EXTREME POINTS OF THE CREDAL SETS GENERATED BY COMPARATIVE PROBABILITIES

ENRIQUE MIRANDA AND SEBASTIEN DESTERCKE

ABSTRACT. When using convex probability sets (or, equivalently, lower previsions) as uncertainty models, identifying extreme points can help simplifying various computations or the use of some algorithms. In general, sets induced by specific models such as possibility distributions, linear vacuous mixtures or 2-monotone measures may have extreme points easier to compute than generic convex sets. In this paper, we study extreme points of another specific model: comparative probability orderings between the singletons of a finite space. We characterise these extreme points by mean of a graphical representation of the comparative model, and use them to study the properties of the lower probability induced by this set. By doing so, we show that 2-monotone capacities are not informative enough to handle this type of comparisons without a loss of information. In addition, we connect comparative probabilities with other uncertainty models, such as imprecise probability masses.

Keywords: Comparative probabilities, credal sets, 2-monotone capacities, belief functions, extreme points, imprecise probability masses.

1. INTRODUCTION

In the last decades, there has been a growing interest in imprecise probability models as alternative models to probability in situations where the available information is vague or scarce. This type of models include for instance possibility measures [Dubois and Prade , 1988], belief functions [Shafer , 1976], 2- and n-monotone capacities [Choquet , 1953, Denneberg , 1994] and probability boxes [Destercke et al. , 2008, Ferson et al. , 2003]. All the examples above can be seen as particular cases of coherent lower and upper previsions [Walley , 1991].

The adequacy of each of these models for a particular problem depends, among other things, on the interpretation we are giving to our uncertainty. In this paper, we consider a *robust Bayesian* interpretation [Berger , 1994, Good , 1962]: we assume the existence of a precise, but unknown, probability model, and work with the set of probability measures that are compatible with the available information. This gives rise to a *credal set*, as considered by Levi [1980].

Here, we consider the case where the information is expressed by means of a *comparative probability model* [de Finetti , 1931, Koopman , 1940a,b]: we consider a finite possibility space $\mathscr{X}$ and assume that we are given judgements of the type "the probability of A is at least as great as that of B". Comparative probabilities have been deemed of particular interest within the context of subjective probability theory [Fine , 1973, 1979, Fishburn , 1986, Suppes , 1974, Walley and Fine , 1979]; we also refer to [Capotorti and Formisano , 2008, Christian et al. , 2007, Nehring , 2009] for some recent work and to Walley [1991, Section 4.5] for a study from the point of view of coherent lower previsions. One of their advantages is that they seem well suited for modelling qualitative judgements (e.g., expert opinions). Moreover, they have been shown to be more general than classificatory probabilities [Walley and Fine , 1979], and they can also be used to compare random quantities [Cohen , 1991].

In spite of this, there are only few works dealing with the numerical and practical aspects of comparative probabilities [Regoli , 1996]. One reason for this is that it is not easy to summarize the set of probabilities associated with the comparative assessments, for instance by means of a lower and an upper probability, and this renders it difficult to handle the information about the probability of the events of interest. In this paper, we solve this problem for the specific case of comparisons between the probabilities of singletons. We do so by characterizing the comparative probability models by means of the extreme points of their associated sets of probabilities. Characterizing such extreme points is instrumental in a number of applications of imprecise probabilities, including inferences for graphical [Cano and Moral , 2000] and statistical models [Walley , 1991, Sec. 8.5], and also to compute bounds of some functionals such as entropy [Abellán and Moral , 2005]. It is a problem that has been studied for other types of imprecise probability models, such as 2-monotone capacities [Chateauneuf and Jaffray , 1989], possibility measures [Miranda et al. , 2003], probability intervals [de Campos et al. , 1994] and belief functions [Dempster , 1967]; however, as we shall detail later, there is only one partial result for the case we shall consider in this paper [Gulordava , 2010].

After giving some preliminary results in Section 2, we show in Section 3 that, when the comparison judgements are made on the probabilities of the singletons, a graphical representation of these judgements makes it easy to derive the extreme points of the associated credal sets. In Section 4, we use this result to discuss some practical aspects of these models: we establish tight lower and upper bounds of the number of extreme points; investigate their relationship with other imprecise probability models; provide algorithms for the extraction of these extreme points; and discuss the computation of conditional lower probabilities and the merging of multiple

comparison judgments. Some additional remarks related to the practical use of these models and their extensions are provided in Section 5.

2. PRELIMINARIES

Consider a finite space $\mathscr{X} = \{x_1, \ldots, x_n\}$, modelling the set of outcomes of some experiment. In this paper, we assume that our information about these outcomes can be modelled by means of *comparative probability orderings of the states*, i.e., statements of the type "the probability of x_i is at least as great as that of x_j". Hence, we shall represent the available information by means of a subset $\mathscr{L}$ of $\{1, \ldots, n\} \times \{1, \ldots, n\}$ modelling the (pre)order relation between the states.

The set of probability measures compatible with this information is given by

(1) $$\mathscr{P}(\mathscr{L}) = \{P \in \mathbb{P}_{\mathscr{X}} : \forall (i,j) \in \mathscr{L}, P(x_i) \geq P(x_j)\},$$

where $\mathbb{P}_{\mathscr{X}}$ denotes the set of all probability measures on $\mathscr{X}$. This set is called the *natural extension* of the ordering by Walley [1991, Section 4.5.1]. This can be equivalently stated by saying that we consider a preorder $\succeq$ between the singletons and we want to characterise the set of probability measures P that agree with this order, in the sense that

$$x_i \succeq x_j \Rightarrow P(x_i) \geq P(x_j).$$

For the purposes of this paper, it shall be useful to represent these assessments by means of a graph $\mathscr{G} = (\mathscr{X}, \mathscr{L})$ where the nodes are the elements of $\mathscr{X}$ and we draw an edge between x_i and x_j when $(i,j) \in \mathscr{L}$.

Example 1. Consider the space $\mathscr{X} = \{x_1, \ldots, x_5\}$ and the assessments

$$\mathscr{L} = \{(1,3), (1,4), (2,5), (4,5)\}.$$

The acyclic graph $\mathscr{G}$ associated with $\mathscr{L}$ is given by Figure 1.
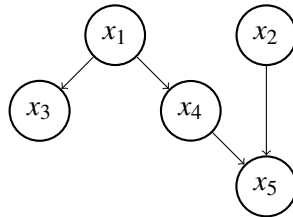

FIGURE 1. Graph $\mathscr{G}$ of Example 1.♦

Note that the credal set $\mathscr{P}(\mathscr{L})$ determined by Eq. (1) is always non-empty, because it includes for instance the uniform probability distribution. Further, the set $\mathscr{P}(\mathscr{L})$ coincides with the set $\mathscr{P}(\mathscr{C}(\mathscr{L}))$ determined by the *transitive closure* $\mathscr{C}(\mathscr{L})$ of $\mathscr{L}$, as the additional constraints of $\mathscr{C}(\mathscr{L})$ are redundant with those of $\mathscr{L}$: since any model in $\mathscr{P}(\mathscr{L})$ (a probability measure) is transitive, so should be the relationship $\succeq$ associated with $\mathscr{L}$.

It is interesting to compare $\mathscr{P}(\mathscr{L})$ with the set

$$\widetilde{\mathscr{P}}(\mathscr{L}) = \{P \in \mathbb{P}_{\mathscr{X}} : \forall (i,j) \in \mathscr{L}, P(x_i) > P(x_j)\},$$

i.e., with the credal set associated with *strict* probability comparisons, which also appear sometimes in the literature [Fishburn , 1986]. Since $\mathscr{P}(\mathscr{L})$ is a convex polytope in $\mathbb{R}^n$, it follows from basic convex analysis that $\mathscr{P}(\mathscr{L})$ corresponds to the closure of $\widetilde{\mathscr{P}}(\mathscr{L})$ when the latter set is non-empty, and this non-emptiness is easy to characterise.

Lemma 1. $\widetilde{\mathscr{P}}(\mathscr{L}) \neq \emptyset$ *if and only if its associated graph $\widetilde{\mathscr{G}}$ is acyclic.*

Proof. "Only if": $\widetilde{\mathscr{G}}$ cyclic means that there are at least two indices i, j such that $P(x_i) > P(x_j)$ and $P(x_i) < P(x_j)$, leading to an inconsistency.

"If": if $\widetilde{\mathscr{G}}$ is acyclic, then it can be associated with a preorder over the probability masses $P(x_i)$. We can then take a linear extension of this pre-order and associate it with a permutation σ of $\{1, \ldots, n\}$ such that

$$P(x_{\sigma(1)}) < P(x_{\sigma(2)}) < \ldots < P(x_{\sigma(n)});$$

then, it is easy to see that there exists a probability satisfying all these constraints: we may for instance consider the probability measure associated with the probability mass $P(x_{\sigma(i)}) = 1/n - (n-i)\varepsilon$ for $i = 1, \ldots, n-1$ and $P(x_{\sigma(n)}) = 1/n + \varepsilon + \ldots + (n-1)\varepsilon$ with $\varepsilon \in (0, \frac{1}{n^2})$. $\square$

Indeed, a cyclic graph $\widetilde{\mathscr{G}}$ is incompatible with the irreflexive property that strict comparative assessments must satisfy. Nevertheless, in this paper we shall focus on non-strict comparative assessments, and for those the associated graph $\mathscr{G}$ may possess cycles, as we shall discuss later.

Note also that we can straightforwardly connect our current model with the axiomatic view of comparative probabilities [de Finetti , 1931]. From $\mathscr{L}$, we can obtain a probability ordering $\succeq$ over subsets of $\mathscr{X}$ such that $A \succeq B$ whenever $P(A) \geq P(B)$ for all $P \in \mathscr{P}(\mathscr{L})$. Using [Walley , 1991, Sec. 4.5.3.], this probability ordering satisfies a number of properties, in particular all axioms required by de Finetti [1931] except for the one of completeness. Hence, while we focus in this paper on the numerical aspects associated with specific comparative probabilities, we are completely in-line with the axiomatic view.

3. EXTREME POINTS OF $\mathscr{P}(\mathscr{L})$

In this section, we characterise the extreme points of the credal set $\mathscr{P}(\mathscr{L})$ associated with a number of probability comparisons on the singletons. Consider a finite space $\mathscr{X} = \{x_1, \ldots, x_n\}$ and a subset $\mathscr{L}$ of $\{1, \ldots, n\} \times \{1, \ldots, n\}$, and let $\mathscr{P}(\mathscr{L})$ be the set it determines by means of Eq. (1). Any of the probability measures in $\mathscr{P}(\mathscr{L})$ is completely determined by its probability mass, and as a consequence it can be seen as an element of the n-dimensional Euclidean space (more specifically of the $n-1$ unit simplex). Then, $\mathscr{P}(\mathscr{L})$ is a closed convex subset of $\mathbb{R}^n$ in the Euclidean topology, which is equal to the closed convex hull of its set of extreme points. That is, we can characterize the set of functions $P \colon \mathscr{X} \to [0,1]$ such that $(i,j) \in \mathscr{L} \Rightarrow P(x_i) \geq P(x_j)$ and $\sum_{x \in \mathscr{X}} P(x) = 1$ by means of its extreme points[1]. Recall that P is an *extreme point* of $\mathscr{P}(\mathscr{L})$ when for every $P_1, P_2 \in \mathscr{P}(\mathscr{L})$ and every $\alpha \in (0,1)$,

$$P = \alpha P_1 + (1-\alpha)P_2 \iff P_1 = P_2 = P,$$

that is, when P cannot be expressed as the convex combination of two other points of $\mathscr{P}(\mathscr{L})$. We shall determine these extreme points by means of the graphical representation we have established in Section 2. In order to do this, we consider a number of lemmas:

Lemma 2. *Any extreme point P of $\mathscr{P}(\mathscr{L})$ corresponds to a uniform probability measure concentrated over some subset $A \subseteq \mathscr{X}$.*

Proof. The (linear) constraints determining $\mathscr{P}(\mathscr{L})$ are:

$$\mathscr{C} = \{P(x_i) \geq P(x_j) \ \forall (i,j) \in \mathscr{L}\}$$
$$\mathscr{B} = \{0 \leq P(x) \leq 1 \ \forall x \in \mathscr{X}\}$$
$$\mathscr{S} = \{\sum_{x \in \mathscr{X}} P(x) = 1\}$$

Also recall that, in order to determine an extreme point P of $\mathscr{P}(\mathscr{L})$, its probability mass must saturate $n = |\mathscr{X}|$ linearly independent constraints delimiting $\mathscr{P}(\mathscr{L})$ [Bertsimas and Tsitsiklis , 1997, Section 2.2.], that is, constraints in $\mathscr{C} \cup \mathscr{B} \cup \mathscr{S}$. Clearly, $\mathscr{S}$ is always saturated.

