore a weaker form of rationality made up by the axioms D1, D2, D4 and a relaxed version of the axiom D3:

D3* : if $g, h \in \mathcal{D}$, then $\gamma g + (1-\gamma)h \in \mathcal{D}$ for each $\gamma \in [0, 1]$ [Convexity].

This axiom adds realism to the concept of rational agent considered. In particular, it gives us the possibility to model more general situations in which the agent judges rewards of gambles with a not necessarily linear utility scale or they have limited financial resources, see Example 4. It was already introduced in [22, 23, 24]. Here we connect it with our classification framework.

We define a set $\mathcal{D} \subseteq \mathcal{L}$ satisfying D1, D2, D3*, D4 a *convex coherent* set of gambles.

Again, a modeller that assumes this notion of rationality will think that an agent, who is willing to accept and reject respectively a finite set of gambles $\mathcal{A}$ and a finite set of gambles $\mathcal{R}$, is *rational* if and only if there exists a convex coherent set of gambles $\mathcal{D}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. If this is the case, the minimal such set is $\overline{\mathrm{ch}(\mathcal{A} \cup T)}$, the *closed convex hull* of $\mathcal{A} \cup T$ under the usual topology of $\mathbb{R}^n$ or, equivalently, under the supremum norm topology (see Lemma 2 in Appendix). Geometrically, the latter corresponds to a convex polyhedron (see Lemma 3 in Appendix):

$$\overline{\mathrm{ch}(\mathcal{A} \cup T)} = \mathrm{ch}^+(\mathcal{A} \cup \{0\}) \coloneqq \{g \in \mathcal{L} : g \geq f,\ f \in \mathrm{ch}(\mathcal{A} \cup \{0\})\}.$$

Moreover, similarly to before, $\overline{\mathrm{ch}(\mathcal{A} \cup T)}$ also corresponds to a deductive closure for the set $\mathcal{A}$, see [33]. Analogously to Section 2, we give the following definition.[23]

---

[22] Note that this is not the only feature mapping that we can use. Given the fact that $\overline{\mathcal{E}(\mathcal{A}_2)}$ is a closed convex cone in 2D indeed, there always exist a linear classifier that classifies $\overline{\mathcal{E}(\mathcal{A}_2)}$ as 1 and $\mathcal{L} \setminus \overline{\mathcal{E}(\mathcal{A}_2)}$ as -1 in the feature space determined by the feature mapping $\eta : g \to [g_2, \mathbb{I}_{g_1<0}(g)g_1, \mathbb{I}_{g_1\geq 0}(g)g_1]^\top$. In this case, it is sufficient to consider $\beta' = [1, 2, \frac{1}{2}]^\top$.

[23] A similar concept for preference relations over monetary lotteries can be found also in [22].

Definition 9 (Finitely generated convex coherent set). A convex coherent set of gambles $\mathcal{D} \subseteq \mathcal{L}$ is called *finitely generated (convex coherent set of gambles)* if and only if $\mathcal{D} = \overline{\mathrm{ch}(\mathcal{A} \cup T)}$, where $\mathcal{A}$ is a finite set of gambles.

Again analogously to the previous sections, we claim that proving whether an agent is rational in the sense of respecting D1,D2,D3*,D4 is equivalent to solving a binary classification task.

Definition 10 (Binary piecewise affine classifier). We use the term *binary piecewise affine classifier* to denote a classifier $PAC(\cdot)$ defined as follows:

$$PAC(g) := \begin{cases} c_1 & \text{if } g^\top \beta_j + \alpha_j \geq 0 \text{ for all } j \in \{1, ...N\}, \\ c_2 & \text{otherwise,} \end{cases} \tag{15}$$

for every $g \in \mathcal{L}$, with labels c_1, c_2, and $\beta_j \in \mathbb{R}^n$, $\alpha_j \in \mathbb{R}$ for all j, $N \geq 1$.

Again, without loss of generality, we assume $c_1 = 1$ and $c_2 = -1$.

Definition 11 (Piecewise affine separability). A pair of sets of gambles (A, B) is *piecewise affine separable* if and only if there exists a binary piecewise affine classifier $PAC(\cdot)$ such that $PAC(A) = 1$ and $PAC(B) = -1$. We indicate the set of these classifiers with $\mathrm{PAC}(A, B)$.

Now we can state the main result of this section.

Proposition 6. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a convex coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is piecewise affine separable.*

It is easy to show moreover that, again similarly to before, we can transform the problem of classifying the two non-finite sets of gambles T and F into constraints to be required on the coefficients of the classifier.

Proposition 7. *Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$. Every classifier $PAC(\cdot) \in PAC(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is a classifier $PAC(\cdot) \in PAC(\mathcal{A}, \mathcal{R})$ with coefficients $\{\beta_j, \alpha_j\}_{j=1}^N$ such that $\beta_j \gneq 0, \alpha_j \geq 0$, for every $j \in \{1, ..., N\}$ with at least an $\alpha_k = 0$ for some $k \in \{1, \ldots, N\}$, and vice versa.*

Corollary 6. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a convex coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is piecewise affine separable and there exists a classifier $PAC(\cdot) \in PAC(\mathcal{A}, \mathcal{R})$ with coefficients $\{\beta_j, \alpha_j\}_{j=1}^N$ such that $\beta_j \gneq 0, \alpha_j \geq 0$, for every $j \in \{1, ..., N\}$ with at least an $\alpha_k = 0$ for some $k \in \{1, \ldots, N\}$.*

Notice that if a binary piecewise affine classifier has coefficients $\{\beta_j, \alpha_j\}_{j=1}^N$ satisfying the constraints provided by Proposition 7 and Corollary 6 ($\forall j, \beta_j \gneq 0,\ \alpha_j \geq 0$ with at least an $\alpha_k = 0$), it is always possible to normalize each β_j and interpret its (normalized) components as the values of a probability mass function on Ω. Coefficients $\{\alpha_j\}_{j=}^N$ instead, can be interpreted as non-negative (with at least a zero one) 'penalty terms' due to, for example, limited financial resources of an agent, see Example 4. Clearly, the converse is also true. From now on therefore, we assume without loss of generality that if coefficients $\{\beta_j, \alpha_j\}_{j=1}^N$ of a binary piecewise affine classifier satisfy $\beta_j \gneq 0$ for every j, $\{\beta_j\}_{j=1}^N$ are normalized.

The proofs of Proposition 6 and Proposition 7, moreover, show that finitely generated convex coherent sets of gambles can always be represented as regions classified as 1 by binary piecewise affine classifiers $PAC(\cdot)$ satisfying constraints provided by Proposition 7 and Corollary 6. Vice versa, these latter regions are always convex coherent sets but in general not finitely generated ones, see Example 10 in Appendix.

The following diagram summarizes the implications highlighted above.

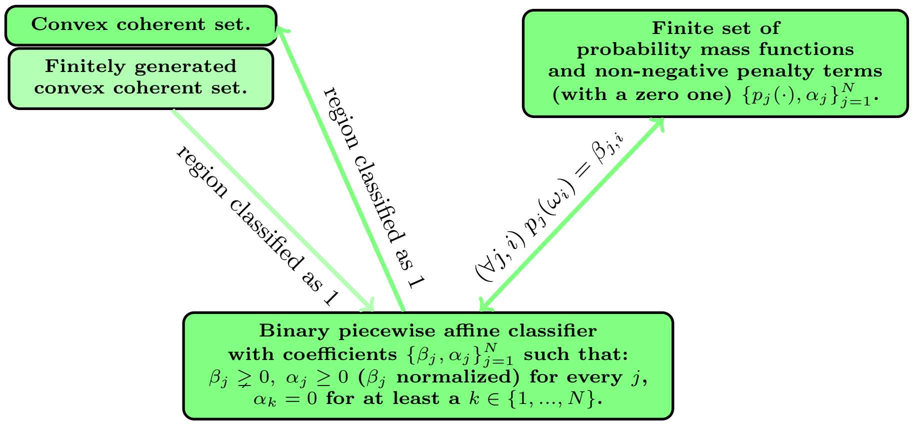

Figure 8: Diagram showing the implications among different models for representing convex coherent sets (finitely generated and more general ones).

Notice however that, analogously to what we discovered in Section 4, we can work only with finitely generated sets. Indeed, if given a pair of finite sets $(\mathcal{A}, \mathcal{R})$ there exists a convex coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, the smallest such set $\overline{\mathrm{ch}(\mathcal{A} \cup T)}$ is finitely generated. In this specific case therefore, we can enrich the previous diagram with the following implication.

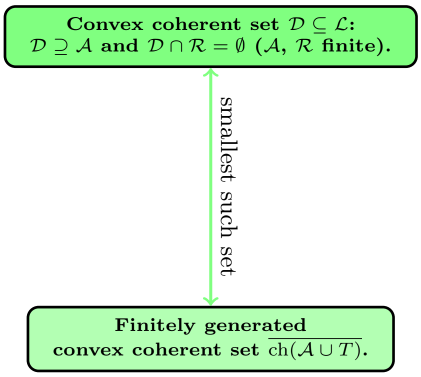

Figure 9: Diagram highlighting the smallest convex coherent set containing a finite set of gambles $\mathcal{A}$ (if it exists).

Given these considerations, Corollary 6 can be reformulated as follows.

Corollary 7. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a convex coherent set (not necessarily finitely generated) $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if equivalently:*

- *there exists a binary piecewise affine classifier $PAC(\cdot) \in PAC(\mathcal{A}, \mathcal{R})$ with coefficients $\{\beta_j, \alpha_j\}_{j=1}^N$ such that $\beta_j \gneq 0, \alpha_j \geq 0$, ($\beta_j$ normalized) for every $j \in \{1, ..., N\}$ with at least an $\alpha_k = 0$ for some $k \in \{1, ..., N\}$;*
- *there exists a finite set of probability mass functions on Ω, $\{p_j(\cdot)\}_{j=1}^N$, and a finite set of non-negative penalty terms (with at least a zero one) $\{\alpha_j\}_{j=1}^N$, such that $E_{p_j}(g) + \alpha_j \geq 0$ for every $g \in \mathcal{A}$, $j \in \{1, ..., N\}$ and, for every $g \in \mathcal{R}$, there exists $k \in \{1, ..., N\}$ such that $E_{p_k}(g) + \alpha_k < 0$;*
- *there exists a finitely generated convex coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$ (e.g., $\mathcal{D} = \overline{\mathrm{ch}(\mathcal{A} \cup T)}$).*

Since, in particular, if there exists a convex coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, the set $\overline{\mathrm{ch}(\mathcal{A} \cup T)}$ is the smallest convex coherent set containing $\mathcal{A}$, a binary piecewise affine classifier $PAC(\cdot)$ such that $\{g \in \mathcal{L} : PAC(g) = 1\} = \overline{\text{ch}(\mathcal{A} \cup T)}$ corresponds to make the minimal set of assumptions on the agent's willingness to accept gambles.

The following example illustrates the results obtained in this section considering a specific numerical case and reports a numerical optimization procedure to solve the classification problem.

Example 4. Consider again the coin tossing experiment of Example 1. In this framework suppose Diana judges gambles on the basis of Claire's coherent lower prevision $(\forall g \in \mathcal{L})\ \underline{P}_2(g) := \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g)\}$, but with the further constraint to not lose more than 1 thousand dollars. So, she is characterized by the following convex coherent set of gambles:

$$\begin{aligned}\mathcal{D}_3 &:= \{g \in \mathcal{L} : \underline{P}_2(g) \geq 0\} \cap \{g \in \mathcal{L} : \min(g) \geq -1\} = \\ &= \{g \in \mathcal{L} : \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g), g_1 + 1, g_2 + 1\} \geq 0\}.\end{aligned}$$

She is then disposed to accept, for example, gambles in $\mathcal{A}_3 = \{[-1, 2]^\top, [2, -1]^\top, [1, -0.5]^\top\}$ and reject gambles in $\mathcal{R}_3 = \{[-3, 2]^\top, [1, -1]^\top, [4, -2]^\top\}$. Notice that Diana is not disposed to accept, for example, the gamble $g' := [4, -2]^\top$, which instead is acceptable for Claire, because it can lead to a loss greater than 1 thousand dollars. Notice moreover that $g' = 2[2, -1]^\top$ where $[2, -1]^\top \in \mathcal{A}_3$. $\mathcal{D}_3$ therefore *does not* respect axiom D3, hence Diana does not judge rewards of gambles with a linear utility scale.

Consider now the modeller Bob who finds an agent rational if they respect D1,D2,D3*,D4. If Bob receives, as information about Diana, only the sets $\mathcal{A}_3$ and $\mathcal{R}_3$, Bob will think she is rational. Indeed, $(\mathcal{A}_3, \mathcal{R}_3)$ is piecewise affine separable through, at least, the binary piecewise affine classifier:

$$(\forall g \in \mathcal{L})\ PAC_3(g) := \begin{cases} 1 & \text{if } E_{\{1/3,2/3\}}(g) \geq 0, E_{\{2/3,1/3\}}(g) \geq 0, \\ & \quad E_{\{1,0\}}(g) + 1 \geq 0, E_{\{0,1\}}(g) + 1 \geq 0, \\ -1 & \text{otherwise,} \end{cases}$$

which identifies, as the region classified as 1, $\overline{\text{ch}(\mathcal{A}_3 \cup T)} = \text{ch}^+(\mathcal{A}_3 \cup \{0\})$, i.e., the minimal set of assumptions the modeller can made on the agent willingness to accept gambles. It is possible to notice that, in particular, it corresponds to Diana's set of desirable gambles $\mathcal{D}_3$.

The following Figure 10 shows gambles in $\mathcal{A}_3$ represented again as (blue) points in the plane (g_1, g_2), gambles in $\mathcal{R}_3$ represented as (red) triangles and, in blue, the region classified as 1 by the classifier $PAC_3(\cdot)$, i.e., Diana's set of desirable gambles.

