J. Math. Anal. Appl. 421 (2015) 1042–1080

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Extreme lower previsions

Jasper De Bock *, Gert de Cooman

SYSTeMS Research Group, Ghent University, Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium

ARTICLE INFO

Article history:
Received 28 January 2014
Available online 30 July 2014
Submitted by R. Curto

Keywords:
Extreme lower prevision
Fully imprecise lower prevision
Minkowski decomposition
Credal set
Extreme point
Finitely generated

ABSTRACT

Coherent lower previsions constitute a convex set that is closed and compact under the topology of point-wise convergence, and Maaß [13] has shown that any coherent lower prevision can be written as a 'countably additive convex combination' of the extreme points of this set. We show that when the possibility space has a finite number n of elements, these extreme points are either degenerate precise probabilities, or fully imprecise and in a one-to-one correspondence with Minkowski indecomposable non-empty convex compact subsets of $\mathbb{R}^{n-1}$. By exploiting this connection, we are able to prove that for $n = 3$, fully imprecise extreme lower previsions are lower envelopes of at most three linear previsions. For $n \geq 4$, 'most' fully imprecise lower previsions are extreme. Finally, we show that in our context, Maaß's result can be strengthened as follows: any coherent lower prevision can be written as, or approximated arbitrarily closely by, a finite convex combination of finitely generated extreme lower previsions.

1. Introduction

In his PhD dissertation, Maaß [13] proved a general, Choquet-like representation result for what he called inequality preserving functionals. When we apply his results to coherent lower previsions, which have an important part in the theory of imprecise probabilities, we find that the set of all coherent lower previsions defined on a subset of the linear space of all bounded real-valued maps (gambles) on a possibility space $\mathscr{X}$ constitute a convex set, that is furthermore closed and compact under the topology of point-wise convergence, and that any coherent lower prevision can be written as a 'countably additive convex combination' of the extreme points of this set; see Refs. [12] and [34, Chapter 10].

It became apparent quite soon, however, that finding these extreme coherent lower previsions was a non-trivial task, and Maaß himself formulated this as an open problem. Contributions to solving this problem were made by Quaeghebeur [22,23], who essentially concentrated on coherent lower previsions defined on

* Corresponding author.
E-mail addresses: jasper.debock@ugent.be (J. De Bock), gert.decooman@ugent.be (G. de Cooman).

http://dx.doi.org/10.1016/j.jmaa.2014.07.067
0022-247X/

finite domains. In this paper, we look at the extreme points of the set of all coherent lower previsions defined on the infinite space of *all* real-valued maps on a finite set $\mathscr{X}$, containing n elements.

We begin in Section 2 by defining (extreme) coherent lower previsions. In Section 3, we recall that coherent lower previsions are in a one-to-one relationship with non-empty convex compact sets of probability mass functions, which allows us, in Sections 4 and 5, to establish a link between extreme coherent lower previsions on the one hand, and (Minkowski) indecomposability on the other. We show that any coherent lower prevision can be uniquely decomposed in a precise and a fully imprecise part, thereby establishing that extreme lower previsions are either linear or fully imprecise. The only extreme linear previsions are the degenerate ones, and the fully imprecise extreme lower previsions are proved to be in one-to-one correspondence with Minkowski indecomposable non-empty convex compact subsets of $\mathbb{R}^{n-1}$.

This link allows us to reduce the problem of finding all extreme coherent lower previsions to a problem that has received quite a bit of attention in the mathematical literature, and to use existing solutions for that problem. Section 6 provides on overview of some of the most relevant results on Minkowski indecomposability and explains some of their implications for extreme credal sets and lower previsions. In Section 6.2, we show that for $n = 3$, fully imprecise lower previsions are extreme if and only if they are lower envelopes of at most three linear previsions. For $n \geq 4$, no such easy characterisation is available and the class of extreme lower previsions is surprisingly diverse; we provide several examples in Section 6.3.

In Section 7, we try to find out how many extreme lower previsions there are. We start with some topological groundwork in Section 7.1. This allows us to show in Section 7.2 that for $n = 3$, the extreme lower previsions are nowhere dense in the set of all fully imprecise ones. In contrast, and rather surprisingly, the answer for $n \geq 4$ is completely different, as the extreme lower previsions are now a dense G_δ subset of the fully imprecise ones. This means that for $n \geq 4$, in a categorical sense, 'most' of the fully imprecise lower previsions are extreme.

Finally, in Section 8, we show that in our context, where the cardinality of the possibility space $\mathscr{X}$ is finite, Maaß's representation result can be strengthened: any coherent lower prevision is, or can be approximated arbitrarily closely by, a finite convex combination of finitely generated extreme lower previsions. For $n \geq 4$, it even suffices to use a convex combination of degenerate linear previsions and a single, finitely generated, fully imprecise extreme lower prevision. We also investigate to which extent it is possible to obtain similar approximation results using smaller subclasses of extreme lower previsions.

We conclude in Section 9 by commenting on the theoretical and practical relevance of our results. Our main conclusion is that from a practical point of view, there are simply too many extreme coherent lower previsions and that, therefore, future research in this area should focus on particular subclasses of coherent lower previsions, either by imposing additional properties, or by restricting their domain. We provide some insight on how our work can be used to further research on these topics, present some open problems, and discuss possible avenues for future research. We end the paper by establishing a link between our results and Quaeghebeur's [22,23] work on extreme coherent lower previsions on finite domains.

In order to make our main argumentation as readable as possible, all technical proofs are collected in Appendix B, which contains additional technical results as well. Appendix A collects topological results that are used in the proofs. For the interested reader, some of these results might be of independent interest as well: for example, for finite $\mathscr{X}$, we show that the topology of pointwise convergence—when imposed on the set of all coherent lower previsions—is induced by a metric.

Part of the material in this paper has been published in an earlier conference version [3]. The present version gives a more detailed exposition of these results, provides them with proofs—omitted in the conference version—and extends them; notable extensions are the examples for the case $n \geq 4$ in Section 6.3, the topological discussion in Section 7 and Appendix A, and the results on approximation in Section 8.