Now, let us show that in order to saturate n constraints, a probability measure $P \in \mathscr{P}(\mathscr{L})$ must be uniform on some subset $A \subseteq \mathscr{X}$. Without loss of generality, assume that there are ℓ elements $P(x_1), \ldots, P(x_\ell)$ that are

---

[1]In this respect, the problem under study is somewhat related to that in *random utility representations*, where a partial rank among the alternatives is established and the goal is to infer from it the probability of each alternative being optimal. See Fishburn [1989], Koopen [1995] and Marschak [1960] for some works in this field.

non-null. This means that, unless $\ell = 1$, none of them saturates a constraint in $\mathscr{B}$. The result then follows taking into account that:

- as the $n - \ell$ probabilities $P(x_{\ell+1}), \ldots, P(x_n)$ are zero, they saturate $n - \ell$ inequalities in $\mathscr{B}$ (that are, by definition, linearly independent, as they bear on disjoint singletons);
- saturating inequalities in $\mathscr{C}$ can only be done by equating some probabilities $P(x_i) = P(x_j)$;
- any set of m equal probabilities will saturate at most $m - 1$ linearly independent inequalities in $\mathscr{C}$, forming the chain

$$P(x_1) = \ldots = P(x_m).$$

Note that, even if two null probabilities also saturate constraints in $\mathscr{C}$, they will only saturate linearly dependent constraints, because the restrictions $P(x_i) - P(x_j) \geq 0, P(x_i) \geq 0$ and $P(x_j) \geq 0$ are linearly dependent: $P(x_i) - P(x_j) = 1 \cdot P(x_i) + (-1) \cdot P(x_j)$. Therefore we can omit those constraints in $\mathscr{C}$ that are saturated by null probabilities. This means that, if the ℓ non-null probabilities are not all equal, they saturate $k < \ell - 1$ linearly independent inequalities in $\mathscr{C}$; hence, in that case the total number of linearly independent inequalities in $\mathscr{C} \cup \mathscr{B} \cup \mathscr{S}$ that are saturated is

$$k + n - \ell + 1 < \ell - 1 + n - \ell + 1 = n.$$

Therefore, a probability measure P that is not uniform on some subset $A \subseteq \mathscr{X}$ cannot be an extreme point. $\qquad \square$

For every subset A of $\mathscr{X}$, we shall denote by P_A the uniform probability measure on A, which is determined by the values

$$(2) \qquad\qquad P_A(x_i) = \begin{cases} \frac{1}{|A|} & \text{if } x_i \in A \\ 0 & \text{otherwise} \end{cases}$$

for any $i \in \{1, \ldots, n\}$. Using the graph $\mathscr{G}$, we can now characterize those subsets $A \subseteq \mathscr{X}$ for which P_A is an extreme point of $\mathscr{P}(\mathscr{L})$. For every $x_j \in \mathscr{X}$, we shall denote by $H(x_j)$ the set of ancestors of x_j, i.e., those nodes x_i such that there is a directed path going from x_i to x_j in $\mathscr{G}$. By an abuse of notation, we shall also consider that x_j is an ancestor of itself, i.e., we shall assume that $x_j \in H(x_j)$ for all j. Finally, for every $A \subseteq \mathscr{X}$, we shall denote $H(A) := \cup_{x \in A} H(x)$. Using the terminology from Cozman [2000], we shall refer to $H(A)$ as the *top subnetwork* of $\mathscr{G}$ generated by A. Our next result shows that we can restrict our attention to uniform probability measures on top subnetworks.

Lemma 3. *If $A \neq H(A)$, then P_A is not an extreme point of $\mathscr{P}(\mathscr{L})$.*

Proof. If A is a proper subset of $H(A)$, then we can find $x_i, x_j \in \mathscr{X}$ such that $(i, j) \in \mathscr{L}$ and $x_i \in H(A) \setminus A$, $x_j \in A$. Since $(i, j) \in \mathscr{L}$, this means that we have made the assessment $P(x_i) \geq P(x_j)$. On the other hand, given the uniform probability distribution P_A, it follows from Eq. (2) that

$$P_A(x_i) = 0 < P_A(x_j) = \frac{1}{|A|}.$$

Hence, P_A does not belong to $\mathscr{P}(\mathscr{L})$ and as a consequence it cannot be an extreme point. $\qquad\square$

Example 2. Consider Example 1 and the set $A = \{x_2, x_4\}$. As $H(A) = \{x_1, x_2, x_4\} \neq A$, the probability measure associated with the probability mass $P(x_2) = P(x_4) = 1/2$ is not an extreme point. $\blacklozenge$

In order to simplify the notation, we shall consider the set

$$\mathscr{H} := \{\emptyset \neq B \subseteq \mathscr{X} : B = H(A) \text{ for some } A\}.$$

Lemma 3 implies that we can focus on the elements of $\mathscr{H}$ in order to identify the extreme points. Our next result shows that not all of those sets will give rise to an extreme point of $\mathscr{P}(\mathscr{L})$.

Lemma 4. *Consider a set $B \in \mathscr{H}$. If there are $B_1, B_2 \in \mathscr{H}$ such that $B_1 \cap B_2 = \emptyset$ and $B_1 \cup B_2 = B$, then P_B is not an extreme point on $\mathscr{P}(\mathscr{L})$.*

Proof. First of all, let us show that for every $B \in \mathscr{H}$, P_B belongs to $\mathscr{P}(\mathscr{L})$. Consider any $(i, j) \in \mathscr{L}$. There are two possibilities: if $x_j \in B$, then since B is a top subnetwork also x_i belongs to B, whence $P_B(x_i) = P_B(x_j) = \frac{1}{|B|}$. On the other hand, if $x_j \notin B$, then $P_B(x_j) = 0 \leq P_B(x_i)$, and therefore the constraint is also satisfied.

Now, if B is the disjoint union of B_1 and B_2, we can deduce from Eq. (2) that

$$P_B = \frac{|B_1|}{|B|} P_{B_1} + \frac{|B_2|}{|B|} P_{B_2};$$

since $|B_1| + |B_2| = |B|$, we deduce that P_B is a convex mixture of two different probabilities in $\mathscr{P}(\mathscr{L})$, and as a consequence it cannot be an extreme point. $\qquad\square$

Example 3. Consider Example 1 and the set $B = \{x_1, x_2, x_4\}$. Then given the sets $B_1 = \{x_1, x_4\}$, $B_2 = \{x_2\}$, it holds that $B, B_1, B_2 \in \mathscr{H}$, $B_1 \cap B_2 = \emptyset$ and $B_1 \cup B_2 = B$. From this we can derive that $P_B = 2/3 P_{B_1} + 1/3 P_{B_2}$ and therefore it is not an extreme point of $\mathscr{P}(\mathscr{L})$. $\blacklozenge$

The top subnetwork associated with $B \in \mathscr{H}$ is not *weakly connected*[2] if and only if we can decompose it in the manner depicted in Lemma 4,

---

[2]By weakly connected we mean that for every two nodes in the graph associated with B there is an undirected path between them.

by considering its weakly connected components (i.e., the maximal weakly connected subsets). Thus, we can restrict our attention to weakly connected top subnetworks of $\mathscr{G}$. This leads to the main result in this section, where we establish that any of these sets determines an extreme point of $\mathscr{P}(\mathscr{L})$:

Theorem 1. *The set of extreme points coincides with the set of probabilities P_B generated by sets $B \in \mathscr{H}$ that cannot be decomposed as in Lemma 4.*

Proof. From Lemmas 2 and 4 any extreme point of $\mathscr{P}(\mathscr{L})$ corresponds to a uniform distribution P_B, where $B \in \mathscr{H}$ cannot be decomposed as the disjoint union of other elements of $\mathscr{H}$. Let us prove that, conversely, given a set B of this type the probability measure P_B it determines is indeed an extreme point.

Assume *ex absurdo* that this is not the case for some set B. This means that there are $m \geq 2$, $B_1, \ldots, B_m \in \mathscr{H}$ and $\alpha_1, \ldots, \alpha_m > 0$ s.t. $\sum_{i=1}^m \alpha_i = 1$ and

$$(3) \qquad\qquad P_B = \alpha_1 P_{B_1} + \ldots + \alpha_m P_{B_m}.$$

It must be $B_i \subseteq B$ for all $i = 1, \ldots, m$: otherwise, there would be some node x_j such that $P_{B_i}(x_j) > 0 = P_B(x_j)$, a contradiction with (3). We deduce also that B_i must be a proper subset of B for all $i = 1, \ldots, m$, or we would obtain that P_B and P_{B_i} coincide.

Denote $C = B \setminus B_i$. Then $H(C) \subseteq B$, whence

$$B = B_i \cup C \subseteq B_i \cup H(C) \subseteq B$$

and therefore $B = B_i \cup H(C)$. Since $H(C) \in \mathscr{H}$, we deduce that it must be $H(C) \cap B_i \neq \emptyset$, since otherwise B would be the disjoint union of two elements of $\mathscr{H}$, a contradiction with the assumptions we have made on B. Therefore, there is some $x_k \in C$ such that $H(x_k) \cap B_i \neq \emptyset$. Given $x_j \in B_i \cap H(x_k)$, we have that $(j,k) \in \mathscr{C}(\mathscr{L})$, and $x_j \in B_i, x_k \notin B_i$. As a consequence,

$$P_{B_i}(x_j) > P_{B_i}(x_k) = 0,$$

whence

$$\sum_{\ell=1}^m \alpha_\ell (P_{B_\ell}(x_j) - P_{B_\ell}(x_k)) > 0 = P_B(x_j) - P_B(x_k),$$

taking into account that $P_{B_\ell}(x_j) - P_{B_\ell}(x_k) \geq 0$ for every $\ell = 1, \ldots, m$ because P_{B_ℓ} belongs to $\mathscr{P}(\mathscr{L})$. This is a contradiction. Hence, Eq. (3) does not hold and therefore P_B is an extreme point. $\qquad\square$

Remark 1. An interesting related result has been established by Gulordava [2010], in the context of credal classification. There, it is considered the credal set determined by the comparisons of the probabilities of the states, and it is computed the lower probability of the set A of elements with no predecessor in $\mathscr{G}$. In order to determine this lower probability, she provides

results analogous to our Lemmas 2 and 3, and then in [Gulordava , 2010, Theorem B.2.2] she establishes which elements in $\mathscr{P}(\mathscr{L})$ attain the lower probability of A.

Our previous result subsumes these results, in the sense that we give the explicit form of the extreme points (from which we may determine also the lower probability of any other set, as well as the lower prevision induced by a comparative probability model). Note moreover that we have shown that not all the uniform probability distributions $P_{H(A)}$ determine an extreme point of $\mathscr{P}(\mathscr{L})$. ♦

Example 4. The extreme points generated by Example 1 are summarized in Table 1:

B	x_1	x_2	x_3	x_4	x_5
$\{x_1\}$	1	0	0	0	0
$\{x_2\}$	0	1	0	0	0
$\{x_1,x_3\}$	1/2	0	1/2	0	0
$\{x_1,x_4\}$	1/2	0	0	1/2	0
$\{x_1,x_2,x_4,x_5\}$	1/4	1/4	0	1/4	1/4
$\{x_1,x_3,x_4\}$	1/3	0	1/3	1/3	0
$\{x_1,x_2,x_3,x_4,x_5\}$	1/5	1/5	1/5	1/5	1/5

TABLE 1. Extreme points of Example 1. ♦

To conclude this section, we discuss briefly a couple of ways in which the procedure above may be simplified. On the one hand, when the graph $\mathscr{G}$ is not weakly connected, we can decompose it as a union of its weakly connected components $\mathscr{G}_1,\ldots,\mathscr{G}_k$. Each of these components $\mathscr{G}_i$ will be a graph on the elements of a respective subset $\mathscr{X}_i$ of $\mathscr{X}$, and the sets $\{\mathscr{X}_1,\ldots,\mathscr{X}_k\}$ will form a partition of $\mathscr{X}$. Then we can characterise the extreme points associated with $\mathscr{G}_i$ in the form given by Theorem 1, and extend them from $\mathscr{X}_i$ to $\mathscr{X}$ by assigning zero mass to the elements of $\mathscr{X} \setminus \mathscr{X}_i$. The extreme points associated with $\mathscr{G}$ will be the union of the sets of extreme points obtained in this way. This is formalised in the following result:

Proposition 2. *Let $\mathscr{P}_i \subseteq \mathbb{P}_{\mathscr{X}_i}$ be the set of extreme points associated with the graph $\mathscr{G}_i$ by means of Theorem 1. Then the extreme points of $\mathscr{P}(\mathscr{L})$ are given by*

$$\cup_{i=1}^k \{ext_i(P_i) : P_i \in \mathscr{P}_i\},$$

where $ext_i(P_i)$ is the element of $\mathbb{P}_{\mathscr{X}}$ given by

$$(4) \qquad ext_i(P_i)(x_j) = \begin{cases} P_i(x_j) & \text{if } x_j \in \mathscr{X}_i \\ 0 & \text{otherwise.} \end{cases}$$

Proof. It suffices to note that a subset of $\mathscr{G}$ can only be weakly connected when it is included in one of the weakly connected components $\mathscr{G}_i$. Then, the set of weakly connected top subnetworks of $\mathscr{G}$ is given by the union of the weakly connected top subnetworks of $\mathscr{G}_1, \ldots, \mathscr{G}_k$. Each of these subnetworks determines a probability measure in one of $\mathscr{X}_1, \ldots, \mathscr{X}_k$ that is in a one-to-one correspondence with a probability measure on $\mathscr{X}$, in the manner depicted in Eq. (4). $\qquad\square$

On the other hand, we can also assume without loss of generality that the graph $\mathscr{G}$ associated with our set of assessments $\mathscr{L}$ is acyclic. When this is not the case, we have some assumptions of equality $P(x_i) = P(x_j)$ between the probabilities of two elements x_i, x_j in our possibility space.