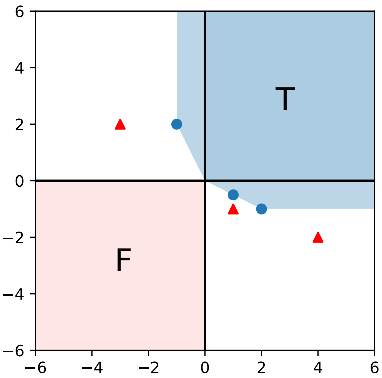

Figure 10: Gambles in $\mathcal{A}_3$, $\mathcal{R}_3$ and the region classified as 1 by $PAC_3(\cdot)$.

The classifier $PAC_3(\cdot)$ can be found numerically as follows:[24] for any gamble g, $PAC_3(g)$ is equal to 1

[24] There are other ways to find numerically the classifier coefficients – the one we provided guarantees to find the exact solution in polynomial-time.

if the following linear programming problem is feasible, otherwise it is equal to -1.

$$LP = \begin{cases} \min_{\{w_1,\dots,w_{|\mathcal{A}_3\cup\{0\}|},\lambda_1,\lambda_2\in\mathbb{R}\}} \lambda_1+\lambda_2 \\ s.t. \\ g = \sum_{f^j\in\mathcal{A}_3\cup\{0\}} w_j f^j + [\lambda_1,\lambda_2]^\top, \\ w_j \geq 0, \text{ for every } j\in\{1,\dots,|\mathcal{A}_3\cup\{0\}|\}, \\ \sum_{j=1}^{|\mathcal{A}_3\cup\{0\}|} w_j = 1, \\ \lambda_1,\lambda_2 \geq 0. \end{cases}$$

The classifier is contained in $\mathrm{PAC}(\mathcal{A}_3\cup T, \mathcal{R}_3\cup F)$ provided that $\mathcal{A}_3\cap F=\emptyset$, $PAC_3(g)=-1$ for each $g\in\mathcal{R}_3$ and $PAC_3([-\epsilon,-\epsilon])=-1$ for some small ϵ (which depends on $\mathcal{A}_3$). Note that, if $g\in \mathrm{ch}(\mathcal{A}_3\cup\{0\})$ then $\lambda_1,\lambda_2=0$, otherwise either λ_1 or λ_2 must be different from zero.

5.1. Feature Mapping

We can reformulate the previous classification problem as a linear classification task in a higher dimensional space, using a feature mapping similar to the one seen in Section 4.1. We can indeed define new partitions $\{\mathcal{B}'_j\}_{j=1}^N$ of $\mathcal{L}=\mathbb{R}^n$, where $\mathcal{B}'_j$ is defined as follows:

$$\begin{aligned} \mathcal{B}'_j := \{g\in\mathcal{L} : g^\top\omega'_{j,1:n}+\omega'_{j,n+1} \leq g^\top\omega'_{k,1:n}+\omega'_{k,n+1}, \\ \text{for } k=1,\dots,N,\quad j\neq k\} \end{aligned} \tag{16}$$

with $\omega'_j\in\mathbb{R}^{n+1}$ such that $\omega'_{j,1:n}$ is the vector containing the first n components of ω'_j, for every $j\in\{1,\dots,N\}$. We can introduce the feature mapping $\psi:\mathbb{R}^n\to\mathbb{R}^{(n+1)N}$, defined as $\psi(g):=[\psi_1(g),\dots,\psi_N(g)]^\top$ for every $g\in\mathcal{L}$, where $\psi_j:\mathbb{R}^n\to\mathbb{R}^{n+1}$ is defined in turn as:

$$\psi_j(g) := \begin{bmatrix} \mathbb{I}_{\mathcal{B}'_j}(g)g_1 \\ \cdots \\ \mathbb{I}_{\mathcal{B}'_j}(g)g_n \\ \mathbb{I}_{\mathcal{B}'_j}(g) \end{bmatrix} \tag{17}$$

for every $g\in\mathcal{L}$ and $j\in\{1,\dots,N\}$. Further, we define the following classifier corresponding to a linear classifier in the feature space:

$$LC_\psi(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N \psi_j(g)^\top\beta'_j \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{18}$$

for every $g\in\mathcal{L}$, with $\beta'_j\in\mathbb{R}^{n+1}$ for all $j=1,\dots,N$. We consider both $\{\beta'_j\}_{j=1}^N$ and $\{\omega'_j\}_{j=1}^N$ as parameters of $LC_\psi(\cdot)$. We can then introduce the following definition.

Definition 12 (Ψ-separability). A pair of sets of gambles (A,B) is Ψ-*separable* if and only if there exists a classifier $LC_\psi(\cdot)$ of type (18) such that $LC_\psi(A)=1$ and $LC_\psi(B)=-1$. We indicate the set of these classifiers with $\mathrm{LC}_\Psi(A,B)$.

We can now state the main result of this section.

Proposition 8. *There is a one-to-one correspondence between binary piecewise affine classifiers with coefficients* $\{\beta_j,\alpha_j\}_{j=1}^N$ *and classifiers of type* (18) *with parameters* $\{\omega'_j,\beta'_j\}_{j=1}^N$ *such that* $\beta'_j=\omega'_j=[\beta_j,\alpha_j]^\top$, *for every* $j\in\{1,..,N\}$. *In particular, the classification provided by the two classifiers is the same.*

The proof of Proposition 8 is analogous to the one of Proposition 5, indeed it is based on the same observation. The only difference is that now we work with classifiers having decision boundaries that are piecewise affine functions instead of piecewise linear ones. Hence, to reduce them to linear classifiers, we need to work in a feature space with more dimensions to take into account also the affine parts.

Hence, the following corollary also holds.

Corollary 8. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a convex coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is $\Psi-$separable and there exists a classifier $LC_\psi(\cdot) \in LC_\Psi(\mathcal{A}, \mathcal{R})$, with parameters $\{\omega'_j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \omega'_j \geq 0$ for every $j = 1, ..., N$ with $\omega'_{j,1:n} \gneq 0$ for every j, and at least a ω'_k with $k \in \{1, .., N\}$ so that $\omega'_{k,n+1} = 0$.*

Similarly to before, we have that, given a classifier $LC_\psi(\cdot)$ of the type specified by Corollary 8, the set $\mathcal{D} \coloneqq \{g \in \mathcal{L} : LC_\psi(g) = 1\}$ is always convex coherent. Vice versa, if $\mathcal{D}$ is a finitely generated convex coherent set, we can always find a classifier $LC_\psi(\cdot)$ of the type specified by Corollary 8 such that $\mathcal{D} = \{g \in \mathcal{L} : LC_\psi(g) = 1\}$.

Example 5. Consider again the framework of Example 4. Consider the classifier $LC^3_\psi(\cdot)$ of type (18) with parameters $\{\omega'_j, \beta'_j\}_{j=1}^4$ such that $\beta'_j = \omega'_j = \left[\begin{smallmatrix}\beta_j \\ \alpha_j\end{smallmatrix}\right]$ for every j, where $\{\beta_j, \alpha_j\}_{j=1}^4$ are the parameters of $PAC_3(\cdot)$ introduced in Example 4:

$$\beta_1 \coloneqq \begin{bmatrix}1/3 \\ 2/3\end{bmatrix}, \beta_2 \coloneqq \begin{bmatrix}2/3 \\ 1/3\end{bmatrix}, \beta_3 \coloneqq \begin{bmatrix}1 \\ 0\end{bmatrix}, \beta_4 \coloneqq \begin{bmatrix}0 \\ 1\end{bmatrix},$$

$$\alpha_1 = \alpha_2 = 0, \alpha_3 = \alpha_4 = 1.$$

It corresponds to $PAC_3(\cdot)$. It therefore classifies $\mathcal{A}_3$ as 1 and $\mathcal{R}_3$ as -1. Moreover, $\{g \in \mathcal{L} : LC^3_\psi(g) = 1\} = \overline{\text{ch}(\mathcal{A}_3 \cup T)}$.

6. Positive additive coherence

We now consider an even weaker relaxation (see Lemma 1 in Appendix) of the linearity axioms D3:

D3** if $g \in \mathcal{D}$, then $g + t \in \mathcal{D}$ for each $t \in T$ [positive additivity].

It can be considered the weakest axiom to add to D1,D2 and D4. Indeed, it forces an agent only to accept those gambles that are 'better' than the ones they are already disposed to accept, i.e., which can provide more gains and/or less losses. It has already been introduced in [24].

We call a set $\mathcal{D}$ satisfying D1, D2, D3**, D4 a *positive additive coherent* set of gambles.

Analogously to the previous sections, a modeller who receives as information about an agent only two finite sets of gambles $\mathcal{A}$ and $\mathcal{R}$ they are respectively disposed to accept and reject, will consider the agent rational in the sense of respecting D1,D2,D3**,D4 if and only if there exists a positive additive coherent set of gambles $\mathcal{D}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. If this is the case, the minimal such set is the *principal up-set* of $(\mathcal{A} \cup \{0\})$: $\uparrow (\mathcal{A} \cup \{0\}) \coloneqq \{g \in \mathcal{L} : (\exists f \in (\mathcal{A} \cup \{0\}))\ g \geq f\}$, see [29, Section 1.27], as Lemma 4 in Appendix proves. It is possible to notice that, geometrically, the latter corresponds to a union of closed, under the usual topology of $\mathbb{R}^n$, orthants centered in the elements of $\mathcal{A} \cup \{0\}$. Further, also in this case, it is a deductive closure for $\mathcal{A}$, see [29, Section 7.5].

Analogously to the other sections, we can give the following definition.

Definition 13 (Finitely generated positive additive coherent set). A positive additive coherent set of gambles $\mathcal{D} \subseteq \mathcal{L}$ is called *finitely generated (positive additive coherent set of gambles)* if and only if $\mathcal{D} = \uparrow (\mathcal{A} \cup \{0\})$, where $\mathcal{A}$ is a finite set of gambles.

As usual, in the rest of the section we show that a way the modeller has to check the rationality of an agent in the sense of respecting D1,D2,D3**,D4 is solving a binary classification task. Let us introduce the following definitions.

Definition 14 (PWP classifier). We use the term *binary piecewise positive affine (PWP) classifier* to denote a classifier $PWPC(\cdot)$ defined as follows:

$$PWPC(g) := \begin{cases} c_1 & \text{if } \exists f^j \in \mathcal{F} \text{ s.t. } g \geq f^j, \\ c_2 & \text{otherwise,} \end{cases} \tag{19}$$

for every $g \in \mathcal{L}$, with labels c_1, c_2 and with $\mathcal{F} := \{f^j\}_{j=1}^N$, a finite set of gambles (generators of the classifier).

Again, without loss of generality, we assume $c_1 = 1$ and $c_2 = -1$.

Definition 15 (PWP separability). A pair of sets of gambles (A, B) is *piecewise positive affine separable* (PWP) if and only if there exists a PWP classifier $PWPC(\cdot)$ such that $PWPC(A) = 1$ and $PWPC(B) = -1$. We indicate the set of these classifiers with $\text{PWPC}(A, B)$.

Note that, for every j, $\{g \geq f^j\}$ defines an orthant centered at f^j, whose border can be expressed as a piecewise affine function. It can easily be proved (by induction on the elements of $\mathcal{F}$) that the decision boundary of (19) is also a piecewise affine function. We can now state the main result of this section.

Proposition 9. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a positive additive coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is PWP separable.*

Once again, we can limit ourselves to classify only $\mathcal{A}$ and $\mathcal{R}$.

Proposition 10. *Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$. Every classifier $PWPC(\cdot) \in PWPC(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is a classifier $PWPC(\cdot) \in PWPC(\mathcal{A}, \mathcal{R})$ with generators $\{f^j\}_{j=1}^N$ such that, $f^j \not< 0$ for every $j \in \{1, ..., N\}$ and there exists $f^k \leq 0$ (but $f^k \not< 0$) for some $k \in \{1, ..., N\}$, and vice versa.*

Corollary 9. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a positive additive coherent set $\mathcal{D} \subseteq \mathcal{L}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is PWP separable and there exists a classifier $PWPC(\cdot) \in PWPC(\mathcal{A}, \mathcal{R})$ with generators $\{f^j\}_{j=1}^N$ such that, $f^j \not< 0$ for every $j \in \{1, ..., N\}$ and there exists $f^k \leq 0$ (but $f^k \not< 0$) for some $k \in \{1, ..., N\}$.*

The proofs of Proposition 9 and Proposition 10, moreover, show that there is a one-to-one correspondence between finitely generated positive additive coherent sets of gambles and regions classified as 1 by PWP classifiers $PWPC(\cdot)$ satisfying constraints provided by Proposition 10 and Corollary 9.

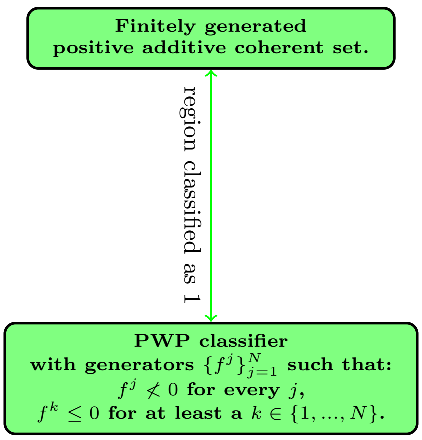

Figure 11: Diagram showing equivalent models for representing finitely generated positive additive coherent sets of gambles.

We can work only with finitely generated sets because, given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, if there exists a positive additive coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, the smallest such set $\uparrow(\mathcal{A} \cup \{0\})$ is finitely generated. Since, in particular, the latter is the smallest positive additive coherent set containing $\mathcal{A}$, a PWP classifier $PWPC(\cdot)$ such that $\{g \in \mathcal{L} : PWPC(g) = 1\} = \uparrow(\mathcal{A} \cup \{0\})$ corresponds to make the minimal set of assumptions on the agent's willingness to accept gambles.