2. Coherent lower previsions

Consider a variable X taking values in some non-empty set $\mathscr{X}$, called *possibility space*. We will restrict ourselves to finite possibility spaces $\mathscr{X} = \{x_1, \dots, x_n\}$, with $n \in \mathbb{N}_{>1}$.[1,2] The theory of coherent lower previsions models a subject's beliefs regarding the uncertain value of X by means of lower and upper previsions of so-called gambles. A *gamble* is a real-valued map on $\mathscr{X}$ and we use $\mathscr{G}(\mathscr{X})$ to denote the set of all of them. A lower prevision $\underline{P}$ is a real-valued functional defined on this set $\mathscr{G}(\mathscr{X})$. $\underline{P}$ is said to be *coherent* if it satisfies the following three conditions:

C1. $\underline{P}(f) \geq \min f$ for all $f \in \mathscr{G}(\mathscr{X})$
C2. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for all $f \in \mathscr{G}(\mathscr{X})$ and all $\lambda \in \mathbb{R}_{>0}$ (positive homogeneity)
C3. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathscr{G}(\mathscr{X})$ (super-additivity)

The set of all coherent lower previsions on $\mathscr{G}(\mathscr{X})$ is denoted by $\underline{\mathbb{P}}(\mathscr{X})$. The conjugate of a lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ is called an *upper prevision*. It is denoted by $\overline{P}$ and defined by $\overline{P}(f) := -\underline{P}(-f)$ for all gambles $f \in \mathscr{G}(\mathscr{X})$. Coherent lower and upper previsions can be given a behavioural interpretation in terms of buying and selling prices, turning the three conditions above into criteria for rational behaviour; see Refs. [34,35] for in-depth studies, and Refs. [1,19] for recent surveys.

2.1. Extreme lower previsions

Coherence is preserved under taking convex combinations [35, Section 2.6.4]. Consider two coherent lower previsions $\underline{P}_1$ and $\underline{P}_2$ in $\underline{\mathbb{P}}(\mathscr{X})$ and any $\lambda \in [0, 1]$. Then the lower prevision $\underline{P} = \lambda \underline{P}_1 + (1 - \lambda)\underline{P}_2$, defined by $\underline{P}(f) := \lambda \underline{P}_1(f) + (1 - \lambda)\underline{P}_2(f)$ for all $f \in \mathscr{G}(\mathscr{X})$, will also be coherent. One can now wonder whether every coherent lower prevision can be written as such a convex combination of others: given a coherent lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$, is it possible to find coherent lower previsions $\underline{P}_1$ and $\underline{P}_2$ in $\underline{\mathbb{P}}(\mathscr{X})$ and $\lambda \in [0, 1]$ such that $\underline{P} = \lambda \underline{P}_1 + (1-\lambda)\underline{P}_2$? If we exclude the trivial decompositions, where $\lambda = 0$, $\lambda = 1$ or $\underline{P}_1 = \underline{P}_2 = \underline{P}$, then the answer can be no. We will refer to those coherent lower previsions for which no non-trivial decomposition exists as *extreme lower previsions*. The goal of this paper is to characterise, and where possible to find, the set $\operatorname{ext} \underline{\mathbb{P}}(\mathscr{X})$ of all extreme lower previsions on $\mathscr{G}(\mathscr{X})$.

2.2. Special kinds of coherent lower previsions

In order to find these extreme lower previsions, it will be useful to split the set $\underline{\mathbb{P}}(\mathscr{X})$ into three disjoint subsets: linear previsions, lower previsions that are fully imprecise and lower previsions that are partially imprecise.

A coherent lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ is called a *linear prevision* if it has the additional property that $\underline{P}(f + g) = \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathscr{G}(\mathscr{X})$ or, equivalently, if $\underline{P}(f) = \overline{P}(f) = -\underline{P}(-f)$ for all $f \in \mathscr{G}(\mathscr{X})$. It is then generically denoted by P and we use $\mathbb{P}(\mathscr{X})$ to denote the set of all of them. For every *mass function* p in the so-called $\mathscr{X}$-simplex

$$\Sigma_{\mathscr{X}} := \left\{ p \in \mathbb{R}^{\mathscr{X}} \colon \sum_{i=1}^{n} p(x_i) = 1 \text{ and } p(x_i) \geq 0 \text{ for all } i \in \mathbb{N}_{\leq n} \right\},$$

[1] $\mathbb{N}$ denotes the positive integers (excluding zero) and $\mathbb{R}$ the real numbers. Subsets are denoted by using predicates as subscripts; e.g., $\mathbb{N}_{\leq n} := \{i \in \mathbb{N} \colon i \leq n\} = \{1, \dots, n\}$ denotes the positive integers up to n and $\mathbb{R}_{>0} := \{r \in \mathbb{R} \colon r > 0\}$ the strictly positive real numbers.

[2] We do not consider $n = 1$ because this case is both trivial and of no practical use. Indeed, a variable that can only assume a single value has no uncertainty associated with it.

the corresponding expectation operator P_p, defined by $P_p(f) := \sum_{i=1}^{n} f(x_i)p(x_i)$ for all $f \in \mathscr{G}(\mathscr{X})$, is a linear prevision in $\mathbb{P}(\mathscr{X})$. Conversely, every linear prevision $P \in \mathbb{P}(\mathscr{X})$ has a unique mass function $p \in \Sigma_{\mathscr{X}}$ for which $P = P_p$. It is defined by $p(x_i) := P(\mathbb{I}_{\{x_i\}})$, $i \in \mathbb{N}_{\leq n}$, where $\mathbb{I}_{\{x_i\}}$ denotes the *indicator* of $\{x_i\}$: for all $x \in \mathscr{X}$, $\mathbb{I}_{\{x_i\}}(x) = 1$ if $x = x_i$ and $\mathbb{I}_{\{x_i\}}(x) = 0$ otherwise.