If we define the equivalence relationship

$$(5) \qquad x_i R x_j \iff P(x_i) = P(x_j) \; \forall P \in \mathscr{P}(\mathscr{L}),$$

then we can consider a representative x_i for each equivalence class, so that we end up with a subset $\mathscr{X}'$ of $\mathscr{X}$. If we denote $n_i = |[x_i]_R|$ the number of elements in the equivalence class associated with x_i, then it holds that $\sum_{x_i \in \mathscr{X}'} n_i = n = |\mathscr{X}|$. Let

$$(6) \qquad \mathscr{L}' := \{(i,j) \in \mathscr{C}(\mathscr{L}) : x_i, x_j \in \mathscr{X}'\},$$

be the set of comparative probability assessments on this subset of $\mathscr{X}$, and let $\mathscr{G}'$ be its associated (acyclic) graph.

Proposition 3. *There is a one-to-one correspondence between the extreme points of $\mathscr{P}(\mathscr{L})$ and of $\mathscr{P}(\mathscr{L}')$.*

Proof. First of all, note that for any $C \subseteq \mathscr{X}$ whose associated graph is a top subnetwork of $\mathscr{G}$, it holds that $C = \cup_{x_i \in C}[x_i]_R$.

For any $B \subseteq \mathscr{X}'$, we can define

$$exp(B) = \cup_{x_i \in B}[x_i]_R \subseteq \mathscr{X}.$$

Then the graph associated with a set $B \subseteq \mathscr{X}'$ is a weakly connected top-subnetwork of $\mathscr{G}'$ if and only if the graph associated with $exp(B) \subseteq \mathscr{X}$ is a weakly connected top-subnetwork of $\mathscr{G}$. Conversely, if $C \subseteq \mathscr{X}$ determines a weakly connected top subnetwork in $\mathscr{G}$, then $B := C \cap \mathscr{X}'$ determines a weakly connected top subnetwork in $\mathscr{G}'$, and moreover $exp(B) = C$.

This shows that there is a one-to-one correspondence between the weakly connected top subnetworks of $\mathscr{G}$ and those of $\mathscr{G}'$, and as a consequence between the extreme points of $\mathscr{P}(\mathscr{L})$ and those of $\mathscr{P}(\mathscr{L}')$.   □

The following example illustrates this procedure:

Example 5. Consider the space $\mathscr{X} = \{x_1, x_2, x_3, x_4\}$, and the assessments
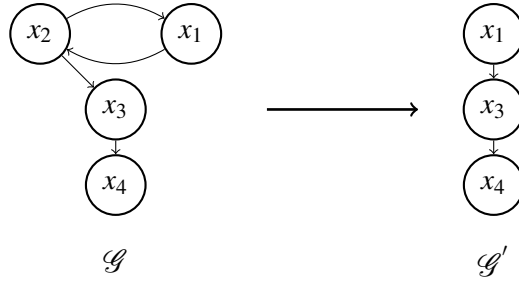
$$\mathscr{L} = \{(1,2),(2,3),(2,1),(3,4)\}.$$

FIGURE 2. Graphs associated with $\mathscr{L}$ and $\mathscr{L}'$

We see that the associated graph (Figure 2) has a cycle between x_1 and x_2. Hence, we can consider the alternative graph $\mathscr{G}'$ related to the space $\mathscr{X}' = \{x_1, x_3, x_4\}$ and the assessments $\mathscr{L}' = \{(1,3),(3,4),(1,4)\}$ determined by Eq. (6) (note that the assessment (1,4) is actually redundant).

It follows from Theorem 1 that the extreme points of $\mathscr{P}(\mathscr{L}')$ are given by Table 2:

		p	
B	x_1	x_3	x_4
$\{x_1\}$	1	0	0
$\{x_1,x_3\}$	$1/2$	$1/2$	0
$\{x_1,x_3,x_4\}$	$1/3$	$1/3$	$1/3$

TABLE 2. Extreme points of $\mathscr{P}(\mathscr{L}')$

and from this we deduce the extreme points of $\mathscr{P}(\mathscr{L})$ given by Table 3.

$exp(B)$	p			
	x_1	x_2	x_3	x_4
$\{x_1,x_2\}$	$1/2$	$1/2$	0	0
$\{x_1,x_2,x_3\}$	$1/3$	$1/3$	$1/3$	0
$\{x_1,x_2,x_3,x_4\}$	$1/4$	$1/4$	$1/4$	$1/4$

TABLE 3. Extreme points of $\mathscr{P}(\mathscr{L})$

Note that all these extreme points correspond again to uniform probability distributions. ♦

Remark 2. The comments above imply that for every $x_i \in \mathscr{X}$ its associated top subnetwork $B_i = H(\{x_i\}) \in \mathscr{H}$ determines an extreme point of $\mathscr{P}(\mathscr{L})$: if it was the disjoint union of two top subnetworks C_1, C_2 then at least one of them should include x_i, and as a consequence also B_i, a contradiction. When $\mathscr{G}$ is acyclic, all these top subnetworks are different. ♦

Remark 3. When $\mathscr{G}$ is weakly connected and acyclic, any top subnetwork $B \in \mathscr{H}$ associated with an extreme point in the manner detailed in Theorem 1 is in correspondence with a set A of strongly disconnected[3] but weakly connected nodes in $\mathscr{G}$, by means of the equality $B = H(A)$: this set is given by the nodes in B that have no successor in B, i.e., the leaves in the associated sub-graph. That this set is non-empty follows from the fact that $\mathscr{G}$ is acyclic, and then it is trivial that $H(A) = B$. To see that any extreme point is associated with only one set of strongly disconnected nodes, note that it is impossible for two distinct sets of strongly disconnected nodes A, A' to satisfy $H(A) = H(A')$ (and thus determine the same extreme point): if $x_i \in A \setminus A'$, then since $x_i \in H(A) = H(A')$ there must be some $x_j \in A'$ such that $x_i \in H(x_j)$; but $x_j \in H(A') = H(A)$ implies that there is some $x_k \in A$ such that $x_j \in H(x_k)$. Note that it cannot be $x_k = x_i$, because then we would obtain that $x_i \in H(x_j)$ and $x_j \in H(x_i)$, meaning that there is a cycle involving x_i, x_j, a contradiction. On the other hand, if $x_i \neq x_j$ then we obtain that $x_i \in H(x_k)$, which contradicts the assumption that the nodes in A are strongly disconnected.

On the other hand, when $\mathscr{G}$ is a weakly connected and acyclic graph, there is a one-to-one correspondence between sets of strongly disconnected nodes and the antichains of the partial order induced by $\mathscr{L}$, not counting $\emptyset$ as an antichain. This means that the number of antichains of $\mathscr{G}$ is an upper bound

---

[3]Two nodes in the graph are strongly disconnected if there is no directed path between them.

of the number of extreme points of $\mathscr{P}(\mathscr{L})$. Note however, that not all sets of strongly disconnected nodes give rise to an extreme point, because it is also necessary that the associated top subnetwork is non-decomposable, as we have established in Lemma 4. ♦

4. PRACTICAL ASPECTS

In this section, we explore various practical aspects (number and extraction of extreme points, n-monotonicity properties, conditioning, merging of information) of comparative probability models on singletons.

4.1. **Number of extreme points.** Since extreme points correspond to uniform distributions over certain subsets $A \subseteq \mathscr{X}$, we immediately see that an upper bound of the number of extreme points is $2^{|\mathscr{X}|} - 1$. Note that this is significantly lower than the maximal number of extreme points generated by lower coherent probabilities, known to be $|\mathscr{X}|!$ [Wallner , 2007]. We next show that this number of extreme points can be reduced even further:

Theorem 4. *The maximal number of extreme points of $\mathscr{P}(\mathscr{L})$ is $2^{(|\mathscr{X}|-1)}$.*

Proof. Consider first of all the case where the graph $\mathscr{G}$ is weakly connected and acyclic.

Remark 3 tells us that the number of extreme points of $\mathscr{P}(\mathscr{L})$ corresponding to a weakly connected and acyclic graph $\mathscr{G}$ is bounded above by its number of antichains. Therefore, determining the maximal number of antichains of $\mathscr{G}$ provides an upper bound on the number of extreme points. Dilworth's Theorem [Dilworth , 1950] tells us that for a finite partially ordered set whose maximal antichain has cardinality a, there exists a minimum number a of disjoint chains of elements partitioning the set $\mathscr{X}$. Every antichain contains at most one element of these disjoint chains, and assuming they have $c_1, \ldots, c_a$ elements, the number of antichains (excluding the empty set) is at most

$$(7) \qquad (c_1 + 1)(c_2 + 1) \ldots (c_a + 1) - 1$$

Taking this idea into account, we can also bound the number of antichains by the number above as soon as there is a partition of ℓ disjoint chains (even if ℓ may be different from a) of elements partitioning the set $\mathscr{X}$, with respective cardinalities $c_1, \ldots, c_\ell$. We shall use this to prove that the number of antichains is bounded by $2^{|\mathscr{X}|-1}$.

In this respect, we shall use that for every natural number $k \geq 3$ it holds that $k + 1 \leq 2^{k-1}$, and for $k = 1, 2$ it holds that $k + 1 \leq 2^k$. There are two possibilities:

(a) Assume first of all that there is a partition of $\mathscr{X}$ made up of chains of respective lengths $c_1, \ldots, c_\ell$, and that $c_j \geq 3$ for $j = 1, \ldots, i$. Then by Eq. (7) the maximal number of antichains is bounded by

$$(c_1 + 1)(c_2 + 1) \ldots (c_\ell + 1) \leq 2^{\sum_{j=1}^{i} c_j - 1} \cdot 2^{\sum_{j=i+1}^{\ell} c_j} \leq 2^{(\sum_{j=1}^{\ell} c_j) - 1} = 2^{|\mathscr{X}| - 1},$$

taking into account that $i \geq 1$.

(b) If there is no chain with length greater than or equal to 3, then it is possible to find a partition of $\mathscr{X}$ made up of chains of respective lengths $c_1, \ldots, c_\ell$, and such that $c_j = 2$ for $j = 1, \ldots, i$ and $c_j = 1$ for $j = i+1, \ldots, \ell$. Note that it must be $i \geq 1$ because we are assuming that the graph $\mathscr{G}$ is connected.

Eq. (7) gives us the upper bound $3^i \cdot 2^{|\mathscr{X}| - 2i} - 1$ for the number of antichains. However, since $\mathscr{G}$ is connected, given $j \in \{i+1, \ldots, \ell\}$ the node in c_j must be a successor of at least one of the top elements of $c_1, \ldots, c_i$. Hence given any non-empty subset of $\{c_{i+1}, \ldots, c_\ell\}$, there is at least one selection of nodes in $\{c_1, \ldots, c_i\}$ for which the combination does not produce an antichain. This allows us to reduce the number of antichains by $2^{|\mathscr{X}| - 2i} - 1$, meaning that the upper bound can be refined into

$$3^i \cdot 2^{|\mathscr{X}| - 2i} - 1 - (2^{|\mathscr{X}| - 2i} - 1) = (3^i - 1)2^{|\mathscr{X}| - 2i} \leq 2^{|\mathscr{X}| - 1},$$

taking into account that $3^i - 1 \leq 2^{2i-1}$ for any natural number $i \geq 1$.