The following example illustrates the results obtained in this section considering a specific numerical case and reports a numerical optimization procedure to solve the classification problem.

Example 6. Consider again the coin tossing experiment of Example 1. In this framework, consider also an agent Elena who is disposed to accept gambles in $\mathcal{A}_4 = \{[-1, 2]^\top, [2, -1]^\top\}$ and reject gambles in $\mathcal{R}_4 = \{[-3, 2]^\top, [1, -1]^\top, [4, -2]^\top, [1, -0.5]^\top\}$.

If the modeller Bob receives only $\mathcal{A}_4$ and $\mathcal{R}_4$ as information about Elena, he can check if at least she respects the minimal rationality axioms D1,D2, D3**,D4. In this case, Elena respects them because $(\mathcal{A}_4, \mathcal{R}_4)$ is PWP separable through the PWP classifier $PWPC_4(\cdot)$:

$$(\forall g \in \mathcal{L})\ PWPC_4(g) := \begin{cases} 1 & \text{if } \exists f \in \mathcal{A}_4 \cup \{0\} \text{ s.t. } g \geq f, \\ -1 & \text{otherwise,} \end{cases} \tag{20}$$

which identifies, as the region classified as 1, the set $\uparrow(\mathcal{A}_4 \cup \{0\})$, i.e., the minimal set of assumptions on the agent's willingness to accept gambles.

The following figure shows gambles in $\mathcal{A}_4$, represented again as (blue) points in the plane (g_1, g_2), gambles in $\mathcal{R}_4$, represented as (red) triangles, and, in blue, the region classified as 1 by the classifier $PWPC_4(\cdot)$. The numerical implementation of this classifier is straightforward: $PWPC_4(g) = 1$ if there exists a gamble f in $\mathcal{A}_4 \cup \{0\}$ such that $g \geq f$ and -1 otherwise, for every $g \in \mathcal{L}$.

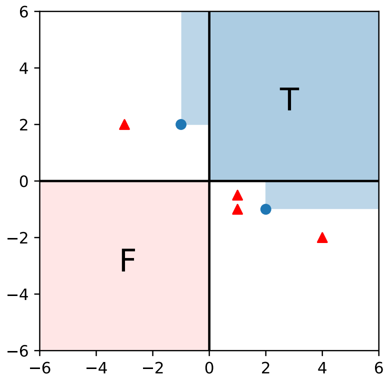

Figure 12: Gambles in $\mathcal{A}_4$, $\mathcal{R}_4$ and the region classified as 1 by $PWPC_4(\cdot)$.

6.1. Feature Mapping

In this section we prove that, also in this case, the previous classification problem can be reformulated as a linear one in a higher dimensional space. Let $\{\zeta_{j,i}\}_{\{j \in \{1,\dots,N\}, i \in \{1,\dots,n\}\}}$ denote another partition of $\mathcal{L} = \mathbb{R}^n$, where $\zeta_{j,i}$ is defined as follows:

$$\zeta_{j,i} := \{g \in \mathcal{L} : (g_i - \omega_i^j) = \max_{\{k \in \{1,\dots,N\}\}} (\min_{\{l \in \{1,\dots,n\}\}} (g_l - \omega_l^k))\} \tag{21}$$

with $\omega^j \in \mathbb{R}^n$, for every $i \in \{1, ..., n\}$, $j \in \{1, ..., N\}$. We can introduce the feature mapping $\rho : \mathbb{R}^n \to \mathbb{R}^{2nN}$, defined as $\rho(g) := [\rho_1(g), \ldots, \rho_N(g)]^\top$ for every $g \in \mathcal{L}$, where $\rho_j : \mathbb{R}^n \to \mathbb{R}^{2n}$ is defined in turn as:

$$\rho_j(g) := \begin{bmatrix} \mathbb{I}_{\zeta_{j,1}}(g) g_1 \\ \cdots \\ \mathbb{I}_{\zeta_{j,n}}(g) g_n \\ \mathbb{I}_{\zeta_{j,1}}(g) \\ \cdots \\ \mathbb{I}_{\zeta_{j,n}}(g) \end{bmatrix} \tag{22}$$

for every $g \in \mathcal{L}$ and $j \in \{1, \ldots, N\}$. Further, we define the following classifier corresponding to a linear classifier in the feature space:

$$LC_\rho(g) := \begin{cases} 1 & \text{if } \sum_{j=1}^N \rho_j(g)^\top \beta'_j \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{23}$$

for every $g \in \mathcal{L}$, with $\beta'_j \in \mathbb{R}^{2n}$ for all $j = 1, ..., N$. We consider both $\{\beta'_j\}_{j=1}^N$ and $\{\omega^j\}_{j=1}^N$ as parameters of $LC_\rho(\cdot)$. Similarly to before, we can introduce the following definition.

Definition 16 (P-separability). A pair of sets of gambles (A, B) is *P-separable* if and only if there exists a classifier $LC_\rho(\cdot)$ of type (23) such that $LC_\rho(A) = 1$ and $LC_\rho(B) = -1$. We indicate the set of these classifiers with $\mathrm{LC}_P(A, B)$.

We can now state the main result of this section.

Proposition 11. *There is a one-to-one correspondence between PWP classifiers with generators $\{f^j\}_{j=1}^N$ and classifiers of type (23) with parameters $\{\omega^j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = [1, \ldots, 1, -\omega^j_1, \ldots, -\omega^j_n]^\top$ and $\omega^j = f^j$, for every $j \in \{1, ..., N\}$.In particular, the classification provided by the two classifiers is the same.*

The proof is based on the following observation, analogous to the previous ones. Given a PWP classifier $PWPC(\cdot)$ with generators $\{f^j\}_{j=1}^N$:

$$PWPC(g) = 1 \iff \max_j(\min_i(g_i - f^j_i)) \geq 0$$
$$\iff \sum_{j=1}^N \left(\begin{bmatrix} \mathbb{I}_{\zeta_{j,1}}(g) g_1 \\ \cdots \\ \mathbb{I}_{\zeta_{j,n}}(g) g_n \\ \mathbb{I}_{\zeta_{j,1}}(g) \\ \cdots \\ \mathbb{I}_{\zeta_{j,n}}(g) \end{bmatrix} \right)^\top \begin{bmatrix} 1 \\ \cdots \\ 1 \\ -f^j_1 \\ \cdots \\ -f^j_n \end{bmatrix} \geq 0,$$

where $\zeta_{j,i} := \{g \in \mathcal{L} : (g_i - f^j_i) = \max_k(\min_l(g_l - f^k_l))\}$, for all j, i.

Corollary 10. *Given a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$, there exists a positive additive coherent set $\mathcal{D} \subseteq \mathcal{L}$ such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, if and only if $(\mathcal{A}, \mathcal{R})$ is $P-$separable and there exists a classifier $LC_\rho(\cdot) \in LC_P(\mathcal{A}, \mathcal{R})$ with parameters $\{\omega^j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = [1, \ldots, 1, -\omega^j_1, \ldots, -\omega^j_n]^\top$ with $\omega^j \not< 0$ for all $j \in \{1, ..., N\}$, and with at least a ω^k such that $\omega^k \leq 0$ (but $\omega^k \not< 0$) for some $k \in \{1, .., N\}$.*

In particular, in this case we have that, if $\mathcal{D}$ is a finitely generated positive additive coherent set, we can find $LC_\rho(\cdot)$ of the type specified in Corollary 10 such that $\mathcal{D} = \{g \in \mathcal{L} : LC_\rho(g) = 1\}$ and vice versa.

Example 7. Consider again the framework of Example 6. If we now introduce a classifier $LC^4_\rho(\cdot)$ of type (23), characterized by the parameters $\{\omega^j, \beta'_j\}_{j=1}^3$ such that $\beta'_j = [1, 1, -\omega^j_1, -\omega^j_2]^\top$ for every j with $\omega^1 = \left[\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right]$, $\omega^2 = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\omega^3 = \left[\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right]$ (i.e, $\{\omega^j\}_{j=1}^3 = (\mathcal{A}_4 \cup \{0\})$), it classifies $\mathcal{A}_4$ as 1 and $\mathcal{R}_4$ as -1. Moreover, $\{g \in \mathcal{L} : LC^4_\rho(g) = 1\} = \uparrow (\mathcal{A}_4 \cup \{0\})$.

7. Preferences and lower previsions

In this section we analyze properties satisfied by a lower prevision and a preference relation over gambles induced by sets of gambles satisfying each of the sets of axioms considered in the previous sections. It remains as an open point to evaluate the possibility to establish one-to-one correspondences between these new models and sets of gambles, analogous to the ones existing among coherent sets of gambles and coherent lower previsions and preferences.

Before starting with this analysis, we recall in the following diagram the existing relations among the different sets of rationality axioms considered. In particular, it is important to notice that {D1,D2,D3**,D4} corresponds to the weakest set of axioms considered. Therefore, maximal coherent sets, coherent and convex coherent ones are in particular positive additive coherent sets.

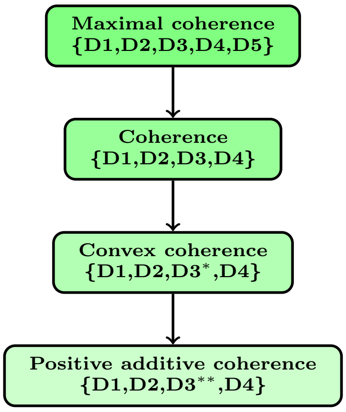

Figure 13: A graphical summary of the existing relations among different concepts of coherence analyzed in the paper.

Let us consider a positive additive coherent set of gambles $\mathcal{D} \subseteq \mathcal{L}$. Analogously to what is done in Section 2 for coherent sets of gambles, we can define from it a preference relation over gambles $\succsim_{\mathcal{D}}$ and a lower prevision $\underline{P}_{\mathcal{D}}(\cdot)$, as follows:

$$(\forall g, h \in \mathcal{L})\ g \succsim_{\mathcal{D}} h \iff g - h \in \mathcal{D}, \tag{24}$$

$$(\forall g \in \mathcal{L})\ \underline{P}_{\mathcal{D}}(g) \coloneqq \sup\{c \in \mathbb{R} : g - c \in \mathcal{D}\}. \tag{25}$$

Notice that, in particular, the definition of the lower prevision $\underline{P}_{\mathcal{D}}(\cdot)$ is consistent. Indeed, the set $\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ is non-empty and bounded from above for every gamble $g \in \mathcal{L}$. For every $g \in \mathcal{L}$, in fact, $g - (\max_{\{\omega\in\Omega\}} g(\omega) + \epsilon) \in F$ for every $\epsilon \in \mathbb{R}^*_+$. Therefore, since $\mathcal{D}$ satisfies D2, $\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ is bounded from above. Moreover, $\{c \in \mathbb{R} : c \leq \min_{\{\omega\in\Omega\}} g(\omega)\} \subseteq \{c \in \mathbb{R} : g - c \in \mathcal{D}\} \neq \emptyset$. If $c \leq \min_{\{\omega\in\Omega\}} g(\omega)$ in fact, by D1, we have $g - c \in T \subseteq \mathcal{D}$.

In Section 2 we concentrated on properties satisfied by $\succsim_{\mathcal{D}}$ and $\underline{P}_{\mathcal{D}}(\cdot)$ constructed starting from a coherent set $\mathcal{D}$. Here we analyze more general cases.

Proposition 12. *Consider a set of gambles $\mathcal{D} \subseteq \mathcal{L}$ satisfying D1,D2,D3**,D4. The preference relation $\succsim_{\mathcal{D}}$ defined from $\mathcal{D}$ through Eq.(24) satisfies the following properties:*

- $-1 \not\succsim_{\mathcal{D}} 0$ *[avoiding sure loss];*
- *if $g \geq h$ then $g \succsim_{\mathcal{D}} h$, for any $g, h \in \mathcal{L}$ [monotonicity];*
- *if $g + \delta \succsim_{\mathcal{D}} h$ for every $\delta \in \mathbb{R}^*_+$ then $g \succsim_{\mathcal{D}} h$, for any $g, h \in \mathcal{L}$ [continuity];*
- *$g \succsim_{\mathcal{D}} h$ if and only if $g - h \succsim_{\mathcal{D}} 0$, for any $g, h \in \mathcal{L}$ [cancellation].*

If $\mathcal{D}$ satisfies also D3*, then $\succsim_{\mathcal{D}}$ satisfies:

- $(g \succsim_{\mathcal{D}} h,\ f \succsim_{\mathcal{D}} l) \Rightarrow \gamma g + (1-\gamma) f \succsim_{\mathcal{D}} \gamma h + (1-\gamma) l$, *for any* $\gamma \in [0,1],\ g,h,f,l \in \mathcal{L}$ *[convexity].*[25]

If $\mathcal{D}$ *satisfies also D3, then* $\succsim_{\mathcal{D}}$ *satisfies:*

- *if* $g \succsim_{\mathcal{D}} h$ *and* $\lambda \in \mathbb{R}^*_+$ *then* $\lambda g \succsim_{\mathcal{D}} \lambda h$, *for any* $g,h \in \mathcal{L}$ *[positive homogeneity];*
- *if* $g \succsim_{\mathcal{D}} h$ *and* $h \succsim_{\mathcal{D}} l$ *then* $g \succsim_{\mathcal{D}} l$, *for any* $g,h,l \in \mathcal{L}$ *[transitivity].*

If $\mathcal{D}$ *satisfies D3 and also D5, then* $\succsim_{\mathcal{D}}$ *satisfies:*

- $g \succsim_{\mathcal{D}} h$ *or* $h \succsim_{\mathcal{D}} g$, *for any* $g,h \in \mathcal{L}$ *[completeness].*

The proof of Proposition 12 directly follows from the axioms satisfied each time by $\mathcal{D}$. It is easy to show that, moreover, a preference relation $\succsim_{\mathcal{D}}$ derived from a convex coherent set $\mathcal{D}$ satisfies axioms introduced in [22] for preference relations over monetary lotteries.