Another special kind of coherent lower previsions are those that are *fully imprecise*. They are uniquely characterised by the property that $\underline{P}(\mathbb{I}_{\{x_i\}}) = 0$ for all $i \in \mathbb{N}_{\leq n}$. As we shall see further on, we can interpret $\underline{P}(\mathbb{I}_{\{x_i\}})$ as the lower probability mass of x_i, thereby making fully imprecise lower previsions those for which the lower probability mass of all elements in the possibility space is zero. We use $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ to denote the set of all such fully imprecise lower previsions. The reason why we call them fully imprecise is because they differ most from the precise, linear previsions. This distinction is already apparent from the following proposition, but will become even clearer in Section 5.1, where we shall prove that every coherent lower prevision that is neither linear nor fully imprecise can be uniquely decomposed into a linear and a fully imprecise part.

Proposition 1. *$\mathbb{P}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ are disjoint subsets of $\underline{\mathbb{P}}(\mathscr{X})$: linear previsions are never fully imprecise.*

We refer to coherent lower previsions in $\underline{\mathbb{P}}(\mathscr{X})$ that are neither fully imprecise nor linear previsions as *partially imprecise*, and we denote by $\utilde{\mathbb{P}}(\mathscr{X})$ the set of all partially imprecise lower previsions. The next corollary is a direct consequence of Proposition 1.

Corollary 2. *$\mathbb{P}(\mathscr{X})$, $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ and $\utilde{\mathbb{P}}(\mathscr{X})$ constitute a partition of $\underline{\mathbb{P}}(\mathscr{X})$.*

3. Credal sets

Linear previsions are not the only coherent lower previsions that can be characterised by means of mass functions in $\Sigma_{\mathscr{X}}$. It is well known that every coherent lower prevision can be uniquely characterised by a so-called *credal set*, which is defined as a non-empty closed (and therefore compact) convex subset of $\Sigma_{\mathscr{X}}$ [10, Section 10.2].[3] We denote a generic credal set by $\mathscr{M}$ and use $\underline{\mathbb{M}}(\mathscr{X})$ to denote the set of all of them. For any $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$, its corresponding credal set $\mathscr{M}_{\underline{P}}$ is the set of all mass functions that define a dominating linear prevision:

$$\mathscr{M}_{\underline{P}} := \bigl\{p \in \Sigma_{\mathscr{X}} \colon\ P_p(f) \geq \underline{P}(f) \text{ for all } f \in \mathscr{G}(\mathscr{X})\bigr\}.$$

The original lower prevision $\underline{P}$ and its conjugate upper prevision $\overline{P}$ can be derived from the credal set $\mathscr{M}_{\underline{P}}$: for all $f \in \mathscr{G}(\mathscr{X})$

$$\underline{P}(f) = \min\bigl\{P_p(f) \colon\ p \in \mathscr{M}_{\underline{P}}\bigr\} \quad \text{and} \quad \overline{P}(f) = \max\bigl\{P_p(f) \colon\ p \in \mathscr{M}_{\underline{P}}\bigr\}. \tag{1}$$

We can use this equation to justify our earlier statement in Section 2.2 that for all $i \in \mathbb{N}_{\leq n}$, we can interpret $\underline{P}(\mathbb{I}_{\{x_i\}})$ as the lower probability mass of x_i. Indeed,

$$\underline{P}(\mathbb{I}_{\{x_i\}}) = \min\bigl\{P_p(\mathbb{I}_{\{x_i\}}) \colon\ p \in \mathscr{M}_{\underline{P}}\bigr\} = \min\bigl\{p(x_i) \colon\ p \in \mathscr{M}_{\underline{P}}\bigr\} \tag{2}$$

is the smallest probability mass of x_i corresponding with the mass functions in $\mathscr{M}_{\underline{P}}$.

[3] Since we only consider finite possibility spaces $\mathscr{X}$, we can use the topology that is induced by the Euclidean metric, or the more intuitive metric d that is discussed in Section 7.1; it makes no difference because, as we show in Appendix A, both metrics induce the same topology. For infinite $\mathscr{X}$, Walley proves a similar equivalence between coherent lower previsions and non-empty compact convex sets of linear previsions using the weak* topology [35, Section 3.6] (the topology of pointwise convergence). For finite $\mathscr{X}$, this weak* topology can be identified with those that are induced by the metrics above; see Ref. [34, Section 5.2] and Appendix A.

Since credal sets are in a one-to-one correspondence with coherent lower previsions, we can think of a coherent lower prevision as a non-empty convex compact set of mass functions rather than as an operator on gambles. This geometric approach will be useful in our search for extreme lower previsions, since it will enable us to establish links with results already proved in fields other than coherent lower prevision theory.

3.1. Extreme credal sets

Similarly to what we have done in Section 2.1 for coherent lower previsions, we can also take convex combinations of credal sets. Consider two credal sets $\mathscr{M}_1$ and $\mathscr{M}_2$ in $\underline{\mathbb{M}}(\mathscr{X})$ and any $\lambda \in [0,1]$. Then the set $\mathscr{M} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, defined by

$$\mathscr{M} := \{\lambda p_1 + (1-\lambda)p_2\colon\ p_1 \in \mathscr{M}_1 \text{ and } p_2 \in \mathscr{M}_2\},$$

will again be a credal set in $\underline{\mathbb{M}}(\mathscr{X})$.[4] Due to the equivalence between credal sets and coherent lower previsions, the following proposition should not cause any surprise.

Proposition 3. *Consider coherent lower previsions $\underline{P}$, $\underline{P}_1$ and $\underline{P}_2$ in $\underline{\mathbb{P}}(\mathscr{X})$ and their corresponding credal sets $\mathscr{M}_{\underline{P}}$, $\mathscr{M}_{\underline{P}_1}$ and $\mathscr{M}_{\underline{P}_2}$ in $\underline{\mathbb{M}}(\mathscr{X})$. Then for all $\lambda \in [0,1]$:*

$$\underline{P} = \lambda\underline{P}_1 + (1-\lambda)\underline{P}_2 \quad \Leftrightarrow \quad \mathscr{M}_{\underline{P}} = \lambda\mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}.$$

We now define an *extreme credal set* as a credal set $\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})$ that cannot be written as a convex combination of two other credal sets $\mathscr{M}_1$ and $\mathscr{M}_2$ other than in a trivial way, trivial meaning that $\lambda = 0$, $\lambda = 1$ or $\mathscr{M}_1 = \mathscr{M}_2 = \mathscr{M}$. We will denote the set of all such extreme credal sets as $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$. The following immediate corollary of Proposition 3 shows that they are in a one-to-one correspondence with extreme lower previsions.