To see that the bound given by the theorem above can indeed be reached, consider the case where a single modal value is provided[4], that is, of $\mathscr{L} = \{(1, j) : j = 2, \ldots, |\mathscr{X}|\}$. Figure 3 illustrates the situation for five elements. In this case, the extreme points are obtained by the top subnetwork induced by x_1 and all non-empty subsets of $\{x_2, \ldots, x_{|\mathscr{X}|}\}$ (whose number is $2^{|\mathscr{X}| - 1} - 1$); that is there are $2^{|\mathscr{X}| - 1}$ extreme points.
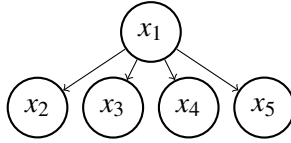

FIGURE 3. Graph $\mathscr{G}$ for $x_1 =$ modal value

Next, if $\mathscr{G}$ is weakly connected but has cycles, we consider the equivalence relationship in Eq. (5), and let $\mathscr{X}'$ be a subset of $\mathscr{X}$ obtained by considering one element from each equivalence class. If we let $\mathscr{L}'$ be given by Eq. (6), then there is a one-to-one correspondence between the

---

[4]This is an instance of a *naive tree*.

extreme points of $\mathscr{P}(\mathscr{L})$ and those of $\mathscr{P}(\mathscr{L}')$. Applying the first part of the proof, we conclude that the number of extreme points is bounded by $2^{|\mathscr{X}'|-1} \leq 2^{|\mathscr{X}|-1}$.

Finally, if the graph $\mathscr{G}$ is not weakly connected, we can express it as the union of its weakly connected components $\mathscr{G}_1, \ldots, \mathscr{G}_k$, which determine a partition $\{\mathscr{X}_1, \ldots, \mathscr{X}_k\}$. The number of extreme points of $\mathscr{P}(\mathscr{L})$ is equal to the sum of the number of extreme points of $\mathscr{P}(\mathscr{L}_i)$ for $i = 1, \ldots, k$, where $\mathscr{P}(\mathscr{L}_i)$ is the set of probability measures associated with $\mathscr{G}_i$. Applying the previous part, this number is bounded by

$$2^{|\mathscr{X}_1|-1} + \cdots + 2^{|\mathscr{X}_k|-1} \leq 2^{\sum_{i=1}^k |\mathscr{X}_i|-1} = 2^{|\mathscr{X}|-1},$$

taking into account that the function $f(x) = 2^x$ is subadditive.   □

Interestingly, the bound given in Theorem 4 is the same as the number of extreme points of the set of probability measures dominated by a possibility measure, as shown by Miranda et al. [2003, Section 5], and it is smaller than the one ($|\mathscr{X}|!$) for the number of extreme points associated with belief functions [Dempster , 1967]. Our intuition for this is that possibility measures also determine an order between the singletons, by means of their associated possibility distributions.

On the other hand, the number of extreme points can be much lower. To see this, consider the equivalence relationship given by Eq. (5), and let $\mathscr{X}'$ be a subset of $\mathscr{X}$ obtained by considering one element from each equivalent class. The ideas in Remark 2 lead easily to the following result:

Proposition 5. *The minimum number of extreme points of $\mathscr{P}(\mathscr{L})$ is $|\mathscr{X}'|$.*

Proof. Let $\mathscr{L}'$ be given by Eq. (6). Then we have seen that there is a one-to-one correspondence between the extreme points of $\mathscr{P}(\mathscr{L})$ and those of $\mathscr{P}(\mathscr{L}')$. On the other hand, by Remark 2 every $x_i \in \mathscr{X}'$ determines a different extreme point of $\mathscr{P}(\mathscr{L}')$ by means of the top subnetwork it generates in the graph $\mathscr{G}'$. Thus, there are at least $|\mathscr{X}'|$ extreme points for any $\mathscr{L}$.   □

Two typical examples where this minimal number of extreme points is reached is the case where $\mathscr{L} = \emptyset$, in which case the extreme points are all Dirac measures, and the case where $\mathscr{L}$ forms a complete ordering of singletons $\{x_1, \ldots, x_n\}$ (treated by Kofler [1989, P. 26] and Walley [1991, P. 195]).

Remark 4. In particular it follows from our results that if we consider a comparative probability ordering between the singletons, the only case in which the set $\mathscr{P}(\mathscr{L})$ has only one element is when

$$\{(1,2),(2,3),\ldots,(n-1,n),(n,1)\} \subseteq \mathscr{C}(\mathscr{L}),$$

that is, when all the elements $x_1, \ldots, x_n$ of $\mathscr{X}$ are believed to be equally likely. In that case, the only agreeing probability measure is the uniform one, given by $P(x_1) = P(x_2) = \cdots = P(x_n) = \frac{1}{n}$.

In any other case, there is not a unique probability compatible with the assessments; in other words, a comparative probability model on the singletons is not *complete*, in the sense of de Finetti [1931] and Scott [1964], unless the associated graph has a cycle involving all the nodes. ♦

4.2. **Extraction Algorithm.** Using the results of Section 3, we can propose a pseudo-algorithm to extract extreme points from a weakly-connected acyclic graph $\mathscr{G}$, summarised in Algorithm 1. It uses the fact that non-decomposable top subnetworks are in correspondence with sets of strongly disconnected nodes, as we have explained in Remark 3. Graphs with cycles or that are not weakly connected can be handled by considering equivalence classes or weakly-connected sub-graphs, as explained in Section 3 (Propositions 2 and 3). Finding such weakly connected components is a well-known and easily solvable problem [Hopcroft and Tarjan , 1973].

Algorithm 1: Extreme point search

Input: Set $\mathscr{L}$ of comparisons inducing a weakly-connected acyclic graph $\mathscr{G}$

Output: Extreme points of $\mathscr{P}(\mathscr{L})$

1   List $\leftarrow \emptyset$;

2   **for** $i = 1, \ldots, n$ **do**

3   Build extreme points corresponding to $H(x_i)$;

4   List $\leftarrow \{x_i\}$;

5   Candidate set $\leftarrow$ List ;

6   **for** $i = 2, \ldots, n$ **do**

7   List $\leftarrow \emptyset$;

8   **foreach** *set B in Candidate set* **do**

9   **for** $i = 1, \ldots, n$ **do**

10   **if** x_i *is strongly disconnected from the elements B and* $H(B \cup \{x_i\})$ *corresponds to a new extreme point* **then**

11   Add $P_{H(B \cup \{x_i\})}$ to extreme points ;

12   List $\leftarrow B \cup \{x_i\}$;

13   Candidate set $\leftarrow$ List ;

Implementing Algorithm 1 mainly requires being able, for a given set B, to check whether elements of B are strongly disconnected and to compute $H(B)$. An instrumental tool to do this is the matrix M corresponding to the

transitive closure $\mathscr{C}(\mathscr{L}) \subseteq \{1,\ldots,n\} \times \{1,\ldots,n\}$ of $\mathscr{L}$, with $M(i,j) = 1$ iff $(i,j) \in \mathscr{C}(\mathscr{L})$. M can be efficiently computed by applying Warshall's algorithm [Warshall , 1962] to the matrix L given by $L(i,j) = 1$ iff $(i,j) \in \mathscr{L}$.

Once this is done, checking whether two elements x_i, x_j are strongly disconnected can be done in linear time. Checking that B is made of strongly disconnected elements is equivalent to check whether all pairs of elements $x_i, x_j \in B$ are strongly disconnected, hence at most in quadratic time. As $H(B) = \cup_{x \in B} H(x)$, computing $H(B)$ is also linear. This means that the complexity of the loop going from Line 10 to 13 in Algorithm 1 is quadratic.

Algorithm 1 also tries to minimize the number of sets of nodes to check by reducing the search to top subnetworks generated by sets of strongly disconnected nodes, rather than making a naive search among all subsets $B \subseteq \mathscr{X}$. However, the algorithm would still have to check, at worst, an exponential number of sets. An interesting study would be to compare Algorithm 1, which takes advantage of the particular structure of the model, with classical vertex enumeration techniques [Avis and Fukuda , 1992] also known to be polynomial for a fixed number of vertices.

4.3. n-monotonicity.

Next, we investigate in more detail the set of probabilities $\mathscr{P}(\mathscr{L})$ from the point of view of the theory of coherent lower previsions developed by Walley [1991]. Since the set $\mathscr{P}(\mathscr{L})$ is a closed convex set of probabilities, its lower envelope $\underline{P}$ on events, given by

(8) $$\underline{P}(A) = \min\{P(A) : P \in \mathscr{P}(\mathscr{L})\} \; \forall A \subseteq \mathscr{X}$$

is a *coherent lower probability*. As such, it can be given a behavioural interpretation in terms of acceptable betting rates.

Coherent lower probabilities include as particular cases most of the imprecise probability models that we can find in the literature, such as 2-monotone capacities, belief functions, or necessity measures; see Walley [2000] for more details. In particular, a coherent lower probability is 2-monotone when for any $A, B \subseteq \mathscr{X}$ we have

(9) $$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B).$$

It is called *n*-monotone when for any $A_1, \ldots, A_n \subseteq \mathscr{X}$ we have that

$$\underline{P}(\cup_{i=1}^n A_i) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \underline{P}(\cap_{i \in I} A_i),$$

and if it is *n*-monotone for every *n*, then it is called a *belief function* [Shafer , 1976], and it is determined by a *Möbius inverse* $m : 2^{\mathscr{X}} \to [0,1]$ by means

of the formula

$$(10) \qquad \underline{P}(A) = \sum_{B \subseteq A} m(B) \ \forall A \subseteq \mathscr{X}.$$

The mapping m is such that $m(\emptyset) = 0$ and $\sum_{E \subseteq \mathscr{X}} m(E) = 1$, and it can be seen as a probability mass defined over subsets. The sets B for which $m(B) > 0$ are called the *focal elements* of the belief function $\underline{P}$; when they are nested, the belief function is minimum-preserving (i.e., it satisfies $\underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}$ for every pair of subsets A, B), and it is called a *necessity measure*.

Belief functions have been shown of interest in the context of evidential reasoning, and also in connection with random sets [Dempster , 1967]. On the other hand, 2-monotone capacities are also called *convex* functions or Choquet capacities of order 2 [Choquet , 1953, Denneberg , 1994]. When $|\mathscr{X}| \leq 3$, it was showed by Walley [1981] that a coherent lower probability on $2^{\mathscr{X}}$ is always 2-monotone but not necessarily 3-monotone.