Proposition 13. *Consider a set of gambles* $\mathcal{D} \subseteq \mathcal{L}$ *satisfying D1,D2,D3**,D4. The lower prevision* $\underline{P}_{\mathcal{D}}(\cdot)$ *defined from* $\mathcal{D}$ *through Eq.*(25) *satisfies the following properties.*

- $\underline{P}_{\mathcal{D}}(g) \geq \min_{\omega \in \Omega} g(\omega)$, *for any* $g \in \mathcal{L}$*;*
- $\underline{P}_{\mathcal{D}}(g) \leq \max_{\omega \in \Omega} g(\omega)$, *for any* $g \in \mathcal{L}$*;*
- $\underline{P}_{\mathcal{D}}(g) = \max\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$, *for any* $g \in \mathcal{L}$*;*
- $\underline{P}_{\mathcal{D}}(g - \underline{P}_{\mathcal{D}}(g)) = 0$, *for any* $g \in \mathcal{L}$*;*
- *if* $g \geq h$ *then* $\underline{P}_{\mathcal{D}}(g) \geq \underline{P}_{\mathcal{D}}(h)$ *and if* $g > h$ *then* $\underline{P}_{\mathcal{D}}(g) > \underline{P}_{\mathcal{D}}(h)$, *for any* $g,h \in \mathcal{L}$ *[monotonicity];*
- $\underline{P}_{\mathcal{D}}(g + r) = \underline{P}_{\mathcal{D}}(g) + r$, *for any* $g \in \mathcal{L}$ *and any* $r \in \mathbb{R}$ *[translation invariance];*
- $\underline{P}_{\mathcal{D}}(0) = 0$*;*
- $\underline{P}_{\mathcal{D}}(\cdot)$ *is a uniformly continuous functional with respect to the supremum norm [uniform continuity].*

If $\mathcal{D}$ *satisfies also D3*, then* $\underline{P}_{\mathcal{D}}(\cdot)$ *satisfies:*

- $\underline{P}_{\mathcal{D}}(\gamma g + (1-\gamma) h) \geq \gamma \underline{P}_{\mathcal{D}}(g) + (1-\gamma)\underline{P}_{\mathcal{D}}(h)$, *for any* $\gamma \in [0,1],\ g,h \in \mathcal{L}$ *[concavity].*

If $\mathcal{D}$ *satisfies also D3, then* $\underline{P}_{\mathcal{D}}(\cdot)$ *satisfies:*

- $\underline{P}_{\mathcal{D}}(\lambda g) = \lambda \underline{P}_{\mathcal{D}}(g)$, *for any* $\lambda \in \mathbb{R}^*_+$ *and* $g \in \mathcal{L}$ *[positive homogeneity];*
- $\underline{P}_{\mathcal{D}}(g + h) \geq \underline{P}_{\mathcal{D}}(g) + \underline{P}_{\mathcal{D}}(h)$, *for any* $g,h \in \mathcal{L}$ *[superlinearity].*

If $\mathcal{D}$ *satisfies D3 and also D5, then* $\underline{P}_{\mathcal{D}}(\cdot)$ *satisfies:*

- $\underline{P}_{\mathcal{D}}(g) = -\underline{P}_{\mathcal{D}}(-g)$, *for any* $g \in \mathcal{L}$ *[self-conjugacy].*

Notice that, in particular, if $\mathcal{D}$ satisfies the standard axioms D1,D2,D3,D4, $\underline{P}_{\mathcal{D}}(\cdot)$ is a coherent lower prevision, if it satisfies also D5, $\underline{P}_{\mathcal{D}}(\cdot)$ is linear. If instead $\mathcal{D}$ satisfies only D1,D2,D3**,D4, $\underline{P}_{\mathcal{D}}(\cdot)$ is a *centered 2-convex lower prevision* and if $\mathcal{D}$ satisfies also D3*, $\underline{P}_{\mathcal{D}}(\cdot)$ is a *centered convex lower prevision.* Centered convex and centered 2-convex lower previsions are extensions of the concept of coherent lower prevision proposed in [24]. We recall here their characterizing properties on $\mathcal{L}$, see [24, Proposition 1, Remark 1, Proposition 2].

Proposition 14. *Consider a map* $\underline{P} : \mathcal{L} \to \mathbb{R}$.

1. $\underline{P}(\cdot)$ *is a* centered convex lower prevision *if and only if it satisfies:*
   - *A1.* $\underline{P}(0) = 0$*;*
   - *A2. if* $g \geq h$ *then* $\underline{P}(g) \geq \underline{P}(h)$, *for any* $g,h \in \mathcal{L}$*;*
   - *A3.* $\underline{P}(g + r) = \underline{P}(g) + r$, *for any* $g \in \mathcal{L}$ *and any* $r \in \mathbb{R}$ *[translation invariance];*
   - *A4.* $\underline{P}(g - \underline{P}(g)) = 0$, $\forall g \in \mathcal{L}$*;*
   - *A5.* $\underline{P}(\lambda g + (1-\lambda) h) \geq \lambda \underline{P}(g) + (1-\lambda)\underline{P}(h)$, *for any* $\lambda \in [0,1],\ g,h \in \mathcal{L}$.
2. $\underline{P}(\cdot)$ *is a* centered 2-convex lower prevision *if and only if A1, A2, A3 and A4 hold.*

The following example illustrates some of the results of this section.

Example 8. Consider the following sets of desirable gambles introduced in the previous examples:

- $\mathcal{D}_1 :=:= \{g \in \mathcal{L} : E_{\{2/3,1/3\}}(g) \geq 0\}$ of Example 1;

[25] This property is introduced also in [22].

- $\mathcal{D}_2 \coloneqq \{g \in \mathcal{L} : \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g)\} \geq 0\}$ of Example 2;
- $\mathcal{D}_3 \coloneqq \{g \in \mathcal{L} : \min\{E_{\{1/3,2/3\}}(g), E_{\{2/3,1/3\}}(g), g_1 + 1, g_2 + 1\} \geq 0\}$ of Example 4;
- $\mathcal{D}_4 \coloneqq \uparrow (\mathcal{A}_4 \cup \{0\}) = \{g \in \mathcal{L} : g \geq [-1, 2]^\top \text{ or } g \geq [0, 0]^\top \text{ or } g \geq [2, -1]^\top\}$ of Example 6.

From each one of these sets, we can induce a preference relation over gambles and a lower prevision.

Let us just focus on lower previsions.

Consider the gambles $0 = [0, 0]^\top$, $g = [2, -1]^\top$ and $h = [-1, 0]^\top$. The following table summarizes the values of the four lower previsions induced respectively from $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4$ for 0, g,h, $\frac{1}{2}g$, $2g$,$g + h$.

	$\underline{P}_{\mathcal{D}_1}(\cdot)$	$\underline{P}_{\mathcal{D}_2}(\cdot)$	$\underline{P}_{\mathcal{D}_3}(\cdot)$	$\underline{P}_{\mathcal{D}_4}(\cdot)$
0	0	0	0	0
g	1.34	0	0	0
h	-0.89	-0.89	-0.89	-1
$\frac{1}{2}g$	0.67	0	0	-0.5
$2g$	2.68	0	-1	-1
$g + h$	0.45	-0.45	-0.45	-1

From these values, it is possible to notice that $\underline{P}_{\mathcal{D}_4}(\frac{1}{2}g) \not\geq \frac{1}{2}\underline{P}_{\mathcal{D}_4}(g)$, hence it is not even a concave functional. $\underline{P}_{\mathcal{D}_3}(\cdot)$ does not respect positive homogeneity with $\lambda > 1$ since $\underline{P}_{\mathcal{D}_3}(2g) \neq 2\underline{P}_{\mathcal{D}_3}(g)$. Only $\underline{P}_{\mathcal{D}_1}(\cdot)$ and $\underline{P}_{\mathcal{D}_2}(\cdot)$ are standard coherent lower previsions. Moreover, $\underline{P}_{\mathcal{D}_1}(\cdot)$ is also a linear one.

8. General coherence

Consider now a new agent, Florence, again providing a finite set of acceptable gambles $\mathcal{A}$ and a finite set of rejectable gambles $\mathcal{R}$. In the previous sections we have analyzed conditions for different concepts of coherence and rationality. In each case we have also redefined them as linear classification problems in feature spaces. To conclude the paper, we would like to outline a general framework for modelling rationality starting directly from this common view.

Consider a general feature mapping $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^{MN}$ with $M \geq n$ and $N \geq 1$, such that $\hat{\phi}(\cdot) \coloneqq [\hat{\phi}_1(\cdot), \ldots, \hat{\phi}_N(\cdot)]^\top$, where $\hat{\phi}_j : \mathbb{R}^n \to \mathbb{R}^M$ for every $j \in \{1, ..., N\}$.

Starting from this feature mapping, we can define the following classifier, which corresponds to a linear classifier in the feature space:

$$LC_{\hat{\phi}}(g) \coloneqq \begin{cases} 1 & \text{if } \sum_{j=1}^{N} \hat{\phi}_j(g)^\top {\beta'}_j \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{26}$$

for every $g \in \mathcal{L}$ with $\beta'_j \in \mathbb{R}^M$ for every $j \in \{1, ..., N\}$.

We then consider sets separable through this general type of classifier.

Definition 17 ($\hat{\Phi}$-separability). A pair of sets of gambles (A, B) is $\hat{\Phi}$-*separable* if and only if there exist a classifier $LC_{\hat{\phi}}(\cdot)$ of type (26), such that $LC_{\hat{\phi}}(A) = 1$ and $LC_{\hat{\phi}}(B) = -1$. We indicate the set of these classifiers with $\mathrm{LC}_{\hat{\Phi}}(A, B)$.

From these elements we can derive a general definition of coherence, and hence of rationality for an agent.

Definition 18. Consider a general feature mapping $\hat{\phi} : \mathbb{R}^n \to \mathbb{R}^{MN}$ with $M \geq n$ and $N \geq 1$. A set of gambles $\mathcal{D} \subseteq \mathcal{L}$ is said $\hat{\phi}$-*coherent* if and only if there exists a classifier $LC_{\hat{\phi}}(\cdot)$ constructed from $\hat{\phi}(\cdot)$ through (26), such that $\mathcal{D} = \{g \in \mathcal{L} : LC_{\hat{\phi}}(g) = 1\}$.

An agent is considered *rational* in this sense, if their set of desirable gambles is $\hat{\phi}$-coherent. This definition of coherence is very general. By selecting the functions $\hat{\phi}_j(\cdot)$ in (26) as in (13), (17), (22), and imposing conditions on the coefficients of the classifier as in Proposition 5, 8 and 11, we obtain the piecewise linear, piecewise affine and, respectively, orthant-based classifers discussed in the previous sections, whose coherence-notion has been linked to the desirability axioms D1,D2,D3,D3*,D3** and D4. Choosing instead the identity function as a feature mapping, we obtain the linear classifiers discussed in Section 3.

In the more general setting treated in this section, we can provide sufficient conditions to guarantee that a $\hat{\phi}$-coherent set $\mathcal{D}$ satisfies at least the axioms D1,D2,D4, which we consider to be basic rationality axioms.

1. D4. A sufficient condition to guarantee that $\mathcal{D}$ satisfies D4 is that $\hat{\phi}(\cdot)$ is a continuous function. This is the case treated in the following Example 9. This condition is not necessary. Indeed it is not satisfied when $\hat{\phi}(\cdot)$ coincides with one of the feature mappings considered in the previous Sections 4.1, 5.1and 6.1. However, in all those cases, D4 is always satisfied.

2. D1 and D2. To guarantee that D1 and D2 are satisfied is sufficient to ask that the classifier $LC_{\hat{\phi}}(\cdot)$ classifies T as 1 and F as -1. Since these sets are infinite it can be useful, as before, to constrain the possible feature mappings in such a way that this task boils down to conditions to be required to the coefficients of the classifier, as it will be discussed in Example 9. This can be more easily achieved considering particular bases functions for the feature mappings, such as for example exponential bases, or odd polynomial functions.

Example 9. Consider, as before, a possibility space of size $n = 2$. Then, every gamble g has two components g_1 and g_2. Consider the feature mapping $\hat{\phi}(\cdot) \coloneqq \hat{\phi}_1(\cdot)$, where $\hat{\phi}_1 : \mathbb{R}^2 \to \mathbb{R}^3$ and is defined as:

$$\hat{\phi}_1(g) = \begin{bmatrix} -e^{-d_1 g_1} \\ -e^{-d_2 g_2} \\ 1 \end{bmatrix} \tag{27}$$

for every $g \in \mathcal{L}$, where, for every $i \in \{1, 2\}$, $d_i \geq 0$ denotes two constants (the parameters of the classifier). We consider then the classifier defined as:

$$LC_{\hat{\phi}}(g) \coloneqq \begin{cases} 1 & \text{if } -\beta'_{1,1}e^{-d_1 g_1} - \beta'_{1,2}e^{-d_2 g_2} + \beta'_{1,3} \geq 0, \\ -1 & \text{otherwise,} \end{cases} \tag{28}$$

for every $g \in \mathcal{L}$. We constrain $\beta'_{1,1}, \beta'_{1,2} \geq 0$, $\beta'_{1,1} + \beta'_{1,2} = 1$ and $\beta'_{1,3} = 1$ so that $LC_{\hat{\phi}}(T) = 1$ and $LC_{\hat{\phi}}(F) = -1$. Therefore, for $\mathcal{D} \coloneqq \{g \in \mathcal{L} : LC_{\hat{\phi}}(g) = 1\}$, D1,D2 hold and D4 follows by the continuity of the function $\hat{\phi}_1(\cdot)$.