Corollary 4. *A coherent lower prevision is extreme if and only if its credal set is. For all $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$:*

$$\underline{P} \in \operatorname{ext}\underline{\mathbb{P}}(\mathscr{X}) \quad \Leftrightarrow \quad \mathscr{M}_{\underline{P}} \in \operatorname{ext}\underline{\mathbb{M}}(\mathscr{X}).$$

3.2. Special kinds of credal sets

Because of the one-to-one correspondence between coherent lower previsions and credal sets, the special subsets of $\underline{\mathbb{P}}(\mathscr{X})$ that were introduced in Section 2.2 immediately lead to corresponding subsets of $\underline{\mathbb{M}}(\mathscr{X})$. The set

$$\mathbb{M}(\mathscr{X}) := \{\mathscr{M}_P\colon\ P \in \mathbb{P}(\mathscr{X})\} = \{\{p\}\colon\ p \in \Sigma_{\mathscr{X}}\}$$

of credal sets that correspond to linear previsions in $\mathbb{P}(\mathscr{X})$ is the most obvious one.

Another subset of $\underline{\mathbb{M}}(\mathscr{X})$, which will become very important further on, contains those credal sets that correspond to fully imprecise coherent lower previsions:

$$\underline{\underline{\mathbb{M}}}(\mathscr{X}) := \{\mathscr{M}_{\underline{P}}\colon\ \underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})\} = \{\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})\colon\ \min\{p(x_i)\colon\ p \in \mathscr{M}\} = 0 \text{ for all } i \in \mathbb{N}_{\leq n}\},$$

where the second equality is a consequence of Eq. (2) and the definition of fully imprecise lower previsions. It should also clarify our statement in Section 2.2 that for fully imprecise lower previsions the lower probability

[4] Since $\mathscr{M}_1$ and $\mathscr{M}_2$ are convex subsets of $\Sigma_{\mathscr{X}}$, $\mathscr{M} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ is clearly also a convex subset of $\Sigma_{\mathscr{X}}$. Since compactness is preserved as well [26, Section 1.7], we infer that $\mathscr{M}$ is indeed a credal set.

mass of all elements of the possibility space is zero. We refer to elements of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ as *fully imprecise credal sets*.

The final subset of $\underline{\mathbb{M}}(\mathscr{X})$ that we need to consider contains the *partially imprecise credal sets*, corresponding to partially imprecise lower previsions in $\underset{\sim}{\mathbb{P}}(\mathscr{X})$:

$$\underset{\sim}{\mathbb{M}}(\mathscr{X}) := \left\{\mathscr{M}_{\underline{P}}\colon\ \underline{P} \in \underset{\sim}{\mathbb{P}}(\mathscr{X})\right\} = \underline{\mathbb{M}}(\mathscr{X}) \setminus \left(\mathbb{M}(\mathscr{X}) \cup \underline{\underline{\mathbb{M}}}(\mathscr{X})\right).$$

Finally, the following result is a direct consequence of Corollary 2.

Corollary 5. *$\mathbb{M}(\mathscr{X})$, $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ and $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ constitute a partition of $\underline{\mathbb{M}}(\mathscr{X})$.*

3.3. Projected credal sets

Mass functions on the possibility space $\mathscr{X} = \{x_1, \ldots, x_n\}$ are uniquely characterised by the probability of the first $n-1$ elements because the final probability follows from the requirement that $\sum_{i=1}^{n} p(x_i) = 1$. This leads us to identify a mass function p on $\mathscr{X}$ with a point v_p in $\mathbb{R}^{n-1}$, defined by $(v_p)_i := p(x_i)$ for all $i \in \mathbb{N}_{<n}$. Similarly, a credal set $\mathscr{M}$ can be identified with a subset of $\mathbb{R}^{n-1}$ by letting

$$K_{\mathscr{M}} := \{v_p\colon\ p \in \mathscr{M}\}.$$

We call $K_{\mathscr{M}}$ the *projected credal set* of $\mathscr{M}$. We will also use $K_{\underline{P}}$ as a shorthand notation for $K_{\mathscr{M}_{\underline{P}}}$ and call it the projected credal set of $\underline{P}$. For all $\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})$, $K_{\mathscr{M}}$ is a non-empty convex compact[5] (and therefore closed) subset of the so-called *projected $\mathscr{X}$-simplex*

$$\mathbf{K}_{\mathscr{X}} := \left\{v \in \mathbb{R}^{n-1}\colon\ \sum_{i=1}^{n-1} v_i \leq 1 \text{ and } v_i \geq 0 \text{ for all } i \in \mathbb{N}_{<n}\right\}, \tag{3}$$

which is a non-empty convex compact subset of $\mathbb{R}^{n-1}$. The set of all non-empty convex compact subsets of $\mathbf{K}_{\mathscr{X}}$ is denoted by $\underline{\mathbb{K}}(\mathscr{X})$. To show that both representations are indeed equivalent, let us define for every point $v \in \mathbf{K}_{\mathscr{X}}$ a corresponding mass function p_v on $\mathscr{X}$, defined by $p_v(x_i) := v_i$ for all $i \in \mathbb{N}_{<n}$ and $p_v(x_n) := 1 - \sum_{i=1}^{n-1} v_i$. It should be clear that $v_{p_v} = v$ and $p_{v_p} = p$, whence the equivalence. Similarly, we can define for all $K \in \mathbf{K}_{\mathscr{X}}$ a corresponding credal set

$$\mathscr{M}_K := \{p_v\colon\ v \in K\}.$$

Again, we have that $K_{\mathscr{M}_K} = K$ and $\mathscr{M}_{K_{\mathscr{M}}} = \mathscr{M}$. Finally, the following intuitive result shows that projecting credal sets on $\mathbf{K}_{\mathscr{X}}$ preserves convex combinations.