Interestingly, the particular coherent lower probabilities induced by a comparative probability model on the singletons are always belief functions (i.e., *n*-monotone for every *n*) when $|\mathscr{X}| = 3$:

Proposition 6. *If $|\mathscr{X}| = 3$, then the lower probability induced by a comparative probability ordering on the singletons by means of* (8) *is a belief function.*

Proof. First of all, if the graph $\mathscr{G}$ associated with $\mathscr{L}$ is weakly connected and acyclic, there are three possibilities (except for permutations of the states):

- $\mathscr{L} = \{(1,2), (2,3)\}$. From Theorem 1, we deduce that the extreme points of $\mathscr{P}(\mathscr{L})$ are determined by the set of probability masses $\{(1,0,0), (0.5, 0.5, 0), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\}$. Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1\}) = \frac{1}{3} = m(\{x_1, x_2\}), \ m(\{x_1, x_3\}) = \frac{1}{6} = m(\{x_1, x_2, x_3\}).$$

- $\mathscr{L} = \{(1,2), (1,3)\}$. Again by Theorem 1, the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\left\{ (1,0,0), (0.5, 0.5, 0), (0.5, 0, 0.5), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \right\}.$$

  Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1\}) = \frac{1}{3} = m(\{x_1, x_2, x_3\}), \ m(\{x_1, x_2\}) = \frac{1}{6} = m(\{x_1, x_3\}).$$

- $\mathscr{L} = \{(1,3),(2,3)\}$. By Theorem 1, the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\left\{ (1,0,0),(0,1,0),(\frac{1}{3},\frac{1}{3},\frac{1}{3}) \right\}.$$

Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1,x_2\}) = \frac{2}{3}, \ m(\{x_1,x_2,x_3\}) = \frac{1}{3}.$$

On the other hand, if the graph $\mathscr{G}$ is acyclic but is not connected, there are two possible scenarios (again, except for permutations of the states):

- $\mathscr{L} = \emptyset$. Then the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\{(1,0,0),(0,1,0),(0,0,1)\}.$$

The lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1,x_2,x_3\}) = 1.$$

- $\mathscr{L} = \{(1,2)\}$. Then the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\{(1,0,0),(0.5,0.5,0),(0,0,1)\}.$$

Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1,x_3\}) = 0.5, m(\{x_1,x_2,x_3\}) = 0.5.$$

Finally, when the graph has cycles, there can be four possible cases:

- $\mathscr{L} = \{(1,2),(2,3),(3,1)\}$. Then there is only one compatible probability measure (that is in particular a belief function): the one determined by $P(x_1) = P(x_2) = P(x_3) = \frac{1}{3}$.
- $\mathscr{L} = \{(1,2),(2,1)\}$. In that case the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\{(0.5,0.5,0),(0,0,1)\}.$$

Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1,x_3\}) = m(\{x_2,x_3\}) = 0.5.$$

- $\mathscr{L} = \{(1,2),(2,1),(2,3)\}$. Then the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\left\{ (0.5,0.5,0),(\frac{1}{3},\frac{1}{3},\frac{1}{3}) \right\}.$$

Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_1\}) = m(\{x_2\}) = \frac{1}{3}, m(\{x_1,x_3\}) = m(\{x_2,x_3\}) = \frac{1}{6}.$$

- $\mathscr{L} = \{(1,2),(2,1),(3,2)\}$. Then the extreme points of $\mathscr{P}(\mathscr{L})$ are associated with the probability masses

$$\left\{ (0,0,1), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \right\}.$$

Then the lower probability associated with this credal set is the belief function with Möbius inverse

$$m(\{x_3\}) = \frac{1}{3}, m(\{x_1,x_3\}) = m(\{x_2,x_3\}) = \frac{1}{3}. \qquad \square$$

Note that the lower probability determined by (8) need not be a necessity measure: as we can see in the proof of Proposition 6, in not all cases the focal elements of the belief function are nested.

Next we show that Proposition 6 does not extend to the case when $|\mathscr{X}| \geq 4$, because the coherent lower probabilities induced by comparative probability models need not even be 2-monotone.

Example 6. Consider the space $\mathscr{X} = \{x_1,x_2,x_3,x_4\}$ and the assessments $\mathscr{L}' = \{(1,2),(1,3),(2,4),(3,4)\}$ represented by Figure 4.
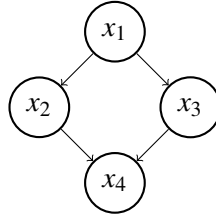

FIGURE 4. Graph $\mathscr{G}$ of Example 6

From Theorem 1, the extreme points of $\mathscr{P}(\mathscr{L})$ are the probability measures determined by the probability masses

$$\left\{ (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}), (\frac{1}{2}, \frac{1}{2}, 0, 0), (\frac{1}{2}, 0, \frac{1}{2}, 0), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0), (1,0,0,0) \right\};$$

as a consequence, if we consider the events $A = \{x_1,x_3\}$ and $B = \{x_1,x_4\}$, we see that

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) = 1/2 + 1/4 < \underline{P}(A) + \underline{P}(B) = 1/2 + 1/3.$$

Hence, $\underline{P}$ violates the 2-monotonicity condition. ♦

This implies that belief functions, which are in particular 2-monotone, are not expressive enough to represent comparative probability models.

On the other hand, from a convex set of probability measures we can also determine lower and upper expectation functionals. Similarly to Eq. (8), the real-valued functional $\underline{P}$ given by

(11)
$$\underline{P}(f) = \min\{P(f) : P \in \mathscr{P}(\mathscr{L})\}$$

for any function $f : \mathscr{X} \to \mathbb{R}$ is called a *coherent lower prevision*. Here, we are also using P to denote the expectation functional associated with the probability measure P, given by $P(f) = \sum_{x \in \mathscr{X}} f(x)p(x)$.

Coherent lower previsions can also be given a behavioural interpretation, in terms of acceptable buying prices. Within that context, the coherent lower prevision $\underline{P}$ defined above corresponds to the *natural extension* of the assessment that the functions $\{I_{x_i} - I_{x_j} : (i,j) \in \mathscr{L}\}$ are *almost-desirable*; see Walley [1991, Section 3.7] for more details on the correspondence with desirability.

Interestingly, although a coherent lower prevision always determines a unique coherent lower probability (its restriction to events), a coherent lower probability may have more than one extension to a coherent lower prevision, and in this sense the latter can be seen as a more expressive model. This is the reason why lower previsions are often preferred to lower probabilities. Next, we investigate in more detail the properties of the coherent lower previsions induced by a comparative probability model.

Similarly to Eq. (9), a coherent lower prevision is called 2-monotone when for any $f, g : \mathscr{X} \to \mathbb{R}$,

$$\underline{P}(f \vee g) + \underline{P}(f \wedge g) \geq \underline{P}(f) + \underline{P}(g),$$

where $\vee$ denotes the point-wise maximum and $\wedge$ denotes the point-wise minimum. 2-monotone lower previsions have been studied in detail by de Cooman et al. [2008] and Walley [1981]. In particular, it was established by de Cooman et al. [2008, Theorem 15] that the property of 2-monotonicity is equivalent to *comonotone additivity*, which means that

$$\underline{P}(f + g) = \underline{P}(f) + \underline{P}(g) \ \forall f, g \text{ comonotone},$$

and where f and g are said to be comonotone when for every $x_i, x_j \in \mathscr{X}$ it holds that

$$(f(x_i) - f(x_j))(g(x_i) - g(x_j)) \geq 0.$$

2-monotone lower previsions are interesting, because, unlike general coherent lower previsions, they can be calculated as the Choquet integral with respect to the lower probability that is their restriction to events (hence easing computations). However, we can prove that if $|\mathscr{X}| \geq 3$ the coherent lower

prevision associated with a comparative probability is not 2-monotone except for trivial cases:

Theorem 7. *Consider a space $\mathscr{X}$ with $|\mathscr{X}| \geq 3$, and let $\mathscr{L}$ be a number of probability comparisons on the elements of $\mathscr{X}$. Let $\underline{P}$ be the coherent lower prevision determined by* (11). *Then:*

$\underline{P}$ *2-monotone $\iff$ either (i)$\mathscr{L} = \emptyset$ or (ii) $(i,j) \in \mathscr{C}(\mathscr{L}) \; \forall i, j \in \{1, \ldots, n\}$.*

Proof. First of all, if $\mathscr{L} = \emptyset$, then the lower prevision it induces is the vacuous one, given by $\underline{P}(f) = \min f$ for every gamble f, which is 2-monotone. On the other hand, if $(i,j) \in \mathscr{C}(\mathscr{L})$ for every $i, j \in \{1, \ldots, n\}$ then the only element of $\mathscr{P}(\mathscr{L})$ is the uniform probability distribution, associated with the probability mass $p(x_1) = p(x_2) = \cdots = p(x_n) = \frac{1}{n}$, which is also 2-monotone.

Let us show that in any other case $\underline{P}$ is not 2-monotone. Assume ex-absurdo that it is. Then by de Cooman et al. [2008] it corresponds to the Choquet integral with respect to its restriction to events, which is a 2-monotone lower probability. Let $\underline{Q}$ denote this lower probability. Given (i,j) in the non-empty set $\mathscr{L}$, it should be

$$0 \leq \underline{P}(I_{x_i} - I_{x_j}) = (C) \int I_{x_i} - I_{x_j} \mathrm{d}\underline{Q}$$
$$= 1\underline{Q}(\{x_i\}) + 1\underline{Q}(\mathscr{X} \setminus \{x_j\}) - 1 = \underline{Q}(\{x_i\}) - \overline{Q}(\{x_j\}),$$

where $\overline{Q}$ is the upper probability that is conjugate to $\underline{Q}$, and corresponds to the upper envelope of the set $\mathscr{P}(\mathscr{L})$.

Now, it follows from Theorem 1 that

$$\underline{Q}(\{x_i\}) = \min\{P_B(\{x_i\}) : B \text{ non-decomposable top subnetwork}\}.$$

This means in particular that $\underline{Q}(\{x_i\}) \in \{0, \frac{1}{j} : j \in 1, \ldots, |\mathscr{X}|\}$. As a consequence, if there is a top subnetwork B that does not include x_i, neither does any of the non-decomposable top subnetworks it includes, and we obtain that $\underline{Q}(\{x_i\}) = P_B(\{x_i\}) = 0$. In particular, this happens when $\mathscr{G}$ is not weakly connected: then given x_k in a different weakly connected component of $\mathscr{G}$ than x_i, we have that $\underline{Q}(\{x_i\}) \leq P_{H(x_k)}(\{x_i\}) = 0$.

On the other hand, if $\mathscr{G}$ is weakly connected and x_i belongs to all the non-decomposable top subnetworks, then $\underline{Q}(\{x_i\}) = \frac{1}{|B|}$, where B is the largest non-decomposable top subnetwork. This is given by $\mathscr{X}$: it is obviously a top subnetwork (generated by itself) and it is non-decomposable because we are assuming that the graph $\mathscr{G}$ is weakly connected. Thus, in that case $\underline{Q}(\{x_i\}) = P_{\mathscr{X}}(\{x_i\}) = \frac{1}{n}$.
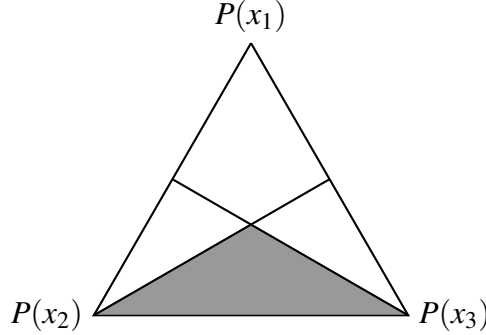
FIGURE 5. Remark 5 illustration for $\mathscr{L} = \{(3,1),(2,1)\}$

We conclude that

$$\underline{Q}(\{x_i\}) = \begin{cases} \frac{1}{n} & \text{if } (i,k) \in \mathscr{C}(\mathscr{L}) \ \forall k \in \{1,\ldots,n\} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, $\overline{Q}(\{x_j\}) = P_{H(x_j)}(\{x_j\}) = \frac{1}{|H(x_j)|} \geq \frac{1}{n}$, and from this we deduce that a necessary condition for the 2-monotonicity of $\underline{P}$ is that

(12) $\qquad (i,j) \in \mathscr{L} \Rightarrow (i,k),(k,j) \in \mathscr{C}(\mathscr{L}) \ \forall k \in \{1,\ldots,n\},$

taking into account that we should have $H(x_j) = \mathscr{X}$ and that x_i is a predecessor of any node in $\mathscr{G}$ in order to get $\underline{Q}(\{x_i\}) = \overline{Q}(\{x_j\}) = \frac{1}{n}$.

Let us now use this condition to show that if $\mathscr{L}$ is non-empty and $\mathscr{P}(\mathscr{L})$ has more than one element, then its associated lower prevision is not 2-monotone. Take $(i,j) \in \mathscr{L}$. We reason ex-absurdo. Assume that the associated lower prevision $\underline{P}$ is 2-monotone, and that therefore Eq. (12) is satisfied. Then it must be $\mathscr{C}(\mathscr{L}) \supseteq \{(\ell,j),(\ell,k) : \ell,k \in \{1,\ldots,n\} \setminus \{j\}\}$.