In this context suppose that an agent, Diana, provides two finite sets of gambles $\mathcal{A}$ and $\mathcal{R}$ that are respectively acceptable and rejectable for her. Suppose moreover that they are $\hat{\phi}$-separable through a classifier $LC_{\hat{\phi}}(\cdot)$ as defined in (28). Among the classifiers that satisfy $LC_{\hat{\phi}}(\mathcal{A}) = 1$ and $LC_{\hat{\phi}}(\mathcal{R}) = -1$, we select the one which minimises the objective function:

$$\sum_{g^j \in \mathcal{A}} -\beta'_{1,1}e^{-d_1 g_1^j} - \beta'_{1,2}e^{-d_2 g_2^j} + 1. \tag{29}$$

Note that $-\beta'_{1,1}e^{-d_1 g_1} - \beta'_{1,2}e^{-d_2 g_2} + 1 \geq 0$ for all $g \in \mathcal{A}$ (coherence constraint). Therefore, by minimizing the objective function, we are minimizing the sum of the slack variables $s^j \coloneqq -\beta'_{1,1}e^{-d_1 g_1^j} - \beta'_{1,2}e^{-d_2 g_2^j} + 1 \geq 0$, forcing the classifier to interpolate as many points in $\mathcal{A}$ as possible (while satisfying $LC_{\hat{\phi}}(\mathcal{A}) = 1$). This means we are looking for the smallest extension in the feature space: the smallest set compatible with the assessments $\mathcal{A}$. The above nonlinear nonconvex optimisation problem can be solved numerically, by minimising (29) subject to

$$\begin{aligned} -\beta'_{1,1}e^{-d_1 g_1^j} - \beta'_{1,2}e^{-d_2 g_2^j} + 1 \geq 0, \quad \forall g^j \in \mathcal{A}, \\ -\beta'_{1,1}e^{-d_1 g_1^j} - \beta'_{1,2}e^{-d_2 g_2^j} + 1 < 0, \quad \forall g^j \in \mathcal{R}, \\ d_1, d_2, \beta'_{1,1} \geq 0, \quad \beta'_{1,2} = 1 - \beta'_{1,1}. \end{aligned} \tag{30}$$

Let us show a numerical example. Suppose Diana considers to be acceptable and rejectable respectively the following finite sets of gambles: $\mathcal{A}_3 = \{[-1,2]^\top, [2,-1]^\top, [1,-0.5]^\top\}$ and $\mathcal{R}_3 = \{[-3,2]^\top, [1,-1]^\top, [4,-2]^\top\}$. The optimal nonlinear classifier is obtained for $\beta'_{1,1} = 0.082$, $d_1 = 0.90, d_2 = 0.07$.

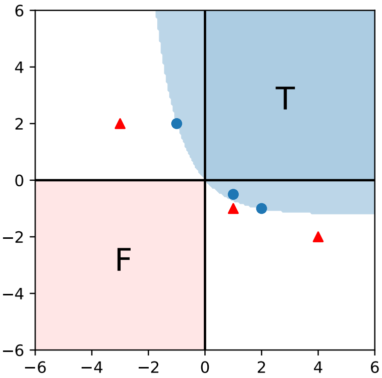

Figure 14: Nonlinear smooth classifier.

This classifier can be interpreted as the minimal set of assumptions the modeller Bob can make on Diana assuming she only accepts gambles g such that $-\beta'_{1,1}e^{-d_1 g_1} - \beta'_{1,2}e^{-d_2 g_2} + 1 \geq 0$, for some probability distribution $\{\beta'_{1,1}, \beta'_{1,2}\}$. The basis functions in (27) are commonly used to construct risk aversion models, see for instance [34]. Therefore, this example shows how our general framework allows us to express risk aversion models as binary classifiers. The set of desirable gambles identified by this classifier satisfies also D3*. However, it is not the smallest convex coherent set which is compatible with $\mathcal{A}_3, \mathcal{R}_3$ because, as discussed in Section 5, the smallest convex coherent set is identified through a piecewise-affine classifer.

9. Conclusions and future works

In this paper we have shown that standard desirability and other instances of nonlinear desirability (convex coherence and positive additive coherence) that relax the linearity axiom, can be formulated as a classification problem (at least in the finitely generated case). In this respect, we provided a unifying algorithmic framework (classification) to deal with different variants of desirability. Our contribution to rationality is to show that the type of rationality of an agent determines the type of classifier and vice-versa.

There are several research directions we aim to pursue in future works. First of all, Nau, in his work [34], characterizes set of desirable gambles of agents having inseparable subjective probabilities and utilities through a twice differentiable utility function $U(\cdot)$. In particular he defines sets of desirable gambles as $\mathcal{D} \coloneqq \{g \in \mathcal{L} : U(g) \geq 0\}$. In this context, he is also able to define *risk neutral probabilities* for the agent, i.e., agent's subjective probabilities adjusted for risk, and provide a measure of their *local risk aversion.* It could be interesting trying to apply Nau's results in our context. A starting point to do this could be to obtain an equivalent representation of convex coherent and positive additive coherent sets in terms of lower previsions, similarly to the one found for coherent sets in Section 2.

Another compelling line of research could be to investigate the *dual cones* of sets of desirable gambles defined in our work. In the standard case indeed, the probabilistic interpretation of a coherent set of gambles derives from the construction of the dual cone of $\mathcal{D}$, see [5]. Dual cones however, can be constructed also for subsets of $\mathbb{R}^n$ that are not necessarily convex cones. We plan therefore to construct them and analyzing possible connections with representation of $\mathcal{D}$ in the feature space, where essentially we recover the standard interpretation of $\mathcal{D}$ as the set classified as 1 by a linear classifier.

We are also interested in finding possible minimal coherence desiderata and characterize marginalisation and conditioning. In this regard, it could be interesting to see if the correspondence with *centered convex* and *centered 2-convex* lower previsions, as defined and analyzed in [24], remains valid also in the conditional case.

Besides, given the connection between standard desirability and Von Neumann - Morgenstern's axiomatization of rationality [7, 8, 9], we plan to investigate if the general framework provided here allows us to represent also non-expected utility theories, see for example [35, 36, 37, 38], as classification problems. This is not straightforward given the above axiomatization is defined over horse-lotteries instead of gambles.

Appendix

Proposition 15. *Consider a set of gambles $\mathcal{D} \subseteq \mathcal{L}$.*

If it is closed under the supremum norm topology, then it satisfies D4. Vice versa, if $\mathcal{D}$ satisfies also the following property:

$$f \geq g,\ g \in \mathcal{D} \Rightarrow f \in \mathcal{D} \tag{31}$$

then D4 implies closure under the supremum norm topology.

Proof. It is well-known that $\mathcal{L}$ is a Banach space under the supremum norm (see [5]).

Consider $\mathcal{D} \subseteq \mathcal{L}$ closed under the supremum norm topology. Then, given a sequence $(f_n)_{\{n \in \mathbb{N}\}}$ with $f_n \in \mathcal{D}$ for every n, convergent to $f \in \mathcal{L}$ (with respect to the supremum norm), we have $f \in \mathcal{D}$. Consider thus a gamble f such that $f + \delta \in \mathcal{D}$ for every $\delta \in \mathbb{R}^*_+$. We have therefore $f + \frac{1}{n} \in \mathcal{D}$ for every $n \in \mathbb{N}^*$. Its limit with respect to the supremum norm is f, hence $f \in \mathcal{D}$.

On the other hand, suppose $\mathcal{D}$ satisfies D4 and (31). Let us consider a sequence $(f_n)_{\{n \in \mathbb{N}\}}$ with $f_n \in \mathcal{D}$ for every n, convergent with respect to the supremum norm to a gamble $f \in \mathcal{L}$. We know that for every $\epsilon \in \mathbb{R}^*_+$ there exists $N \in \mathbb{N}$ such that $\sup |f_n - f| < \epsilon$ for all $n \geq N$. Hence $f_n < f + \epsilon$ for every $n \geq N$, from which it follows that, by (31), $f + \epsilon \in \mathcal{D}$. This procedure can be repeated for every $\epsilon \in \mathbb{R}^*_+$. Then by D4, we have $f \in \mathcal{D}$.

Proof (Proof of Proposition 3). Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$ for which there exists a coherent set of gambles $\mathcal{D}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. Then, the minimal coherent set $\mathcal{D}$ that satisfies these conditions is $\overline{\mathcal{E}(\mathcal{A})} := \overline{\mathsf{posi}(\mathcal{A} \cup T)}$. In fact, $\mathcal{E}(\mathcal{A})$ is the minimal set $\mathcal{D} \subseteq \mathcal{L}$ that satisfies D1–D3 such that $\mathcal{D} \supseteq \mathcal{A}$. Hence, thanks to Proposition 15 in Appendix, $\overline{\mathcal{E}(\mathcal{A})}$ is the minimal coherent set $\mathcal{D}'$ such that $\mathcal{D}' \supseteq \mathcal{A}$. Moreover, by hypothesis, we know also that $\overline{\mathcal{E}(\mathcal{A})} \cap \mathcal{R} = \emptyset$. This proof is also well-known in literature, see Section 2 or [5].

However, $\overline{\mathcal{E}(\mathcal{A})}$, by definition, is a *finitely generated cone* [39, Definition 2.3.2]. Indeed $\overline{\mathcal{E}(\mathcal{A})}$ can be rewritten as:

$$\overline{\mathcal{E}(\mathcal{A})} := \overline{\mathsf{posi}(\mathcal{A} \cup T)} =$$

$$C := \left\{ g \in \mathcal{L} : g = \sum_{j=1}^{r} \lambda_j f_j,\ f_j \in (\mathcal{A} \cup \{\mathbb{I}_{\omega_i}\}_{i=1}^n),\ \lambda_j \in \mathbb{R}_+,\ r \geq 1 \right\}.$$

The last equality derives from the following reasoning: $\mathcal{E}(\mathcal{A}) := \mathsf{posi}(\mathcal{A} \cup T)$; $\mathsf{posi}(\mathcal{A} \cup T) = C$; C is already closed under the usual topology of $\mathbb{R}^n$, hence under the supremum norm topology [5]. The latter follows indeed because, thanks to [39, Theorem 2.3.4], it is also a *polyhedral cone* in $\mathbb{R}^n$, hence an intersection of a finite number of closed halfspaces whose bounding hyperspaces pass through the origin:

$$C = \{g : g^\top \beta_j \geq 0,\ j = 1, ..., N\} \tag{32}$$

with $\beta_j \in \mathbb{R}^n$ for every $j \in \{1, ..., N\}$, $N \geq 1$, with at least a $\beta_k \neq 0$ for some $k \in \{1, ..., N\}$. This concludes this part of the proof since it tells us that there exists a binary piecewise linear classifier $PLC(\cdot)$ with parameters $\{\beta_j\}_{j=1}^N$, which classifies $\mathcal{A} \cup T \subseteq \overline{\mathcal{E}(\mathcal{A})} = C =: \{g \in \mathcal{L} : PLC(g) = 1\}$ as 1 and $(\mathcal{R} \cup F)$, which has empty intersection with C, as -1.

Vice versa, consider a piecewise linearly separable pair $(\mathcal{A} \cup T, \mathcal{R} \cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ are finite sets of gambles. Consider also a classifier $PLC(\cdot) \in \mathrm{PLC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$. Then:

$$\{g \in \mathcal{L} : PLC(g) = 1\} = \{g \in \mathcal{L} : g^\top \beta_j \geq 0,\ \text{for all } j = 1, .., N\} \tag{33}$$

for some $\beta_j \in \mathbb{R}^n$ such that $\beta_j \gneq 0$ for every $j \in \{1, ..., N\}$, $N \geq 1$ (constraints on β_j easily follow from the fact that $PLC(\cdot)$ classifies T as 1 and F as -1, see Proposition 4). We can then normalize every β_j, obtaining β'_j such that $\sum_i \beta'_{j,i} = 1$, for all j. Hence there exists a linear prevision $P_j(\cdot)$, such that $P_j(g) := g^\top \beta'_j$, for all g and for all $j = 1, ..., N$ [5, Section 2.8,Section 3.3]. Therefore we have:

$$\{g \in \mathcal{L} : PLC(g) = 1\} = \{g \in \mathcal{L} : g^\top \beta_j \geq 0, \text{ for all } j = 1, .., N\} =$$
$$\{g \in \mathcal{L} : P_j(g) \geq 0, \text{for all } j = 1, .., N\} = \{g \in \mathcal{L} : \underline{P}(g) \geq 0\},$$

where $\underline{P}(g) := \min_j\{P_j(g)\}$ for any $g \in \mathcal{L}$ is a coherent lower prevision, see Section 2 or [5, Theorem 3.6.2]. Hence, $\mathcal{D} := \{g \in \mathcal{L} : PLC(g) = 1\}$ is a (finitely generated) coherent set of gambles, see Eq. 2 or [5, Theorem 3.8.1].

In particular, we have also that $\mathcal{A} \subseteq \{g \in \mathcal{L} : PLC(g) = 1\} = \mathcal{D}$ and $\mathcal{R} \cap \{g \in \mathcal{L} : PLC(g) = 1\} = \mathcal{R} \cap \mathcal{D} = \emptyset$ by hypothesis.

Proof (Proof of Proposition 4). Consider a piecewise linearly separable pair $(\mathcal{A} \cup T, \mathcal{R} \cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ are finite sets of gambles. Consider also a classifier $PLC(\cdot) \in \mathrm{PLC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$ characterized by the coefficients $\{\beta_j\}_{j=1}^N$ such that $\beta_j \in \mathbb{R}^n$, for every $j \in \{1, ..., N\}$. If $\beta_j = 0$ for every $j \in \{1, .., N\}$, then $PLC(F) = 1$ that is a contradiction. Hence there is at least a coefficient $\beta_k \neq 0$ for some $k \in \{1, ..., N\}$. Without loss of generality, we suppose $\beta_j \neq 0$ for every j since the vanishing coefficients do not change the classification and therefore can be excluded from the ones characterizing the classifier.