Proposition 6. *Consider credal sets $\mathscr{M}$, $\mathscr{M}_1$ and $\mathscr{M}_2$ in $\underline{\mathbb{M}}(\mathscr{X})$ and their corresponding projected credal sets $K_{\mathscr{M}}$, $K_{\mathscr{M}_1}$ and $K_{\mathscr{M}_2}$ in $\underline{\mathbb{K}}(\mathscr{X})$. Then for all $\lambda \in [0,1]$:*

$$\mathscr{M} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2 \quad \Leftrightarrow \quad K_{\mathscr{M}} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}.$$

[5] With respect to the topology induced by the Euclidean metric. See Appendix A for more information and a formal explanation of why $K_{\mathscr{M}}$ is compact with respect to this topology.

3.4. Special kinds of projected credal sets

Due the equivalence between credal sets and their projected versions, we can use the partition of $\underline{\mathbb{M}}(\mathscr{X})$ in Corollary 5 to construct a similar partition of $\underline{\mathbb{K}}(\mathscr{X})$. The first set in that partition corresponds to the credal sets of linear previsions and is equal to

$$\mathbb{K}(\mathscr{X}) := \{K_{\mathscr{M}}\colon\ \mathscr{M} \in \mathbb{M}(\mathscr{X})\} = \{K \in \underline{\mathbb{K}}(\mathscr{X})\colon\ K = \{v\},\ \text{with } v \in \mathbf{K}_{\mathscr{X}}\}. \tag{4}$$

The second set consists of the projections of the credal sets in $\underline{\underline{\mathbb{M}}}(\mathscr{X})$:

$$\underline{\underline{\mathbb{K}}}(\mathscr{X}) := \{K_{\mathscr{M}}\colon\ \mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})\} = \left\{K \in \underline{\mathbb{K}}(\mathscr{X})\colon\ \min_{v \in K} v_i = 0 \text{ for all } i \in \mathbb{N}_{<n} \text{ and } \max_{v \in K} \sum_{i=1}^{n-1} v_i = 1\right\}. \tag{5}$$

The final set contains the projected credal sets of partially imprecise lower previsions:

$$\utilde{\mathbb{K}}(\mathscr{X}) := \{K_{\mathscr{M}}\colon\ \mathscr{M} \in \utilde{\mathbb{M}}(\mathscr{X})\} = \underline{\mathbb{K}}(\mathscr{X}) \setminus \big(\mathbb{K}(\mathscr{X}) \cup \underline{\underline{\mathbb{K}}}(\mathscr{X})\big).$$

That these sets indeed form a partition of $\underline{\mathbb{K}}(\mathscr{X})$ follows trivially from Corollary 5.

Corollary 7. *$\mathbb{K}(\mathscr{X})$, $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ and $\utilde{\mathbb{K}}(\mathscr{X})$ constitute a partition of $\underline{\mathbb{K}}(\mathscr{X})$.*

4. Minkowski decomposition

Given two non-empty convex compact subsets A_1 and A_2 of $\mathbb{R}^{n-1}$, their *Minkowski sum* or *vector sum* is given by $A_1 + A_2 := \{a_1 + a_2\colon\ a_1 \in A_1 \text{ and } a_2 \in A_2\}$. They are called *homothetic* if $A_1 = v + \lambda A_2 := \{v + \lambda a_2\colon\ a_2 \in A_2\}$ for some $\lambda > 0$ and $v \in \mathbb{R}^{n-1}$. If $A = A_1 + A_2$, then A_1 and A_2 are called *summands* of A. We say that a non-empty convex compact subset A of $\mathbb{R}^{n-1}$ is written as a Minkowski sum in a non-trivial way—or has a non-trivial Minkowski decomposition—if neither of its summands is a singleton or homothetic to A. If there is such a non-trivial Minkowski decomposition, we say that A is *Minkowski decomposable*. Otherwise, A is called *Minkowski indecomposable*.[6] Sections 6.2 and 6.3 point to relevant literature, where, incidentally, the prefix "Minkowski" is not always used. We add it in the present paper to avoid confusion with the decomposition of credal sets and lower previsions.

4.1. Connecting both theories

One of the main contributions of this paper consists in showing how the extensive literature on Minkowski decomposition of non-empty convex compact sets can be related to the search for extreme lower previsions in imprecise probability theory. The results in this section take the first step towards doing so, and will turn out to be crucial for our results further on.

We start by associating with any non-empty compact set $A \subseteq \mathbb{R}^{n-1}$ a point $m(A) \in \mathbb{R}^{n-1}$, defined by

$$m_i(A) := \min\{v_i\colon\ v \in A\} \quad \text{for all } i \in \mathbb{N}_{<n}$$

and a real number $\mu(A)$, given by

[6] Gale was the first to introduce Minkowski indecomposable sets; he called them irreducible [5], a term which is now used to refer to a different, although related property [8, p. 322]. Ref. [7, Chapter 7] speaks of morphologically indecomposable shapes, which are not required to be convex.

$$\mu(A) := \max\left\{\sum_{i=1}^{n-1} v_i \colon v \in A\right\} - \sum_{i=1}^{n-1} m_i(A).$$

Both $m(A)$ and $\mu(A)$ are well defined due to the non-emptiness and compactness of A. If A is not a singleton, then it is easy to see that $\mu(A) > 0$ and we can define

$$\underline{\underline{A}} := \frac{1}{\mu(A)}\big[A - m(A)\big] = \left\{\frac{1}{\mu(A)}\big[v - m(A)\big] \colon v \in A\right\}. \tag{6}$$

Proposition 8. *For any non-empty convex compact subset A of $\mathbb{R}^{n-1}$ that is not a singleton, the corresponding set $\underline{\underline{A}}$ is an element of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$.*

Proposition 9. *A non-empty convex compact subset A of $\mathbb{R}^{n-1}$ that is not a singleton is Minkowski decomposable if and only if the corresponding set $\underline{\underline{A}}$ is Minkowski decomposable.*

The following result shows how the transformation that we have just introduced can be usefully exploited to reformulate the property of Minkowski decomposability.