On the other hand, for any $k \neq j$ it must be $(j,k) \notin \mathscr{C}(\mathscr{L})$, since otherwise we would obtain $\mathscr{C}(\mathscr{L}) = \{(\ell,k) : \ell,k \in \{1,\ldots,n\}\}$ and $\mathscr{P}(\mathscr{L})$ would have only one element, which is ruled out by assumption. From this we can conclude that

(13) $\qquad \mathscr{C}(\mathscr{L}) = \{(\ell,j),(\ell,k) : \ell,k \in \{1,\ldots,n\} \setminus \{j\}\}.$

But this is incompatible with Eq. (12): given $k \neq i,j$, the fact that $(i,k) \in \mathscr{C}(\mathscr{L})$ implies that there is some $\ell \neq k$ such that $(\ell,k) \in \mathscr{L}$, and then if Eq. (12) held we should have $(j,k) \in \mathscr{C}(\mathscr{L})$, a contradiction with Eq. (13). We conclude that Eq. (12) cannot hold and as a consequence that $\underline{P}$ is not 2-monotone. $\qquad \square$

Remark 5. Using a geometrical approach, we can easily see that the coherent lower prevision associated with a comparative probability model is

not 2-monotone when $\mathscr{X}$ has three elements and $\mathscr{G}$ is weakly connected and acyclic[5]: $\underline{P}$ will be 2-monotone if and only if its associated credal set $\{P : P \geq \underline{P}\}$ coincides with the one determined by the restriction of $\underline{P}$ to events, i.e., if it is the subset of the 2-unit simplex determined by the restrictions $P(A) \in [\underline{P}(A), \overline{P}(A)]$ for every $A \subseteq \mathscr{X}$. Then it can be checked that any of these restrictions is parallel to one of the sides of the simplex, while the restrictions of the comparative probability model are of the type $p(x_i) - p(x_j) \geq 0$, and correspond to a bisector; and the intersections of the subsets of the simplex determined by these bisectors will not determine a credal set whose sides are parallel to the sides of the simplex, except in the two particular cases considered in Theorem 7: when the credal set is equal to $\mathscr{P}(\mathscr{L})$ (meaning that $\mathscr{L} = \emptyset$) or when $\mathscr{P}(\mathscr{L})$ has only one element (meaning that $\mathscr{C}(\mathscr{L})$ includes $\{(1,2), (2,3), (3,1)\}$). These two cases cannot hold when the graph $\mathscr{G}$ is weakly connected and acyclic. A similar reasoning may be done on higher dimensions. Figure 5 provides an illustration for the assessments $\mathscr{L} = \{(3,1), (2,1)\}$, where it is easy to see that the bisector lines delimiting of the probability set are not parallel to the triangle sides.

We refer to [Fujishige , 1991, Kroupa , 2008, Shapley , 1971] for more information on the geometric approach to credal sets. ♦

4.4. **Conditioning.** A classical operation when dealing with uncertainty is that of conditioning. Here we will study the problem of computing lower conditional probabilities $\underline{P}(A|B)$ from the credal set $\mathscr{P}(\mathscr{L})$. Out of the many possible notions we can consider in this case, we think that the most intuitive under the robust Bayesian interpretation we are considering in this paper is that of *regular extension* [Walley , 1991, Appendix J], that produces

$$(14) \qquad \underline{P}(A|B) = \inf_{P \in \mathscr{P}(\mathscr{L})} \{P(A|B) : P(B) > 0\},$$

where $P(A|B)$ is obtained from P by means of Bayes' rule. We refer to [de Campos et al. , 1990, Fagin and Halpern , 1991b] and [Jaffray , 1992] for some works on this conditioning rule.

In order to apply this definition, we need the existence of some probability measure P in $\mathscr{P}(\mathscr{L})$ such that $P(B) > 0$ (or, in other words, that the *upper* probability $\overline{P}(B)$ is positive); but this is no restriction in the case of comparative probabilities, because there will always be an extreme point P of $\mathscr{P}(\mathscr{L})$ for which $P(B) > 0$: it suffices to consider $P_{H(x_i)}$ with $x_i \in B$. On the contrary, the lower probability $\underline{P}(B)$ will be positive if and only if B intersects all top subnetworks.

---

[5]We thank Erik Quaeghebeur for pointing this out to us.

To reach the conditional lower probability $\underline{P}(A|B)$ given by Eq. (14), we need to find the extreme point for which $P(B)$ is positive and the fraction $P(A \cap B)/P(B)$ minimal. Provided $\mathscr{G}$ is acyclic (graphs with cycles can be dealt with by considering equivalence classes, as described at the end of Section 3), this can be done easily by the procedure described in Algorithm 2. Finding the value (14) comes down to finding the non-decomposable top subnetwork $D \in \mathscr{H}$ minimizing the value

$$(15) \qquad \frac{P_D(A \cap B)}{P_D(B)} = \frac{|D \cap B \cap A|}{|D \cap B|}.$$

Taking into account Remark 3, it holds that $D = H(C)$, for some set C of strongly disconnected nodes in $\mathscr{G}$. Note that it is sufficient to focus on sets C such that $C \subseteq B \setminus A$ or $C \subseteq B \cap A$:

- as $H(C) \cap B = H(H(C) \cap B) \cap B$, it is sufficient to focus on sets $C \subseteq B$;
- for any set C such that $C \cap (B \setminus A) \neq \emptyset$ and $C \cap B \cap A \neq \emptyset$, we can decrease the ratio $|H(C) \cap B \cap A|/|H(C) \cap B|$ by picking an element $x \in C \cap B \cap A$ and considering the set $C' = (C \setminus \{x\}) \cup (H(\{x\}) \setminus \{x\})$. Note that $C' \neq \emptyset$ because it includes the set $C \setminus \{x\}$, which is itself non-empty because it must have a non-empty intersection with $B \setminus A$. Taking into account that x is strongly disconnected from $C \setminus \{x\}$, we obtain that $H(C') = H(C) \setminus \{x\}$. Then since $x \in A \cap B$, both the numerator and denominator are decreased by one, meaning that the ratio (15) decreases. This means that focusing on subsets of $B \setminus A$ and of $B \cap A$ is sufficient to retrieve the solution.

Algorithm 2: Conditional Probability computation

Input: Set $\mathscr{L}$ of comparisons
Output: Lower conditional probability $\underline{P}(A|B)$ with $A \subset B$
1 Cond $\leftarrow 1$;
2 **foreach** *Non-empty set $C \subseteq B \setminus A$ or $C \subseteq A \cap B$* **do**
3 $\quad$ Value $\leftarrow |H(C) \cap (A \cap B)|/|H(C) \cap B|$;
4 $\quad$ **if** *Value $<$ Cond* **then** Cond$\leftarrow$ Value ;
5 Return Cond ;

Next, we will characterize the conditional set

$$(16) \qquad \mathscr{P}_B(\mathscr{L}) = \{P(\cdot|B) : P \in \mathscr{P}(\mathscr{L}), P(B) > 0\}$$

obtained by the regular extension. Any of the elements of the set $\mathscr{P}_B(\mathscr{L})$ can equivalently be seen as a probability measure on $\mathscr{P}(B)$. Using this

correspondence, we shall proceed to show that $\mathscr{P}_B(\mathscr{L})$ is again induced by a comparative probability.

Given a set $\mathscr{L}$ of comparisons, we will denote by $\mathscr{C}(\mathscr{G})$ the graph associated with its transitive closure $\mathscr{C}(\mathscr{L})$. We will denote by $\mathscr{L}_B \subseteq B \times B$ the relation obtained from $\mathscr{C}(\mathscr{L})$ by

$$(17) \qquad \mathscr{L}_B = \{(i,j) : (i,j) \in \mathscr{C}(\mathscr{L}), x_i, x_j \in B\},$$

that is, $\mathscr{L}_B$ is the relation $\mathscr{C}(\mathscr{L})$ restricted to elements in B. We will denote by $\mathscr{G}_B$ the graph corresponding to $\mathscr{L}_B$. Figure 6 shows the graph obtained by transitive closure for Example 1, as well as $\mathscr{G}_{\{x_1,x_2,x_5\}}$.
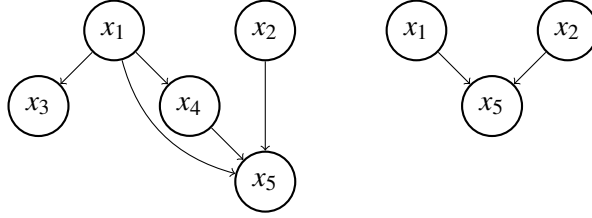

FIGURE 6. Graphs $\mathscr{C}(\mathscr{G})$ and $\mathscr{G}_{\{x_1,x_2,x_5\}}$ obtained from Example 1

Theorem 8. *Consider a space $\mathscr{X}$ and let $\mathscr{L}$ be a set of probability comparisons. Consider a conditioning event B, and let $\mathscr{L}_B$ be given by Eq. (17). Then*

$$\mathscr{P}_B(\mathscr{L}) = \mathscr{P}(\mathscr{L}_B),$$

where $\mathscr{P}(\mathscr{L}_B)$ is the credal set associated with $\mathscr{L}_B$ by (1) and $\mathscr{P}_B(\mathscr{L})$ is given by Eq. (16).

Proof. Let us show the double inclusion between these two sets.

On the one hand, given $x_i, x_j \in B$ such that $(i,j) \in \mathscr{C}(\mathscr{L})$, it follows from transitivity that $P(x_i) \geq P(x_j)$ for every $P \in \mathscr{P}(\mathscr{L})$. In particular, if $P \in \mathscr{P}(\mathscr{L})$ satisfies $P(B) > 0$ then $P(\cdot|B)$ will satisfy

$$P(\{x_i\}|B) = \frac{P(\{x_i\})}{P(B)} \geq \frac{P(\{x_j\})}{P(B)} = P(\{x_j\}|B) \ \forall x_i, x_j \in B \text{ s.t. } (i,j) \in \mathscr{C}(\mathscr{L}),$$

and since $P(B|B) = 1$ we deduce that $P(\cdot|B)$ belongs to $\mathscr{P}(\mathscr{L}_B)$. Thus, $\mathscr{P}_B(\mathscr{L}) \subseteq \mathscr{P}(\mathscr{L}_B)$.

Conversely, consider $P \in \mathscr{P}(\mathscr{L}_B)$ and let $p' : \mathscr{X} \to [0,1]$ be given by

$$p'(x_i) := \begin{cases} \max\{P(\{x_j\}) : x_j \in B, (i,j) \in \mathscr{C}(\mathscr{L})\} & \text{if } x_i \in H(B) \\ 0 & \text{otherwise.} \end{cases}$$

This function satisfies by construction that $p'(x_i) \geq p'(x_j)$ for every $(i,j) \in \mathscr{L}$. If we now consider the probability measure Q given by

$$Q(C) = \frac{\sum_{x_i \in C} p'(x_i)}{\sum_{x_j \in \mathscr{X}} p'(x_j)} \quad \forall C \subseteq \mathscr{X},$$

it follows from the reasoning above that $Q \in \mathscr{P}(\mathscr{L})$. Moreover, by construction we have that $Q(\cdot|B) = P$. Thus, $\mathscr{P}(\mathscr{L}_B) \subseteq \mathscr{P}_B(\mathscr{L})$ and from this we deduce that the two sets are equal. $\qquad \square$

Theorem 8 tells us that computing the conditional lower prevision obtained through regular extension is easy for the specific model studied here.

4.5. **Multiple source merging.** When multiple sources provide different comparisons, for instance when two different experts provide assessments $\mathscr{L}_1$ and $\mathscr{L}_2$, it becomes necessary to merge them in a single representation. The two most common ways are the conjunction and disjunction rules, which respectively come down to computing $\mathscr{P}(\mathscr{L}_1) \cap \mathscr{P}(\mathscr{L}_2)$ and $CH(\mathscr{P}(\mathscr{L}_1) \cup \mathscr{P}(\mathscr{L}_2))$ where CH denotes the convex hull (note that the disjunction usually produces non-convex probability sets). Our next two results demonstrate that simple operations on $\mathscr{L}_1$ and $\mathscr{L}_2$ can provide exact or approximate results of these operations.

Proposition 9. *The conjunctively merged set $\mathscr{P}(\mathscr{L}_1) \cap \mathscr{P}(\mathscr{L}_2)$ is such that*

$$\mathscr{P}(\mathscr{L}_1) \cap \mathscr{P}(\mathscr{L}_2) = \mathscr{P}(\mathscr{L}_1 \cup \mathscr{L}_2)$$

with $\mathscr{L}_1 \cup \mathscr{L}_2 = \{(i,j) : (i,j) \in \mathscr{L}_1 \text{ or } (i,j) \in \mathscr{L}_2\}$

Proof. It is sufficient to note that $\mathscr{L}_1 \cup \mathscr{L}_2$ is the set of all constraints inducing $\mathscr{P}(\mathscr{L}_1)$ and $\mathscr{P}(\mathscr{L}_2)$. $\qquad \square$

For the disjunction, the same strategy only gives us an approximation.