Clearly $PLC(\cdot) \in \mathrm{PLC}(\mathcal{A}, \mathcal{R})$. Moreover, let us suppose that there exists a $k \in \{1, .., N\}$ such that $\beta_{k,i} < 0$ for some $i \in \{1, ..., n\}$. Consider $t \in T$ such that:

$$t(\omega_l) = \begin{cases} 1 & \text{if } l = i, \\ 0 & \text{otherwise.} \end{cases}$$

Then $t^\top \beta_k < 0$ and $PLC(t) = -1$. This is a contradiction, therefore $\beta_j \gneq 0$ for every $j \in \{1, ..., N\}$.

The converse immediately follows.

Proof (Proof of Corollary 3). The proof directly follows from Proposition 3 and Proposition 4.

Proof (Proof of Proposition 5). Consider a binary piecewise linear classifier $PLC(\cdot)$ with coefficients $\{\beta_j\}_{j=1}^N$ and a classifier $LC_\phi(\cdot)$ of type (14) with parameters $\{\omega_j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \omega_j = \beta_j$ for all $j = 1, ..., N$. They classify gambles in the same way. Indeed, consider $g \in \mathcal{L}$ and let us define $m := \min(g^\top \beta_1, \ldots, g^\top \beta_N)$. Then:

$$\sum_{j=1}^{N} \phi_j(g)^\top \beta_j := \sum_{j=1}^{N} (\mathbb{I}_{\mathcal{B}_j}(g) g)^\top \beta_j = \sum_{k=1}^{K} g^\top \beta_k = Km,$$

where $\{\mathcal{B}_j\}_{j=1}^N$ is the partition whose elements are specified by Eq.(12), with $\omega_j = \beta_j$ for every $j \in \{1, ..., N\}$ and $g^\top \beta_k = m$, for all $k = 1, ..., K$, with $1 \leq K \leq N$. Hence, g is classified in the same way by the classifiers $PLC(\cdot)$ and $LC_\phi(\cdot)$ because $m \geq 0 \iff \sum_{j=1}^N \phi_j(g)^\top \beta_j \geq 0$.

Proof (Proof of Corollary 5). The proof follows immediately from Corollary 3 and Proposition 5.

Lemma 1. *If a set $\mathcal{D} \subseteq \mathcal{L}$, satisfies D1, D3* and D4 then it satisfies* (31).

Proof. Consider $f \geq g$ with $g \in \mathcal{D}$. Then $f = g + t$ with $t \in T$.

Now, for every $\epsilon \in \mathbb{R}^*_+$ there exists a $\lambda_\epsilon \in (0, 1)$ such that $\lambda_\epsilon g \leq g + \epsilon$. Thus, given $\epsilon \in \mathbb{R}^*_+$ and $\lambda_\epsilon \in (0, 1)$, we have $f + \epsilon = \lambda_\epsilon g + (1 - \lambda_\epsilon) \frac{(g + \epsilon - \lambda_\epsilon g) + t}{1 - \lambda_\epsilon}$. Now, $g \in \mathcal{D}$ by hypothesis and $\frac{(g + \epsilon - \lambda_\epsilon g) + t}{1 - \lambda_\epsilon} \in T$, so $f + \epsilon \in \mathcal{D}$ by D1 and D3*. This can be repeated for every $\epsilon \in \mathbb{R}^*_+$, hence $f + \epsilon \in \mathcal{D}$ for all $\epsilon \in \mathbb{R}^*_+$ that implies that, by D4, $f \in \mathcal{D}$.

Lemma 2. *Given a pair of finite sets $(\mathcal{A},\mathcal{R})$ for which there exists a convex coherent set of gambles $\mathcal{D}$ such that $\mathcal{D}\supseteq\mathcal{A}$ and $\mathcal{D}\cap\mathcal{R}=\emptyset$, then the minimal such set is $\mathcal{D}=\overline{\mathrm{ch}(\mathcal{A}\cup T)}$.*

Proof. $\overline{\mathrm{ch}(\mathcal{A}\cup T)}$ satisfies D1 by definition and D3* [40, Theorem 6.2] and D4, thanks to Proposition 15 in Appendix.

Let us indicate with $\mathrm{D}(\mathcal{A},\mathcal{R})$, the class of convex coherent sets of gambles $\mathcal{D}$ such that $\mathcal{D}\supseteq\mathcal{A}$ and $\mathcal{D}\cap\mathcal{R}=\emptyset$. Thanks to Lemma 1 in Appendix and Proposition 15 in Appendix, every $\mathcal{D}\in\mathrm{D}(\mathcal{A},\mathcal{R})$, is a convex closed set (under the topology of $\mathbb{R}^n$ or equivalently under the supremum norm topology) that contains $(\mathcal{A}\cup T)$.

Given the fact that $\overline{\mathrm{ch}(\mathcal{A}\cup T)}\supseteq\mathcal{A}\cup T$ and, by definition, it is the intersection of all the closed (under the topology of $\mathbb{R}^n$ or equivalently under the supremum norm topology) and convex sets containing $(\mathcal{A}\cup T)$, see for example [41, Section 3.4.1], we have that $\overline{\mathrm{ch}(\mathcal{A}\cup T)}\subseteq\mathcal{D}$, for all $\mathcal{D}\in\mathrm{D}(\mathcal{A},\mathcal{R})$.

But, every $\mathcal{D}\in\mathrm{D}(\mathcal{A},\mathcal{R})$, satisfies $\mathcal{D}\cap(\mathcal{R}\cup F)=\emptyset$. Therefore, $\overline{\mathrm{ch}(\mathcal{A}\cup T)}\cap(\mathcal{R}\cup F)=\emptyset$, and hence it is also the smallest set $\mathcal{D}\in\mathrm{D}(\mathcal{A},\mathcal{R})$. This concludes the proof.

Lemma 3. *Consider a finite set $\mathcal{A}\subseteq\mathcal{L}$. Then:*

$$\overline{\mathrm{ch}(\mathcal{A}\cup T)}=\mathrm{ch}^+(\mathcal{A}\cup\{0\}):=\{g\in\mathcal{L}: g\geq f,\ f\in\mathrm{ch}(\mathcal{A}\cup\{0\})\}.$$

Proof. First of all, we can observe that:

$$\begin{aligned}\mathrm{ch}^+(\mathcal{A}\cup\{0\}) &:= \{g\in\mathcal{L}: g\geq f,\ f\in\mathrm{ch}(\mathcal{A}\cup\{0\})\} =\\ &= \sum_{i\in I}\alpha_i g^i+\sum_{j\in J}\gamma_j e^j =: \mathrm{ch}(\mathcal{A}\cup\{0\})+\mathrm{posi}(e^1,\dots,e^n)\end{aligned}$$

with I,J finite and $g^i\in\mathcal{A}\cup\{0\}$, $\alpha_i,\gamma_i\in\mathbb{R}_+$ for every i and $\sum_i\alpha_i=1$, where $\{e^j\}_{j=1}^n$ is the canonical basis in $\mathbb{R}^n$ and $\mathrm{posi}(e^1,\dots,e^n)$ is a fintely generated cone, hence a polyhedral cone. From [42, Corollary 7.1.b], it follows that $\mathrm{ch}^+(\mathcal{A}\cup\{0\})$ is a convex (closed) polyhedron. Hence $\overline{\mathrm{ch}^+(\mathcal{A}\cup\{0\})}=\mathrm{ch}^+(\mathcal{A}\cup\{0\})$. Now, we divide the proof in two parts.

- $\overline{\mathrm{ch}(\mathcal{A}\cup T)}\subseteq\mathrm{ch}^+(\mathcal{A}\cup\{0\})$. Notice that, thanks to the previous observation, it is sufficient to show that $\mathrm{ch}(\mathcal{A}\cup T)\subseteq\mathrm{ch}^+(\mathcal{A}\cup\{0\})$. So, let us consider $g\in\mathrm{ch}(\mathcal{A}\cup T)$. By definition, we have:

  $$g=\sum_{k=1}^{r}\alpha_k g^k$$

  with $\alpha_k\in\mathbb{R}_+, g^k\in(\mathcal{A}\cup T)$, for all $k=1,...,r,\ r\geq 1,\ \sum_{k=1}^r\alpha_k=1$. Let us indicate with $Ind_{A\setminus T}:=\{k\in\{1,...,r\}$ such that : $g^k\in\mathcal{A}\setminus T\}$ and $Ind_T:=\{k\in\{1,...,r\}$ such that : $g^k\in T\}$. We then have:

  $$g\geq\sum_{k\in Ind_{A\setminus T}}\alpha_k g^k+\sum_{k\in Ind_T}\alpha_k 0,$$

  hence $g\in\mathrm{ch}^+(\mathcal{A}\cup\{0\})$.
- $\mathrm{ch}^+(\mathcal{A}\cup\{0\})\subseteq\overline{\mathrm{ch}(\mathcal{A}\cup T)}$. By definition, $\overline{\mathrm{ch}(\mathcal{A}\cup T)}$ is a closed convex set that contains T. Therefore, from Proposition 15 in Appendix and Lemma 1 in Appendix, we have:

  $$\begin{gathered}\mathrm{ch}(\mathcal{A}\cup\{0\})\subseteq\overline{\mathrm{ch}(\mathcal{A}\cup T)}\Rightarrow\\ \mathrm{ch}^+(\mathcal{A}\cup\{0\})\subseteq\overline{\mathrm{ch}(\mathcal{A}\cup T)}.\end{gathered}$$

Proof (Proof of Proposition 6). Consider a pair of finite sets of gambles $(\mathcal{A},\mathcal{R})$ for which there exists a convex coherent set of gambles $\mathcal{D}$, such that $\mathcal{D}\supseteq\mathcal{A}$ and $\mathcal{D}\cap\mathcal{R}=\emptyset$. Then the minimal convex coherent set $\mathcal{D}$ which satisfies these conditions is $\overline{\mathrm{ch}(\mathcal{A}\cup T)}$, see Lemma 2 in Appendix. Thanks to Lemma 3 in Appendix, we know that it can be rewritten as:

$$\overline{\mathrm{ch}(\mathcal{A}\cup T)}=\mathrm{ch}^+(\mathcal{A}\cup\{0\}),\tag{34}$$

where $\mathrm{ch}^+(\mathcal{A}\cup\{0\})$ is a convex polyhedron. Any convex polyhedron can be written as an intersection of hyper-spaces, whose border is a piecewise affine function, see [42]. Therefore, there exists a binary piecewise affine classifier $PAC(\cdot)$, such that $\overline{\mathrm{ch}(\mathcal{A}\cup T)} = \overline{\mathrm{ch}^+(\mathcal{A}\cup\{0\})} = \{g\in\mathcal{L} : PAC(g)=1\}$. Note moreover that $\overline{\mathrm{ch}(\mathcal{A}\cup T)} = \{g\in\mathcal{L} : PAC(g)=1\} \supseteq (\mathcal{A}\cup T)$ and $\overline{\mathrm{ch}(\mathcal{A}\cup T)}\cap(\mathcal{R}\cup F) = \{g\in\mathcal{L} : PAC(g)=1\}\cap(\mathcal{R}\cup F)=\emptyset$ by hypothesis.

Vice versa, consider a piecewise affine separable pair $(\mathcal{A}\cup T, \mathcal{R}\cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ are finite sets of gambles. Let us consider a binary piecewise affine classifier $PAC(\cdot)\in \mathrm{PAC}(\mathcal{A}\cup T,\mathcal{R}\cup F)$. Now, the set:

$$\begin{aligned}\mathcal{D} :=& \{g\in\mathcal{L} : PAC(g)=1\} = \\ & \{g\in\mathcal{L} : g^\top\beta_j+\alpha_j\geq 0, \text{ for all } j=1,...,N\},\end{aligned}$$

for some $0\lneq\beta_j\in\mathbb{R}^n$ and $0\leq\alpha_j\in\mathbb{R}$ for all $j\in\{1,\dots N\}$ with at least a $k\in\{1,..,N\}$ such that $\alpha_k=0$ and $N\geq 1$ (constraints on $\beta_j,\ \alpha_j$ easily follow from the fact that $PAC(\cdot)$ classifies T as 1 and F as -1 - see the reasoning provided in the following Proposition 7 -), is a convex coherent set of gambles such that $\mathcal{D}\supseteq\mathcal{A}$ and $\mathcal{D}\cap\mathcal{R}=\emptyset$. Indeed:

- $T\subseteq\mathcal{D}$ and $\mathcal{D}\cap F=\emptyset$, by definition, hence it satisfies D1 and D2;
- $\mathcal{D}$ satisfies D3*. Consider $g,h\in\mathcal{D}$. Then $\gamma g+(1-\gamma)h\in\mathcal{D}$, for all $\gamma\in[0,1]$. Indeed,

$$\begin{aligned}&(\gamma g+(1-\gamma)h)^\top\beta_j+\alpha_j = \\ (\gamma g)^\top\beta_j+((1-\gamma)h)^\top\beta_j+\gamma\alpha_j+(1-\gamma)\alpha_j &= \\ \gamma(g^\top\beta_j+\alpha_j)+(1-\gamma)(h^\top\beta_j+\alpha_j)&\geq 0\end{aligned}$$

  for all $j\in\{1,...,N\}$ and $\gamma\in[0,1]$.
- $\mathcal{D}$ is closed under the usual topology of $\mathbb{R}^n$ because it is the intersection of a finite number of closed half-spaces hence, thanks to Proposition 15 in Appendix, it satisfies D4;
- clearly, by the fact that $PAC(\cdot)\in\mathrm{PAC}(\mathcal{A}\cup T,\mathcal{R}\cup F)$, it is also true that $\mathcal{A}\subseteq\mathcal{D}$ and $\mathcal{D}\cap\mathcal{R}=\emptyset$.