Theorem 10. *A non-empty convex compact subset A of $\mathbb{R}^{n-1}$ that is not a singleton is Minkowski decomposable if and only if its corresponding set $\underline{\underline{A}}$ can be written as a non-trivial convex combination $\lambda K_1 + (1-\lambda) K_2$, with $K_1, K_2 \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, $K_1 \neq K_2$ and $0 < \lambda < 1$.*

5. Characterising extreme lower previsions

We now have all the tools needed to characterise the set $\operatorname{ext} \underline{\mathbb{P}}(\mathscr{X})$ of all extreme lower previsions on $\mathscr{G}(\mathscr{X})$, or equivalently, the set $\operatorname{ext} \underline{\mathbb{M}}(\mathscr{X})$ of all extreme credal sets. We will show that partially imprecise lower previsions are never extreme as they can be split up into a linear and a fully imprecise part. The only extreme linear previsions are the degenerate ones, and the fully imprecise extreme lower previsions will turn out to be closely related to the Minkowski indecomposable non-empty convex compact sets of Section 4.

5.1. Partially imprecise lower previsions

We claimed earlier on in Section 2.2 that every partially imprecise lower prevision can be uniquely decomposed in a linear and a fully imprecise part. To see why this is true, first consider the following proposition, which is the counterpart of that statement in the language of credal sets.

Proposition 11. *Any partially imprecise credal set $\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X})$ can be uniquely written as a convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ of a credal set $\mathscr{M}_1 \in \mathbb{M}(\mathscr{X})$ that contains only a single mass function $p_1 \in \Sigma_{\mathscr{X}}$ and a fully imprecise credal set $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Moreover, $0 < \lambda = \sum_{i=1}^{n} \min\{p(x_i) \colon p \in \mathscr{M}\} < 1$, the mass function p_1 that characterises $\mathscr{M}_1$ is given by $p_1(x_i) = \frac{1}{\lambda} \min\{p(x_i) \colon p \in \mathscr{M}\}$ for all $i \in \mathbb{N}_{\leq n}$, and*

$$\mathscr{M}_2 = \left\{\frac{1}{1-\lambda}p - \frac{\lambda}{1-\lambda}p_1 \colon p \in \mathscr{M}\right\}.$$

Fig. 1 should provide this result with some graphical intuition. It presents an example of a partially imprecise credal set and its decomposition into a singleton and a fully imprecise credal set. Since $n = 3$ in this example, we can depict the credal sets using the well-known simplex representation [35, Section 4.2.3].[7]

[7] Mass functions on a ternary possibility space $\mathscr{X}$ can be conveniently represented by points in the 2-dimensional probability simplex, which is an equilateral triangle with unit height. The probabilities assigned to the three elements of $\mathscr{X}$ are identified with the perpendicular distances from the three sides of the triangle.

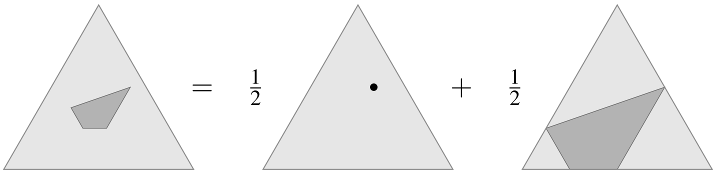

Fig. 1. Decomposition of a partially imprecise credal set ($n = 3$).

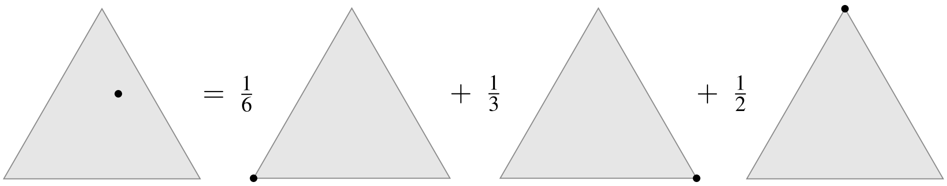

Fig. 2. Decomposition of a mass function in three degenerate ones ($n = 3$).

Corollary 12. *Any partially imprecise lower prevision* $\underline{P} \in \underset{\sim}{\underline{\mathbb{P}}}(\mathscr{X})$ *can be uniquely written as a convex combination* $\lambda P_1 + (1-\lambda)\underline{P}_2$ *of a linear prevision* $P_1 \in \mathbb{P}(\mathscr{X})$ *and a fully imprecise lower prevision* $\underline{P}_2 \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$. *Moreover,* $0 < \lambda = \sum_{i=1}^{n} \underline{P}(\mathbb{I}_{\{x_i\}}) < 1$ *and*

$$P_1(f) = \frac{1}{\lambda}\sum_{i=1}^{n} f(x_i)\underline{P}(\mathbb{I}_{\{x_i\}}) \quad \textit{and} \quad \underline{P}_2(f) = \frac{1}{1-\lambda}\underline{P}(f) - \frac{\lambda}{1-\lambda}P_1(f) \quad \textit{for all } f \in \mathscr{G}(\mathscr{X}).$$

The fact that partially imprecise models can be decomposed in this way has some immediate important consequences for extreme credal sets and lower previsions.

Corollary 13. *Extreme credal sets and lower previsions are never partially imprecise:*

$$\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X}) \quad\Rightarrow\quad \mathscr{M} \notin \operatorname{ext}\underline{\mathbb{M}}(\mathscr{X}) \quad \textit{and} \quad \underline{P} \in \underset{\sim}{\underline{\mathbb{P}}}(\mathscr{X}) \quad\Rightarrow\quad \underline{P} \notin \operatorname{ext}\underline{\mathbb{P}}(\mathscr{X}).$$

In our search for extreme lower previsions, we therefore only need to look at those lower previsions that are either linear or fully imprecise.

5.2. Linear previsions

A special class of linear previsions are those that correspond to degenerate mass functions. For every $i \in \mathbb{N}_{\leq n}$, the corresponding *degenerate mass function* $p_i^\circ \in \Sigma_{\mathscr{X}}$ has all its probability mass in x_i and is therefore defined by $p_i^\circ := \mathbb{I}_{\{x_i\}}$. These degenerate mass functions satisfy the following important property.