Proposition 10. *The disjunctively merged set $CH(\mathscr{P}(\mathscr{L}_1) \cup \mathscr{P}(\mathscr{L}_2))$ is such that*

$$(18) \qquad CH(\mathscr{P}(\mathscr{L}_1) \cup \mathscr{P}(\mathscr{L}_2)) \subseteq \mathscr{P}(\mathscr{C}(\mathscr{L}_1) \cap \mathscr{C}(\mathscr{L}_2)).$$

Proof. It suffices to note that $\mathscr{P}(\mathscr{L}_i) = \mathscr{P}(\mathscr{C}(\mathscr{L}_i)) \subseteq \mathscr{P}(\mathscr{C}(\mathscr{L}_1) \cap \mathscr{C}(\mathscr{L}_2))$ for $i = 1, 2$. $\qquad \square$

The inclusion in Eq. (18) can be strict, as our next example shows. In fact, $CH(\mathscr{P}(\mathscr{L}_1) \cup \mathscr{P}(\mathscr{L}_2))$ will usually not be representable by means of comparative probabilities.

Example 7. Consider the possibility space $\mathscr{X} = \{x_1, x_2, x_3\}$ and the assessments $\mathscr{L}_1 = \{(1,2),(2,3)\}$ and $\mathscr{L}_2 = \{(3,2),(2,1)\}$. We have $\mathscr{C}(\mathscr{L}_1) \cap \mathscr{C}(\mathscr{L}_2) = \emptyset$, hence the induced probability set $\mathscr{P}(\mathscr{C}(\mathscr{L}_1) \cap \mathscr{C}(\mathscr{L}_2))$ is the

set of all possible probability masses, while $CH(\mathscr{P}(\mathscr{L}_1) \cup \mathscr{P}(\mathscr{L}_2))$ does not include the Dirac measure $P(\{x_2\}) = 1$ (that is neither an extreme point of $\mathscr{P}(\mathscr{L}_1)$ nor of $\mathscr{P}(\mathscr{L}_2)$). ♦

It is interesting to note that, from Zaffalon and Miranda [2003, Theorem 6] the union of two credal sets $\mathscr{P}(\mathscr{L}_1)$ and $\mathscr{P}(\mathscr{L}_2)$ is again a convex set if and only if the lower envelope of the intersection $\mathscr{P}(\mathscr{L}_1) \cap \mathscr{P}(\mathscr{L}_2)$ equals the maximum of the lower envelopes of $\mathscr{P}(\mathscr{L}_1)$ and $\mathscr{P}(\mathscr{L}_2)$.

5. EXAMPLES AND EXTENSIONS

In this section, we propose some particular examples of situations where singleton comparisons can be used, and discuss some possible extensions.

5.1. **Modal and least probable values.** An immediate example is the case where we know only the modal value of the underlying probability distribution. In this case, if x_i is the modal value, the set of one-to-one comparisons is given by $\mathscr{L} = \{(i,j) : j \in \{1,\dots,n\}, j \neq i\}$ and the corresponding graph is illustrated by Figure 3. Note that if an expert provides a set M of most probable values, the corresponding comparative probability model would be given by $\mathscr{L} = \{(i,j) : x_i \in M, x_j \in \mathscr{X} \setminus M\}$.

Another example is when a least probable value is given, for example x_i, in which case $\mathscr{L} = \{(j,i) : j \in \{1,\dots,n\}, j \neq i\}$ and the number of extreme points is $|\mathscr{X}|$ (the uniform generated by $H(x_i)$ and the Dirac measure on other points). See Figure 7 for an illustration. In the case where the known least probable singletons belong to a set M, the corresponding model is $\mathscr{L} = \{(j,i) : x_i \in M, x_j \in \mathscr{X} \setminus M\}$.
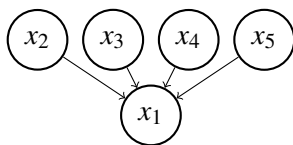

FIGURE 7. Graph $\mathscr{G}$ for $x_1 =$ least probable value

5.2. **Imprecise mass functions.** Comparative probability models can also be related to the work on imprecise mass functions discussed for example by Augustin [2005] and Denoeux [1999].

Recall that a belief function $\underline{P}$ on the power set of $\mathscr{X}$ is uniquely determined by its associated *Möbius inverse m*, by means of Eq. (10). The Möbius inverse $m(A)$ of a set A represents the weight of the available evidence supporting that the outcome of the experiment belongs to E. It holds

that $\sum_{A \subseteq \mathscr{X}} m(A) = 1$, so we may regard m as the probability mass of some probability measure on $2^{2^{\mathscr{X}}}$.

We can then use our results to build imprecise probability masses. If we have assessments of the type $m(A_i) \geq m(A_j)$, we may consider the set of functions compatible with these assessments. This is a convex set (bearing on subsets of $\mathscr{X}$) whose extreme points can be determined by means of Theorem 1. By means of Eq. (10), each of these extreme points is the Möbius inverse of some belief function, that in turn is equivalent to a convex set of probability measures on $2^{\mathscr{X}}$. Note that this convex set of masses also induces a convex set of probabilities. To see that, simply note that a convex set of convex sets is convex, and that each Möbius inverse corresponds to a convex set of probabilities (we refer to Augustin [2005] for further details about convex sets of Möbius inverses).

Let us illustrate this idea with two examples.

Inner/outer measures. [Fagin and Halpern , 1991a] The situation above can arise when $\mathscr{X}$ is partitioned into n sets $A_1, \ldots, A_n$ and where a probability value $P(A_i)$ is associated with each set A_i. Practical cases where this partitioning may occur include continuous variables discretized in (multivariate) histograms, or cases where some grouping of cases makes sense.

The assessments $P(A_i), i = 1, \ldots, n$ induce a set of probabilities (whose lower and upper envelopes are called inner/outer measures) on the power set of $\mathscr{X}$ that can be described by $m(A_i) = P(A_i)$: the set of probabilities that dominate the belief function determined by the Möbius inverse m on the power set of $\{A_1, \ldots, A_n\}$. Indeed, it can be checked that the lower and upper envelopes of this set correspond to the inner and outer measures determined by the belief function.

In this framework, making comparative statements between the probabilities $P(A_i)$ of the elements of the discretized space comes down to making comparative statements between the masses $m(A_i)$, and the set of extreme masses can then be derived using our results.

Example 8. Let $\mathscr{X} = \{x_1, \ldots, x_5\}$ with the partition $A_1 = \{x_1, x_5\}$, $A_2 = \{x_2\}$ and $A_3 = \{x_3, x_4\}$. Then the statements $m(A_3) \geq m(A_2)$, $m(A_3) \geq m(A_1)$ induce the set $\mathscr{M}$ with the following extreme points

	m		
B	A_1	A_2	A_3
$\{A_3\}$	0	0	1
$\{A_1, A_3\}$	1/2	0	1/2
$\{A_2, A_3\}$	0	1/2	1/2
$\{A_1, A_2, A_3\}$	1/3	1/3	1/3

Note that, in the case where the partition counts n elements and is such that $A_i = \{x_i\}$, we retrieve the comparative probabilities studied in the previous sections. The results of Section 4.3 and Example 6 then show that, in general, a convex set of Möbius inverses will not induce a belief function (because it will not induce in general a 2-monotone set function, as Example 6 shows, and a belief function is n-monotone for every n).

Possibility measures. A second example is the case of possibility measures. A *possibility measure* is the upper probability that is conjugate to a necessity measure. It is thus induced by a Möbius inverse that gives positive weights only to a collection of nested sets. Without loss of generality, we can consider that these sets are $A_i = \{x_1, \ldots, x_i\}$ and that $m(A) > 0$ only if $A = A_i$ for some i.

Such a model may be derived by ordering elements from the most plausible (x_1) to the least plausible (x_n) and then constructing sets A_i. However, in practice, specifying possibility values for each element (which is equivalent to specifying the values $m(A_i)$) can be difficult. To avoid this issue, we can use the previous results to get a set $\mathscr{M}$ of Möbius inverses. In this specific case, the obtained model can still be represented by a possibility measure, as the supremum of possibility measures is still a possibility measure (note that possibility measures correspond to *upper* envelopes of a set of probability measures, instead of the lower envelopes we have considered throughout).

Imagine for instance that an expert is asked to provide such a ranking of plausible values, and is then asked about his own confidence about this ranking. If the expert is quite confident, it seems fair to conclude that he would give more weight to the precise information $m(A_1)$ than to more imprecise ones. In practice, we can translate this information by $m(A_1) \geq \ldots \geq m(A_n)$, from which it is easy to extract extreme points. Sets of possibility measures have also been recently investigated from a qualitative perspective by Dubois et al. [2013].

5.3. **Numerical modelling of partial preferences.** A problem where our results may prove useful is the numerical handling of partial preferences; these arise often in multi-criteria decision making [Boutilier et al. , 2004, Labreuche , 2010] or in preference learning [Fürnkranz and Hüllermeier , 2010, Lu and Boutilier , 2011].

Consider a set $\mathscr{X} = \{x_1, \ldots, x_n\}$ of n different items, and a binary relation $\mathscr{L}$ over them defining a preorder that can be represented through a directed graph $\mathscr{G}$. Our results allow to provide a numerical representation of such a preorder as a set of extreme points, or in terms of a set of complete pre-orders compatible with the preorder $\mathscr{L}$: each element within $\mathscr{P}(\mathscr{L})$ provides a pre-order over the elements of $\mathscr{X}$ that is compatible with $\mathscr{L}$.

This also provides us with a straightforward method to sample linear extensions of $\mathscr{L}$ (that is, linear orders compatible with $\mathscr{L}$): if $p_1, \ldots, p_E$ are the extreme points of $\mathscr{P}(\mathscr{L})$, a sampling of E positive weights $w_1, \ldots, w_E$ summing up to one (by sampling a uniform Dirichlet distribution, for instance) provides a mass $p = \sum_{i=1}^{E} w_i p_i$ corresponding to a linear extension.

This view of our results can also be linked to the d-permutaedron, that is the closed convex hull of all d-dimensional permutation vectors, as our results allow to characterize the set of complete preorders consistent with $\mathscr{L}$ by means of extreme points in the unit simplex. Exploring such links would bring us closer to results related to binary choices or random utilities [Zhang , 2004], the latter concerning probabilities over ranks, while our results focus on ranks over singleton probabilities.

5.4. **Extension to general comparative probability models: some comments.** The most important extension of our work would be to consider arbitrary comparative probability models, where we allow for comparisons between any pair of events (the case of partitions is treated in Section 5.2), that is to allow any comparison $P(A) \geq P(B)$ with $A, B \subseteq \mathscr{X}$. These are the models studied extensively by Koopman [1940a], Suppes [1974] and Walley and Fine [1979], amongst others. Most of the axiomatic work is based on de Finetti [1931].

When considering comparative probability models, we can assume that the sets A, B we compare are disjoint, since the assessments $P(A) \geq P(B)$ and $P(A \setminus B) \geq P(B \setminus A)$ are equivalent by additivity of P. However, the existence of a probability compatible with the assessments is no longer trivial, and therefore the associated set $\mathscr{P}(\mathscr{L})$ may be empty [Kraft et al. , 1959, Section 4]. Using the terminology from Kaplan [1977], this means that it does not necessarily hold that the comparative ordering is *additive*.[6]

When $\mathscr{P}(\mathscr{L})$ is non-empty, then it is a closed convex set that is characterized by its finite number of extreme points. However, as the next example shows, we cannot expect the extreme points of such assessments to be as simple as the extreme points generated by the comparison of the probabilities of the states. In particular, the extreme points of the associated credal sets will not be necessarily related to uniform probability distributions over some subsets, and finding an easy graphical representation from which they could be extracted seems hard.

Example 9. Consider the space $\mathscr{X} = \{x_1, x_2, x_3\}$ and the following assessments:

- $P(\{x_2\}) \geq P(\{x_1\})$

---

[6]The notion of additivity is referred to as the existence of an *almost agreeing probability* by Fishburn [1986].

• $P(\{x_1,x_2\}) \geq P(\{x_3\})$ (or, equivalently, $P(\{x_3\}) \leq 1/2$).

And let $\mathscr{P}$ be the credal set of the probabilities satisfying these assessments. The extreme points of $\mathscr{P}$ are summarised in Table 4.

	p	
x_1	x_2	x_3
0	1	0
$1/2$	$1/2$	0
$1/4$	$1/4$	$1/2$
0	$1/2$	$1/2$

TABLE 4. Extreme points of Example 9. ♦

The set of probabilities $\mathscr{P}(\mathscr{L})$ is uniquely determined by the lower prevision $\underline{P}$ it generates by means of Eq. (11). However, Theorem 7 implies that this lower prevision derived from $\mathscr{P}(\mathscr{L})$ will not be 2-monotone in general, and as consequence it will not correspond to the Choquet integral with respect to its restriction to events. In other words, the lower probability that we can derive by means of the formula

$$\underline{Q}(A) := \min\{P(A) : P \in \mathscr{P}(\mathscr{L})\}$$

will only allow us to provide in general an outer approximation of the set $\mathscr{P}(\mathscr{L})$: there are probability measures that do not belong to $\mathscr{P}(\mathscr{L})$ but dominate nonetheless $\underline{Q}$.