Proof (Proof of Proposition 7). Consider a piecewise affine separable pair $(\mathcal{A}\cup T,\mathcal{R}\cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ are finite sets of gambles. Consider also a classifier $PAC(\cdot)$ in $\mathrm{PAC}(\mathcal{A}\cup T,\mathcal{R}\cup F)$ characterized by the coefficients $\{\beta_j,\alpha_j\}_{j=1}^N$ such that $\beta_j\in\mathbb{R}^n,\ \alpha_j\in\mathbb{R}$ for every $j\in\{1,...,N\}$. If $\alpha_j=\beta_j=0$ for every $j\in\{1,...,N\}$, $PAC(F)=1$ that is a contradiction. Without loss of generality, we suppose that $\beta_j\neq 0$ for every j since if $\beta_k=0$ for some $k\in\{1,...,N\}$, $g^\top\beta_k+\alpha_k=\alpha_k$ for every $g\in\mathcal{L}$, hence excluding the couple $\{\beta_k,\alpha_k\}$ from the coefficients characterizing the classifier do not change the classification.

Clearly $PAC(\cdot)\in\mathrm{PAC}(\mathcal{A},\mathcal{R})$. Moreover, first of all, let us suppose that there exists a k such that $\alpha_k<0$. Then $0^\top\beta_k+\alpha_k<0$, hence $PAC(0)=-1$ and this is a contradiction. Therefore, $\alpha_j\geq 0$ for every $j\in\{1,...,N\}$.

Now, suppose there is a β_k such that $\beta_{k,i}<0$ for some $i\in\{1,...,n\}$. Then, consider $t\in T$ and $\epsilon\in\mathbb{R}^*_+$ such that:

$$t(\omega_l)=\begin{cases}\frac{\alpha_k+\epsilon}{|\beta_{k,i}|} & \text{if } l=i,\\ 0 & \text{otherwise.}\end{cases}$$

Then, $t^\top\beta_k+\alpha_k=-\alpha_k-\epsilon+\alpha_k<0$, so $PAC(t)=-1$ and this is a contradiction. Therefore $\beta_j\gneq 0$ for every $j\in\{1,...,N\}$.

Finally, let us suppose $\alpha_j>0$ for every $j\in\{1,...,N\}$. Then, consider the gamble $f\in F$ such that $f(\omega_i):=-\frac{\min_k\alpha_k}{n*max_{i,j}\beta_{j,i}}$ for every $i\in\{1,..,n\}$. Then, $f^\top\beta_j+\alpha_j\geq-\min_k\alpha_k+\alpha_j\geq 0$, for every j. Hence, $PAC(f)=1$ and this is a contradiction. So, there exists at least a $k\in\{1,...,N\}$ such that $\alpha_k=0$.

The converse immediately follows.

Proof (Proof of Corollary 6). The proof directly follows from Proposition 6 and Proposition 7.

Example 10. Consider a binary piecewise affine classifier $PAC(\cdot)$ with coefficients $\{\beta_j,\alpha_j\}_{j=1}^N$ satisfying the constraints of Proposition 7 and Corollary 6 ($\forall j,\ \beta_j\gneq 0,\ \alpha_j\geq 0$ with at least an $\alpha_k=0$). From the proof of Proposition 6 and Prposition 7, it follows that $\{g \in \mathcal{L} : PAC(g) = 1\}$ is always a convex coherent set of gambles. In general, however, $\{g \in \mathcal{L} : PAC(g) = 1\}$ is not a finitely generated convex coherent set. Let us consider for example the case in which $\{g \in \mathcal{L} : PAC(g) = 1\} = \{g \in \mathcal{L} : PLC(g) = 1\} \neq T$, where $PLC(\cdot)$ is a binary piecewise linear classifier. Let us suppose $\{g \in \mathcal{L} : PLC(g) = 1\} = \overline{\mathrm{ch}(\mathcal{A}' \cup T)} = \mathrm{ch}^{+}(\mathcal{A}' \cup \{0\})$, where $\mathcal{A}'$ is a finite set of gambles. Since $\{g \in \mathcal{L} : PLC(g) = 1\} \neq T$, we can suppose without loss of generality that there exists at least a gamble $g \in \mathcal{A}'$ such that $g \notin T$, i.e., $g_i < 0$ for some $i \in \{1, ..., n\}$. Let us consider then $g' \in \mathcal{A}'$ having $g'_i = \min_{g \in \mathcal{A}'} g_i$. There exists also $\lambda > 1$ sufficiently big such that $\lambda g' \notin \mathrm{ch}(\mathcal{A}' \cup \{0\})$, but $\lambda g' \in \{g \in \mathcal{L} : PLC(g) = 1\}$ since the latter set is coherent (the coefficients of $PLC(\cdot)$ indeed satisfy also the constraints of Proposition 4). However $\lambda g' \notin \mathrm{ch}^{+}(\mathcal{A}' \cup \{0\})$ either, since $\lambda g'_i < g'_i \leq f_i$, for every $f \in \mathrm{ch}(\mathcal{A}' \cup \{0\})$.

Proof (Proof of Proposition 8). Consider a binary piecewise affine classifier $PAC(\cdot)$ with coefficients $\{\beta_j, \alpha_j\}_{j=1}^{N}$ and a classifier $LC_{\psi}(\cdot)$ of type (18) with parameters $\{\omega'_j, \beta'_j\}_{j=1}^{N}$ such that $\beta'_j = \omega'_j = \begin{bmatrix}\beta_j \\ \alpha_j\end{bmatrix}$ for all $j = 1, ..., N$. They classify gambles in the same way. Indeed, consider $g \in \mathcal{L}$ and let us define $m \coloneqq \min(g^{\top}\beta_1 + \alpha_1, \ldots, g^{\top}\beta_N + \alpha_N)$. Then:

$$\sum_{j=1}^{N} \psi_j(g)^{\top} \begin{bmatrix}\beta_j \\ \alpha_j\end{bmatrix} = \sum_{j=1}^{N} \mathbb{I}_{\mathcal{B}'_j}(g)(g^{\top}\beta_j + \alpha_j) = Km$$

where $\{\mathcal{B}'_j\}_{j=1}^{N}$ is the partition whose elements are specified by Eq.(16) with $\omega'_j = \begin{bmatrix}\beta_j \\ \alpha_j\end{bmatrix}$ for every $j \in \{1, ..., N\}$ and where $1 \leq K \leq N$. Hence, g is classified in the same way by the classifiers $PAC(\cdot)$ and $LC_{\psi}(\cdot)$ because $m \geq 0 \iff \sum_{j=1}^{N} \psi_j(g)^{\top} \begin{bmatrix}\beta_j \\ \alpha_j\end{bmatrix} \geq 0$.

Proof (Proof of Corollary 8). The proof follows immediately from Corollary 6 and Proposition 8.

Lemma 4. *Given a pair of finite sets $(\mathcal{A}, \mathcal{R})$ for which there exists a positive additive coherent set of gambles $\mathcal{D}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$, then the smallest such set is:*

$$\mathcal{D} = \uparrow (\mathcal{A} \cup \{0\}) \coloneqq \{g \in \mathcal{L} : (\exists f \in \mathcal{A} \cup \{0\})\ g \geq f\}.$$

Proof. $\uparrow (\mathcal{A} \cup \{0\})$ satisfies D1, D3** and $\mathcal{A} \subseteq \uparrow (\mathcal{A} \cup \{0\})$ by construction. Moreover, it satisfies also D4 by Proposition 15 in Appendix, because it is closed under the usual topology of $\mathbb{R}^n$ (it is a finite union of closed sets).

Let us indicate with $\mathrm{P}(\mathcal{A}, \mathcal{R})$, the class of positive additive coherent sets of gambles $\mathcal{D}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. Clearly, each $\mathcal{D} \in \mathrm{P}(\mathcal{A}, \mathcal{R})$ satisfies $\mathcal{D} \supseteq \uparrow (\mathcal{A} \cup \{0\})$ because, thanks to Proposition 15 in Appendix, they are also closed under the supremum norm topology, hence under the usual topology of $\mathbb{R}^n$. But, every $\mathcal{D} \in \mathrm{P}(\mathcal{A}, \mathcal{R})$, satisfies also $\mathcal{D} \cap (\mathcal{R} \cup F) = \emptyset$. Therefore, $\uparrow (\mathcal{A} \cup \{0\}) \cap (\mathcal{R} \cup F) = \emptyset$. So, it is also the smallest positive additive coherent set of gambles $\mathcal{D} \in \mathrm{P}(\mathcal{A}, \mathcal{R})$.

Proof (Proof of Proposition 9). Consider a pair of finite sets of gambles $(\mathcal{A}, \mathcal{R})$ for which there exists a positive additive coherent set of gambles $\mathcal{D}$, such that $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$. Then the minimal such set is $\uparrow (\mathcal{A} \cup \{0\})$. However, it can be rewritten as:

$$\uparrow (\mathcal{A} \cup \{0\}) = \{g \in \mathcal{L} : PWPC(g) = 1\}$$

where $PWPC(\cdot)$ is a PWP classifier, defined as:

$$PWPC(g) \coloneqq \begin{cases} 1 & \text{if } \exists f^j \in (\mathcal{A} \cup \{0\}) \text{ s.t. } g \geq f^j, \\ -1 & \text{otherwise.} \end{cases}$$

Therefore, given that $\mathcal{A} \cup T \subseteq \uparrow(\mathcal{A} \cup \{0\}) = \{g \in \mathcal{L} : PWPC(g) = 1\}$ and $\uparrow(\mathcal{A} \cup \{0\}) \cap (\mathcal{R} \cup F) = \{g \in \mathcal{L} : PWPC(g) = 1\} \cap (\mathcal{R} \cup F) = \emptyset$, we have that $(\mathcal{A} \cup T, \mathcal{R} \cup F)$ is PWP separable. Vice versa, consider a PWP separable pair $(\mathcal{A} \cup T, \mathcal{R} \cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ are finite sets of gambles. Consider also a classifier $PWPC(\cdot) \in \text{PWPC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$. Then:

$$\mathcal{D} \coloneqq \{g \in \mathcal{L} : PWPC(g) = 1\}$$

is, by construction, a positive additive coherent set of gambles. Indeed, it clearly satisfies D1, D2, D3**. Further, it is closed because it is a finite union of closed sets (under the usual topology of $\mathbb{R}^n$) hence, by Proposition 15 in Appendix, it satisfies D4. It satisfies also $\mathcal{D} \supseteq \mathcal{A}$ and $\mathcal{D} \cap \mathcal{R} = \emptyset$ by hypothesis. Moreover, since the set of generators $\mathcal{F}$ of a PWP classifier is always finite, we have: $\{g \in \mathcal{L} : PWPC(g) = 1\} = \uparrow(\mathcal{F})$. Therefore, $\mathcal{D}$ is a finitely generated positive additive coherent set ($T \subseteq \{g \in \mathcal{L} : PWPC(g) = 1\}$ so $\{g \in \mathcal{L} : PWPC(g) = 1\} = \uparrow(\mathcal{F} \cup \{0\})$).

Proof (Proof of Proposition 10). Consider a PWP separable pair $(\mathcal{A} \cup T, \mathcal{R} \cup F)$, where $\mathcal{A}$ and $\mathcal{R}$ finite sets of gambles. Consider also a classifier $PWPC(\cdot) \in \text{PWPC}(\mathcal{A} \cup T, \mathcal{R} \cup F)$, characterized by the generators $\{f^j\}_{j=1}^N$ such that $f^j \in \mathbb{R}^n$ for every $j \in \{1, ...N\}$. Clearly, $PWPC(\cdot) \in \text{PWPC}(\mathcal{A}, \mathcal{R})$.

Moreover, let us suppose that there is a f^k such that $f^k < 0$. Consider $0 < \epsilon < \min_{\{i \in \{1,\dots,n\}\}}\{|f^k(\omega_i)|\}$. Then $f \coloneqq -\epsilon \geq f^k$, so $PWPC(f) = 1$. But $f \in F$, hence this is a contradiction. Therefore, for every $j \in \{1, .., N\}$, there exists $i \in \{1, ..., n\}$ such that $f^j(\omega_i) \geq 0$.

Let us suppose now that for every $j \in \{1, ...N\}$ there exists $l \in \{1, ..., n\}$ such that $f^j(\omega_l) > 0$. Consider $t = 0$. Then $t \not\geq f^j$ for every $j \in \{1, ..., N\}$. Hence $PWPC(0) = -1$ and this is a contradiction. Therefore there exists at least a f^k with $k \in \{1, .., N\}$ such that $f^k \leq 0$ and $f^k \not< 0$.

The converse immediately follows.

Proof (Proof of Corollary 9). The proof directly follows from Proposition 9 and Proposition 10.

Proof (Proof of Proposition 11). Consider a PWP classifier $PWPC(\cdot)$ with generators $\mathcal{F} = \{f^j\}_{j=1}^N$ and a classifier $LC_\rho(\cdot)$ of type (23) with parameters $\{\omega^j, \beta'_j\}_{j=1}^N$ such that $\beta'_j = \begin{bmatrix} 1 \\ \dots \\ 1 \\ -\omega_1^j \\ \dots \\ -\omega_n^j \end{bmatrix}$ and $\omega^j = f^j$ for all $j = 1, \dots, N$. They classify gambles in the same way. Indeed, consider $g \in \mathcal{L}$ and let us define $m \coloneqq \max_{\{k \in \{1,..,N\}\}}(\min_{\{l \in \{1,\dots,n\}\}}(g_l - f_l^k))$. Then:

$$\sum_{j=1}^{N} \rho_j(g)^\top \beta'_j = \sum_{j=1}^{N} \sum_{i=1}^{n} \mathbb{I}_{\zeta_{j,i}}(g)(g_i - f_i^j) = KLm$$

where $\{\zeta_{j,i}\}_{\{j \in \{1,\dots,N\}, i \in \{1,\dots,n\}\}}$ is the partition whose elements are specified by Eq.(21) with $\omega_i^j = f_i^j$ for every i, j and where $1 \leq L \leq n$, $1 \leq K \leq N$. Hence, g is classified in the same way by the classifiers $PWPC(\cdot)$ and $LC_\rho(\cdot)$ because $(\exists f^j \in \mathcal{F} : g \geq f^j) \iff m \geq 0 \iff \sum_{j=1}^N \rho_j(g)^\top \beta'_j \geq 0$.