Proposition 14. *A credal set* $\mathscr{M} \in \mathbb{M}(\mathscr{X})$ *containing only a single mass function is extreme if and only if that single mass function is degenerate. Furthermore, any other mass function can be written as a convex combination of those degenerate ones.*

Fig. 2 depicts the decomposition of a non-degenerate mass function into degenerate ones.

The linear previsions that correspond to degenerate mass functions are called *degenerate linear previsions*. For every $i \in \mathbb{N}_{\leq n}$, we have a degenerate linear prevision P_i°, defined for all $f \in \mathscr{G}(\mathscr{X})$ by $P_i^\circ(f) := f(x_i)$. As a direct consequence of Proposition 14, we find that these degenerate linear previsions are the only extreme linear previsions.

Corollary 15. *A linear prevision* $P \in \mathbb{P}(\mathscr{X})$ *is extreme if and only if it is degenerate. Furthermore, any other linear prevision can be written as a convex combination of degenerate ones.*

For coherent lower previsions defined on a finite domain $\mathscr{K} \subset \mathscr{G}(\mathscr{X})$, a result that combines Corollaries 13 and 15 was already mentioned by Quaeghebeur [22, Proposition 1].

5.3. Fully imprecise lower previsions

So far, we have shown that partially imprecise lower previsions are never extreme and that the extreme linear previsions are those that are degenerate. The only lower previsions that are thus left to investigate are the fully imprecise ones. We start with a property of decompositions of fully imprecise credal sets.

Proposition 16. *If a fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be written as a non-trivial convex combination $\lambda\mathscr{M}_1+(1-\lambda)\mathscr{M}_2$, with $\mathscr{M}_1, \mathscr{M}_2 \in \mathbb{M}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$, then $\mathscr{M}_1$ and $\mathscr{M}_2$ are both fully imprecise and therefore elements of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.*

In the language of coherent lower previsions, this turns into the following corollary.

Corollary 17. *If a fully imprecise coherent lower prevision $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$ can be written as a non-trivial convex combination $\lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$, with $\underline{P}_1, \underline{P}_2 \in \underline{\mathbb{P}}(\mathscr{X})$, $\underline{P}_1 \neq \underline{P}_2$ and $0 < \lambda < 1$, then $\underline{P}_1$ and $\underline{P}_2$ are both fully imprecise and therefore elements of $\underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

Combined with Proposition 6 and Theorem 10, Proposition 16 leads to a crucial result.

Theorem 18. *A fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be written as a non-trivial convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, with $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\mathbb{M}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$ if and only if its projected credal set $K_{\mathscr{M}}$ is Minkowski decomposable.*

When stated in terms of coherent lower previsions, this result looks as follows.

Corollary 19. *A fully imprecise coherent lower prevision $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$ can be written as a non-trivial convex combination $\lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$, with $\underline{P}_1, \underline{P}_2 \in \underline{\mathbb{P}}(\mathscr{X})$, $\underline{P}_1 \neq \underline{P}_2$ and $0 < \lambda < 1$ if and only if its projected credal set $K_{\underline{P}}$ is Minkowski decomposable.*

The importance of these two results is that they immediately provide us with an easy characterisation of the extreme models that are fully imprecise.

Corollary 20. *A fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is extreme if and only if its projected credal set $K_{\mathscr{M}}$ is Minkowski indecomposable. Equivalently, a fully imprecise lower prevision $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$ is extreme if and only if its projected credal set $K_{\underline{P}}$ is Minkowski indecomposable.*

These alternative characterisations of fully imprecise extreme credal sets and lower previsions will allow us to import known results from the literature on Minkowski decomposability, and to use them to find the sets $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$ and $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ containing all extreme credal sets and lower previsions, respectively.

To conclude this section, we want to mention a very special fully imprecise credal set. It contains every single mass function in $\Sigma_{\mathscr{X}}$ and will be denoted as $\mathscr{M}_V := \Sigma_{\mathscr{X}}$. It is used to model complete ignorance and is called the *vacuous* credal set. The corresponding (fully imprecise) lower prevision $\underline{P}_V$ is referred to as the *vacuous* lower prevision and is given by $\underline{P}_V(f) = \min f$ for all $f \in \mathscr{G}(\mathscr{X})$.

Proposition 21. *The vacuous credal set is extreme: $\mathscr{M}_V \in \operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$.*

Corollary 22. *The vacuous lower prevision is extreme: $\underline{P}_V \in \operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$.*

6. Finding all extreme lower previsions

The size of $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$ and $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ and the complexity of their elements turn out to depend heavily on the number of elements in the possibility space $\mathscr{X} = \{x_1, \ldots, x_n\}$. We consider three distinct cases: $n = 2$, $n = 3$ and $n \geq 4$. We focus on constructing $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$, since $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ can be derived from it using Corollary 4.

6.1. Possibility spaces with two states

For $n = 2$, constructing $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$ is almost trivial. Nevertheless, it serves as a good didactic exercise to get to know the basic tools in this paper.

It follows from the results in Section 5 that in our search for the extreme credal sets, we do not need to consider the partially imprecise ones. It suffices to look at the precise and the fully imprecise credal sets. We know from Proposition 14 that of all the precise credal sets (those consisting of only a single mass function) the only extreme ones are those that correspond to a degenerate mass function. In the current binary case, with $\mathscr{X} = \{x_1, x_2\}$, this yields the extreme credal sets $\mathscr{M}_1^\circ := \{p_1^\circ\}$ and $\mathscr{M}_2^\circ := \{p_2^\circ\}$. All other extreme credal sets will be fully imprecise. We know from Proposition 21 that $\mathscr{M}_V$ is one of those fully imprecise extreme credal sets, but finding the other ones would normally require the use of Corollary 20. However, in this simple binary case, $\mathscr{M}_V$ is the only fully imprecise credal set (we leave the simple proof of this statement as an exercise for the reader) and we can therefore conclude that for binary possibility spaces:

$$\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X}) = \{\mathscr{M}_1^\circ, \mathscr{M}_2^\circ, \mathscr{M}_V\}.$$

By applying Corollary 4, we obtain the corresponding result for lower previsions:

$$\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X}) = \{P_1^\circ, P_2^\circ, \underline{P}_V\}.$$

6.2. Possibility spaces with three states

For $n = 3$, finding $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$ becomes a bit more involved. As always, the partially imprecise credal sets are never extreme and the only precise extreme credal sets are the degenerate ones. Finding the fully imprecise credal sets that are extreme is more difficult than it was in the binary case. Here, the vacuous credal set $\mathscr{M}_V$ will not be the only fully imprecise extreme credal set. In order to find the others, we rely on Corollary 20, using it to import the following result from the theory of Minkowski decomposability into our framework.