Another (quite rough) outer approximation is to consider the set of comparative probability assessments on the singletons that are implied by $\mathscr{L}$, and to characterise the associated credal set by means of our results in this paper.

6. CONCLUSIONS

Comparative probabilities are a useful approach to modelling uncertain information about a probability measure, especially when the available information is of a qualitative nature. However, most works in the literature about these models have focused on axiomatising those comparative probability models that can be associated with a set of probability measures. In this paper, we have focused on the particular case where the comparisons are established between the probabilities of the singletons, and we have deepened the link between comparative probability models and imprecise probabilities, by: (a) characterising the structure of the set of probability measures associated with a comparative probability model, and (b) studying the properties of the lower probability induced by this set. We have

shown that this lower probability may not be 2-monotone, from which it follows that 2-monotone capacities (and in particular belief functions, or possibility measures) are not expressive enough to be able to deal with this type of qualitative information.

We have characterised the extreme points of the credal set associated with a comparative probability model, and shown that the number of extreme points can be as high as two to the power of $|\mathcal{X}| - 1$. This is comparable to the maximal number of extreme points of credal sets induced by possibility measures, and smaller than those induced by 2-monotone capacities or belief functions.

We have also suggested some practical situations where comparative probabilities can be useful (elicitation of modal or least probable values, imprecise probability masses, . . .). However, this model remains quite simple and of limited expressiveness, and it would be desirable to study to which extent the results presented in this paper can help to solve the same problem applied to general comparisons between disjoint events (illustrated in Section 5.4). Some useful work in this sense may be done by means of Walley's work on *n-coherence* [Walley , 1991, Appendix B]. In particular, it would be interesting to see if our results help to characterise weighted comparative models on the singletons (that is, assessments of the type "the probability of x_i is at least twice that of x_j", for instance), or *probability ratios* [Walley , 1991, Section 4.6.2],[Piatti et al. , 2010]. Although it is not difficult to see that in that case the extreme points may not correspond to uniform distributions and that the set of agreeing probabilities may even be empty, we think that weighted graphs, and the arguments provided at the end of Section 3, may prove useful here.

Another important open problem would be to provide algorithms for the computation of the lower prevision induced by a comparative probability model, and to study in detail the applications of these results in fields such as qualitative decision making.

One interesting issue is to characterise the differences between strict preferences, i.e., $p(x_i) > p(x_j)$ over non-strict ones in terms of uncertainty representation. This would require a richer language than lower previsions, such as for instance sets of desirable gambles [Walley , 1991, Section 3.7], since, as we have mentioned in Section 2, both strict and non-strict preferences determine the same lower prevision when $\mathcal{G}$ is acyclic. In this sense, it would also be interesting to investigate the connections between comparative probability models and the choice functions considered in Seidenfeld et al. [2010], which seem better suited to deal with open sets of probabilities.

Finally, a more practical perspective is to run numerical experiments to check whether Algorithm 1 is competitive when compared to classical algorithms used to extract extreme points of polytopes. This could be done with

the Improb Python Library (`https://pypi.python.org/pypi/improb/`) or directly with the lrs package (`http://cgm.cs.mcgill.ca/~avis/C/lrslib/lrslib.html`).

ACKNOWLEDGEMENTS

REFERENCES

Abellán, J., & Moral, S. (2005). Upper entropy of credal sets. Applications to credal classification. *International Journal of Approximate Reasoning, 39*, 235–255.

Augustin, T. (2005). Generalized basic probability assignments. *International Journal of General Systems, 4*, 451–463.

Avis, D., & Fukuda, K. (1992). A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry, 8*, 295–313.

Berger, J. (1994). An overview of robust Bayesian analysis. *Test, 3*, 5–124.

Bertsimas, D., & Tsitsiklis, J. (1997). *Introduction to Linear Optimization*. Athena Scientific.

Boutilier, C., Brafman, R., Domshlak, C., Hoos, H., & Poole, D. (2004). Cp-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research, 21*, 135–191.

Cano, A., & Moral, S. (2000). Algorithms for imprecise probabilities. In D. Gabbay & P. Smets (Eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (pp. 369–420). Springer.

Capotorti, A., & Formisano, A. (2008). Comparative uncertainty: theory and automation. *Mathematical Structures in Computer Science, 18*, 57–80.

Chateauneuf, A., & Jaffray, J-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences, 17*, 263–283.

Choquet, G. (1953) Theory of capacities. *Annales de l'Institut Fourier, 5*, 131–295.

Christian, R., Conder, M., & Slinko, A. (2007). Flippable pairs and subset comparisons in comparative probability orderings. *Order, 24*, 193–213.

Cohen, M. (1991). Necessary and sufficient conditions for the existence and uniqueness of weak qualitative probability measures. *Journal of Mathematical Psychology, 35*, 242–259.

de Campos, L., Huete, J., & Moral, S. (1994). Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2*, 167–196.

de Campos, L., Lamata, M.T., & Moral, S. (1990). The concept of conditional fuzzy measures. *International Journal of Intelligent Systems, 5*, 237–246.

de Cooman, G., Troffaes, M., & Miranda, E. (2008). *n*-Monotone exact functionals. *Journal of Mathematical Analysis and Applications, 347*, 143–156.

Cozman, F. (2000). Credal networks. *Artificial Intelligence, 120*, 199–233.

de Finetti, B. (1931). Sul significato soggettivo della probabilità. *Fundamenta Mathematicae, 17*, 298–329.

Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics, 38*, 325–339.

Denneberg, D. (1994). *Non-Additive Measure and Integral*. Dordrecht: Kluwer Academic.

Denoeux, T. (1999). Reasoning with imprecise belief structures. *International Journal of Approximate Reasoning, 20*, 79–111.

Destercke, S., Dubois, D., & Chojnacki, E. (2008). Unifying practical uncertainty representations: I generalised p-boxes. *International Journal of Approximate Reasoning, 49*, 649–663.

Dilworth, R. (1950). A decomposition theorem for partially ordered sets. *Annals of Mathematics, 51*, 161–166.

Dubois, D., & Prade, H. (1988). *Possibility Theory*. New York: Plenum Press.

Dubois, D., Prade, H., & Rico, A. (2013). Qualitative Capacities as Imprecise Possibilities. In L. van der Gaag (Ed.), *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2013)* (pp. 169–180).

Fagin, R., & Halpern, J. (1991) Uncertainty, belief, and probability. In *Computational Intelligence, 6*, 160–173.

Fagin, R., & Halpern, J. (1991b). A new approach to updating beliefs. In P. P. Bonissone, M. Henrion, L. N. Kanal & J. F. Lemmer (Eds.), *Uncertainty in Artificial Intelligence*, volume 6, (pp. 347–374). Amsterdam: North-Holland.

Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D., & Sentz, K. (2003). Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories.

Fine, T. (1973). *Theories of Probability*. New York: Academic Press.

Fine, T. (1979). An argument for comparative probability. In R.E. Butts & J. Hintikka (Eds.), *Basic Problems in Methodology and Linguistics* (pp. 105–119). Dordretch: Reidel.

Fishburn, P. (1986). The axioms of subjective probability. *Statistical Science, 1*, 335–345.

Fishburn, P., & Falmagne, J.C. (1989). Binary choice probabilities and rankings. *Economic Letters, 31*, 113–117.

Fujishige, S. (1991). Submodular functions and optimization. *Annals of Discrete Mathematics, 47*. North-Holland.

Fürnkranz, J., & Hüllermeier, E. (2010). *Preference learning*. Springer.

Good, I. (1962) Subjective probability as the measure of a non-measurable set. In E. Nagel, P. Suppes & A. Tarski (Eds.) *Logic, Methodology and Philosophy of Science* (pp.319–329). Stanford: Stanford University Press.

Gulordava, K. (2010) *Empirical evaluation of credal classifiers*. Master's thesis, University of Lugano. Supervised by M. Zaffalon and G. Corani.

Hopcroft, J. & Tarjan, R. (1973) Efficient algorithms for graph manipulation. *Communications of the ACM, 16*, 372–378.

Jaffray, J-Y. (1992). Bayesian updating and belief functions. *IEEE Transactions on Systems, Man and Cybernetics, 22*, 1144–1152.

Kaplan, M., & Fine, T. (1977) Joint orders in comparative probability. *Annals of Probability, 5*, 161–179.

Kofler, E. (1989) *Prognosen und Stabilität bei unvollständiger Informations*. New York: Campuss.

Koopen, M. (1995). Random utility representation of binary choice probabilities: critical graphs yielding critical necessary conditions. *Journal of Mathematical Pyschology, 39*, 21–39.

Koopman, B. (1940a) The axioms and algebra of intuitive probability. *Annals of Mathematics, 41*, 269–292.

Koopman, B. (1940b). The bases of probability. *Bulletin of the American Mathematical Society, 46*, 763–774.

Kraft, C., Pratt, J., & Seidenberg, A. (1959). Intuitive probability on finite sets. *The Annals of Mathematical Statistics, 30*, 408–419.

Kroupa, T. (2008). Geometry of possibility measures on finite sets. *International Journal of Approximate Reasoning, 48*, 237–245.

Labreuche, C. (2010). On the robustness for the Choquet integral. In *Computational Intelligence for Knowledge-Based Systems Design* (pp.

484–493). Springer.

Levi, I. (1980) *The enterprise of knowledge*. Cambridge: MIT Press.

Lu, T., & Boutilier, C. (2011) Learning mallows models with pairwise preferences. In L. Getoor & E. Xing (Eds.) *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (pp. 145–152). New York: ACM.

Marschak, J. (1960) Binary choice constraints and random utility indicators. In K.J. Arrow, S. Karlin & P. Suppes (Eds.) *Mathematical Methods in Social Sciences* (pp. 312–319). Stanford: Stanford University Press.

Miranda, E., Couso, I. & Gil, P. (2003) Extreme points of credal sets generated by 2-alternating capacities. *International Journal of Approximate Reasoning, 33*, 95–115.

Nehring, K. (2009). Imprecise probabilistic beliefs as a context for decision-making under ambiguity. *Journal of Economic Theory, 144*, 1054–1091.

Piatti, P., Antonucci, A. & Zaffalon, M. (2010). Building knowledge-based expert systems by credal networks: a tutorial. In A.R. Baswell (Ed.), *Advances in Mathematics Research, 11*. New York: Nova Science Publishers.

Regoli, G. (1996). Comparative probability and robustness. *Lecture Notes-Monograph Series 29*, 343–352.

Scott, D. (1964). Measurement structures and linear inequalities. *Journal of Mathematical Psychology, 1*, 233–247.

Seidenfeld, T., Schervisch, M. & Kadane, J. (2010). Coherent choice functions under uncertainty. *Synthese, 172*, 157–176.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Shapley, L. (1971). Cores of convex games. *International Journal of Game Theory, 1*, 11–26.

Suppes, P. (1974). The measurement of belief. *Journal of the Royal Statistical Society. Series B (Methodological), 36*, 160–191.

Walley, P. (1981). Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning, 24*, 125–148.

Walley, P., & Fine, T. (1979). Varieties of modal (classificatory) and comparative probability. *Synthese, 41*, 321–374.

Wallner, A. (2007). Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning, 44*, 339–357.

Warshall, S. (1962). A theorem on boolean matrices. *Journal of the ACM, 9*, 11–12.

Zaffalon, M., & Miranda, E. (2013). Probability and time. *Artificial Intelligence, 198*, 1–51.

Zhang, J. (2004). Binary choice, subset choice, random utility, and ranking: A unified perspective using the permutahedron. *Journal of Mathematical Psychology, 48*, 107–134.

UNIVERSITY OF OVIEDO, DEPT. OF STATISTICS AND O.R. C-CALVO SOTELO, S/N, OVIEDO, SPAIN

E-mail address: mirandaenrique@uniovi.es

CNRS, HEUDIASYC JOINT RESEARCH UNIT, UMR 7253, CENTRE DE RECHERCHE DE ROYALLIEU, UTC, F-60205 COMPIEGNE CEDEX, FRANCE

E-mail address: sebastien.destercke@hds.utc.fr