Proof (Proof of Corollary 10). The proof follows immediately from Corollary 9 and Proposition 11.

Proof (Proof of Proposition 13). Consider a set $\mathcal{D} \subseteq \mathcal{L}$ satisfying D1, D2, D3**, D4.

For any $g \in \mathcal{L}$, $\min_{\omega \in \Omega} g(\omega) \in \{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ by D1, hence $\underline{P}_\mathcal{D}(g) \coloneqq \sup\{c \in \mathbb{R} : g - c \in \mathcal{D}\} \geq \min_{\omega \in \Omega} g(\omega)$. Moreover, $\underline{P}_\mathcal{D}(g) \leq \max_{\omega \in \Omega} g(\omega)$, thanks to D2.

Thus, as also observed before, for every $g \in \mathcal{L}$, $\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ is not empty and bounded from above. D3**, D4 and Proposition 15 in Appendix guarantee that $\mathcal{D}$ is closed under the supremum norm topology, then $\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ is also a closed subset of $\mathbb{R}$. Therefore, we have $\underline{P}_\mathcal{D}(g) = \max\{c \in \mathbb{R} : g - c \in \mathcal{D}\}$ for every $g \in \mathcal{L}$.

Clearly now $\underline{P}_{\mathcal{D}}(g - \underline{P}_{\mathcal{D}}(g)) = 0$ for every $g \in \mathcal{L}$, because $g - \underline{P}_{\mathcal{D}}(g) - 0 \in \mathcal{D}$ for every $g \in \mathcal{L}$, and $g - \underline{P}_{\mathcal{D}}(g) - \epsilon \notin \mathcal{D}$ for every $\epsilon \in \mathbb{R}^*_+$, $g \in \mathcal{L}$.

Now, consider $g, h \in \mathcal{L}$ such that $g \geq h$. Then:

$$\{c \in \mathbb{R} : h - c \in \mathcal{D}\} \subseteq \{c \in \mathbb{R} : g - c \in \mathcal{D}\},$$

because if $h - c \in \mathcal{D}$ then $g - c \geq h - c \in \mathcal{D}$ by D3**. Hence $\underline{P}_{\mathcal{D}}(g) \geq \underline{P}_{\mathcal{D}}(h)$. We show that $\underline{P}_{\mathcal{D}}(g) > \underline{P}_{\mathcal{D}}(h)$ if $g > h$ after proving that $\underline{P}_{\mathcal{D}}(g + r) = \underline{P}_{\mathcal{D}}(g) + r$ for every $g \in \mathcal{L}$ and $r \in \mathbb{R}$.

Consider $g + r$ for some $g \in \mathcal{L}$ and $r \in \mathbb{R}$. Then

$$\underline{P}_{\mathcal{D}}(g) + r \in \{c \in \mathbb{R} : g + r - c \in \mathcal{D}\}.$$

Indeed, $g + r - (\underline{P}_{\mathcal{D}}(g) + r) = g - \underline{P}_{\mathcal{D}}(g) \in \mathcal{D}$. However, $g + r - (\underline{P}_{\mathcal{D}}(g) + r + \epsilon) = g - \underline{P}_{\mathcal{D}}(g) - \epsilon \notin \mathcal{D}$ for every $\epsilon \in \mathbb{R}^*_+$. Hence, $\underline{P}_{\mathcal{D}}(g) + r = \max\{c \in \mathbb{R} : g + r - c \in \mathcal{D}\} =: \underline{P}_{\mathcal{D}}(g + r)$.

Consider now $g, h \in \mathcal{L}$ such that $g > h$. There exists a $\delta \in \mathbb{R}^*_+$ such that $g - \delta \geq h$. Then, $\underline{P}_{\mathcal{D}}(g - \delta) = \underline{P}_{\mathcal{D}}(g) - \delta \geq \underline{P}_{\mathcal{D}}(h)$. Hence $\underline{P}_{\mathcal{D}}(g) > \underline{P}_{\mathcal{D}}(h)$.

We have also that $\underline{P}_{\mathcal{D}}(0) = 0$. Indeed, if $\underline{P}_{\mathcal{D}}(0) = \epsilon > 0$, then $\underline{P}_{\mathcal{D}}(-\epsilon) = \underline{P}_{\mathcal{D}}(0 - \epsilon) = \underline{P}_{\mathcal{D}}(0) - \epsilon = \epsilon - \epsilon = 0$, by translation invariance. So, $-\epsilon \in F \cap \mathcal{D}$, which is a contradiction since $\mathcal{D}$ satisfies D2. If instead $\underline{P}_{\mathcal{D}}(0) < 0$, then $0 \notin \mathcal{D}$, which is a contradiction since $\mathcal{D}$ satisfies D1.

Finally, to show that $\underline{P}_{\mathcal{D}}(\cdot)$ is a uniformly continuous functional with respect to the supremum norm, we will show that:

$$\begin{aligned} &\forall \epsilon \in \mathbb{R}^*_+,\ \exists\, \delta = \epsilon \text{ such that } \forall g, h \in \mathcal{L} : \\ &\max_{\{\omega \in \Omega\}} |g(\omega) - h(\omega)| < \delta \text{ implies } |\underline{P}_{\mathcal{D}}(g) - \underline{P}_{\mathcal{D}}(h)| < \epsilon. \end{aligned}$$

Consider therefore a pair of gambles $g, h \in \mathcal{L}$ such that $\max_{\{\omega \in \Omega\}} |g(\omega) - h(\omega)| < \epsilon$ for some $\epsilon \in \mathbb{R}^*_+$. It implies in particular that $g < h + \epsilon$ and $h < g + \epsilon$. Therefore, from the previous points, we have: $\underline{P}_{\mathcal{D}}(g) < \underline{P}_{\mathcal{D}}(h) + \epsilon$ and $\underline{P}_{\mathcal{D}}(h) < \underline{P}_{\mathcal{D}}(g) + \epsilon$. Hence $|\underline{P}_{\mathcal{D}}(g) - \underline{P}_{\mathcal{D}}(h)| < \epsilon$.

Suppose now that $\mathcal{D}$ satisfies also D3*. Consider now $g, h \in \mathcal{L}$ and $\gamma \in [0, 1]$. Then

$$\gamma \underline{P}_{\mathcal{D}}(g) + (1 - \gamma)\underline{P}_{\mathcal{D}}(h) \in \{c \in \mathbb{R} : \gamma g + (1 - \gamma)h - c \in \mathcal{D}\}.$$

Indeed, $\gamma g + (1 - \gamma)h - \gamma \underline{P}_{\mathcal{D}}(g) - (1 - \gamma)\underline{P}_{\mathcal{D}}(h) = \gamma(g - \underline{P}_{\mathcal{D}}(g)) + (1 - \gamma)(h - \underline{P}_{\mathcal{D}}(h)) \in \mathcal{D}$. Hence $\gamma \underline{P}_{\mathcal{D}}(g) + (1 - \gamma)\underline{P}_{\mathcal{D}}(h) \leq \max\{c \in \mathbb{R} : \gamma g + (1 - \gamma)h - c \in \mathcal{D}\} =: \underline{P}_{\mathcal{D}}(\gamma g + (1 - \gamma)h)$.

The other properties follow from [5].

References

[1] F. J. Anscombe, R. J. Aumann, et al., A definition of subjective probability, Annals of Mathematical Statistics 34 (1) (1963) 199–205.

[2] B. de Finetti, La prévision: ses lois logiques, ses sources subjectives, in: Annales de l'institut Henri Poincaré, Vol. 7, 1937, pp. 1–68.

[3] F. P. Ramsey, Truth and probability, in: In R. B. Braithwaite, editor, The Foundations of Mathematics and other Logical Essays, chapter 7, Brace & Co., 1931, pp. 156–198.

[4] L. J. Savage, The Foundations of Statistics, Wiley, New York, 1954.

[5] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1991.

[6] P. M. Williams, Notes on conditional previsions, Tech. rep., University of Sussex (1975).

[7] R. Nau, et al., The shape of incomplete preferences, The Annals of Statistics 34 (5) (2006) 2430–2448.

[8] M. Zaffalon, E. Miranda, Axiomatising incomplete preferences through sets of desirable gambles, Journal of Artificial Intelligence Research 60 (2017) 1057–1126.

[9] M. Zaffalon, E. Miranda, Desirability foundations of robust rational decision making, Synthese 198 (27) (2021) 6529–6570.

[10] L. J. Neumann, O. Morgenstern, et al., Theory of Games and Economic Behavior, Vol. 60, Princeton University Press Princeton, 1947.

[11] F. J. Girón, S. Ríos, Quasi-bayesian behaviour: A more realistic approach to decision making?, Trabajos de Estadistica Y de Investigacion Operativa 31 (1) (1980) 17.

[12] R. J. Aumann, Utility theory without the completeness axiom, Econometrica: Journal of the Econometric Society (1962) 445–462.
[13] T. Galaabaatar, E. Karni, Subjective expected utility with incomplete preferences, Econometrica 81 (1) (2013) 255–284.
[14] T. Seidenfeld, M. J. Schervish, J. B. Kadane, et al., A representation of partially ordered preferences, The Annals of Statistics 23 (6) (1995) 2168–2217.
[15] A. Benavoli, A. Facchini, J. Vicente-Perez, M. Zaffalon, A polarity theory for sets of desirable gambles, in: Proc. ISIPTA'17 Int. Symposium on Imprecise Probability: Theories and Applications, PMLR, 2017, pp. 1–12.
[16] L. Blume, A. Brandenburger, E. Dekel, Lexicographic probabilities and choice under uncertainty, Econometrica: Journal of the Econometric Society (1991) 61–79.
[17] F. G. Cozman, Some remarks on sets of lexicographic probabilities and sets of desirable gambles, in: 9th ISIPTA, Pescara, Italy, 2015, pp. 87–96.
[18] T. Seidenfeld, M. J. Schervish, J. B. Kadane, Decisions without ordering, in: Acting and reflecting, Springer, 1990, pp. 143–170.
[19] A. Benavoli, A. Facchini, D. Piga, M. Zaffalon, Sum-of-squares for bounded rationality, International Journal of Approximate Reasoning 105 (2019) 130 – 152.
[20] A. Benavoli, A. Facchini, M. Zaffalon, The weirdness theorem and the origin of quantum paradoxes, Foundations of Physics 51 (95) (2021). doi:10.1007/s10701-021-00499-w.
[21] M. J. Schervish, T. Seidenfeld, J. B. Kadane, How sets of coherent probabilities may serve as models for degrees of incoherence, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 8 (03) (2000) 347–355.
[22] R. Nau, Indeterminate probabilities on finite sets, The Annals of Statistics (1992) 1737–1767.
[23] R. Pelessoni, P. Vicig, Uncertainty modelling and conditioning with convex imprecise previsions, International Journal of Approximate Reasoning 39 (2-3) (2005) 297–319.
[24] R. Pelessoni, P. Vicig, 2-coherent and 2-convex conditional lower previsions, International Journal of Approximate Reasoning 77 (2016) 66–86.
[25] E. Quaeghebeur, G. de Cooman, F. Hermans, Accept & reject statement-based uncertainty models, International Journal of Approximate Reasoning 57 (2015) 69–102.
[26] T. Augustin, F. P. A. Coolen, G. De Cooman, M. C. M. Troffaes, Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.
[27] G. De Cooman, Belief models: an order-theoretic investigation, Annals of Mathematics and Artificial Intelligence 45 (1-2) (2005) 5–34.
[28] M. C. M. Troffaes, G. De Cooman, Lower Previsions, John Wiley & Sons, 2014.
[29] B. A. Davey, H. A. Priestley, Introduction to Lattices and Order, Cambridge University Press, 2002.
[30] O. L. Mangasarian, E. W. Wild, Nonlinear knowledge-based classification, IEEE Transactions on Neural Networks 19 (10) (2008) 1826–1832.
[31] K. N. Oikonomou, P. D. Grünwald, Explicit bounds for entropy concentration under linear constraints, IEEE Transactions on Information Theory 62 (3) (2015) 1206–1230.
[32] D. Piga, A. Bemporad, A. Benavoli, Rao-Blackwellized sampling for batch and recursive Bayesian inference of Piecewise Affine models, Automatica 117 (2020) 109002.
[33] C. Kiselman, A semigroup of operators in convexity theory, Transactions of the American Mathematical Society 354 (5) (2002) 2035–2053.
[34] R. F. Nau, A generalization of Pratt-Arrow measure to nonexpected-utility preferences and inseparable probability and utility, Management Science 49 (8) (2003) 1089–1104.
[35] J. Quiggin, A theory of anticipated utility, Journal of Economic Behavior & Organization 3 (4) (1982) 323–343.
[36] S. H. Chew, K. MacCrimmon, Alpha utility theory: a generalization of expected utility theory, Faculty of Commerce and Business Administration, Working Paper 669 (1979).
[37] S. H. Chew, K. MacCrimmon, Alpha utility theory, lottery composition, and the Allais paradox, Faculty of Commerce and Business Administration, Working Paper 686 (1979).
[38] D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk, in: Handbook of the Fundamentals of Financial Decision Making: Part I, World Scientific, 2013, pp. 99–127.
[39] R. Greer, Chapter 2: A tutorial on polyhedral convex cones, in: R. Greer (Ed.), Trees and Hills: Methodology for Maximizing Functions of Systems of Linear Relations, Vol. 96 of North-Holland Mathematics Studies, North-Holland, 1984, pp. 15 – 81.
[40] R. T. Rockafellar, Convex Analysis, Vol. 36, Princeton University Press, 1970.
[41] I. E. L. Lewis, J. E. Edward, Geometry of Convex Sets, John Wiley & Sons, 2015.
[42] A. Schrijver, Theory of Linear and Integer Programming, John Wiley & Sons, 1998.