Theorem 23. *A non-empty convex compact subset of $\mathbb{R}^2$ is Minkowski indecomposable if and only if it is a triangle or a line segment [28,29].*[8]

This theorem is highly non-trivial since it holds for general non-empty convex compact subsets of $\mathbb{R}^2$ and not only for convex polygons. It allows us to derive the next result, which concludes our search for the extreme credal sets of ternary possibility spaces.

Corollary 24. *Let $\mathscr{X} = \{x_1, x_2, x_3\}$ be a possibility space that contains only three elements and consider any* fully imprecise *credal set $\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})$. Then $\mathscr{M}$ is extreme if and only if it is the convex hull of three probability mass functions: we can find $p_1, p_2, p_3 \in \Sigma_{\mathscr{X}}$ such that*

$$\mathscr{M} = \left\{\sum_{i=1}^{3} \lambda_i p_i \colon (\lambda_1, \lambda_2, \lambda_3) \in \Sigma_{\mathscr{X}}\right\}.$$

[8] This result was stated without proof by Gale [5]; we refer to the first (analytic) proof by Silverman [28,29]. Meyer [17] provides a geometric proof.

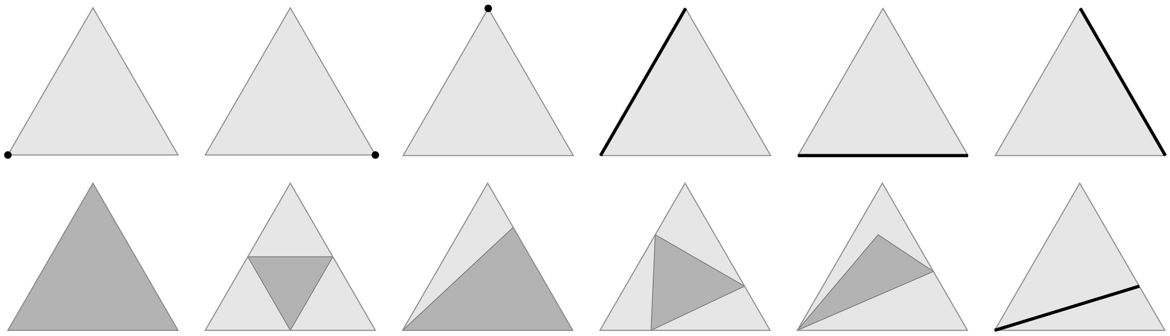

Fig. 3. Examples of extreme credal sets ($n = 3$).

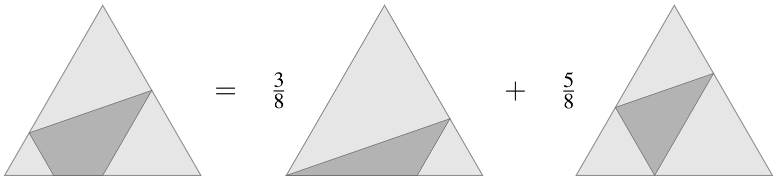

Fig. 4. Decomposition of a fully imprecise credal set into two extreme ones ($n = 3$).

Fig. 3 depicts some examples of extreme credal sets for a ternary possibility space. The first three credal sets on the top line are degenerate mass functions, all the others are fully imprecise credal sets that are the convex hull of three mass functions; the line segments correspond to cases where one of these three mass functions is part of the line segment connecting the other two.

In order to obtain the extreme lower previsions for a ternary possibility space, all we need to do is apply Corollary 4. We find that apart from the three degenerate linear previsions P_1°, P_2° and P_3°, all other extreme lower previsions are characterised by the following translation of Corollary 24.

Corollary 25. *Let $\mathscr{X} = \{x_1, x_2, x_3\}$ be a possibility space that contains only three elements and consider any* fully imprecise *lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$. Then $\underline{P}$ is extreme if and only if it is the lower envelope of three linear previsions: one can find $P_1, P_2, P_3 \in \mathbb{P}(\mathscr{X})$ such that*

$$\underline{P}(f) = \min_{i \in \mathbb{N}_{\leq 3}} P_i(f) \quad \text{for all } f \in \mathscr{G}(\mathscr{X}).$$

Since these results can be perceived as rather counterintuitive, we provide three examples of credal sets that are not extreme, and show how to decompose them into extreme ones. They should provide Corollaries 24 and 25 with some extra intuition.

The first example, which is depicted in Fig. 4, shows how a fully imprecise credal set with four vertices can be decomposed into two extreme credal sets, each of which has three vertices. It serves as a good exercise to try and see that, in this example, this decomposition is unique.

However, as illustrated by our next example, this uniqueness does not hold in general. Fig. 5 provides an example of a fully imprecise credal set that allows for two different decompositions into extreme credal sets.

As a last example, we decompose a coherent lower prevision $\underline{P}$ that cannot be written as the lower envelope of a finite number of linear previsions—whose credal set $\mathscr{M}_{\underline{P}}$ is not a polytope—into an infinite number of extreme lower previsions $\underline{P}_\lambda$, $0 \leq \lambda \leq 1$, each of which is the lower envelope of three linear previsions: P_1°, P_2° and $\lambda P_3^\circ + (1 - \lambda)P_2^\circ$. The credal set $\mathscr{M}$ that corresponds to $\underline{P}$ is depicted on the

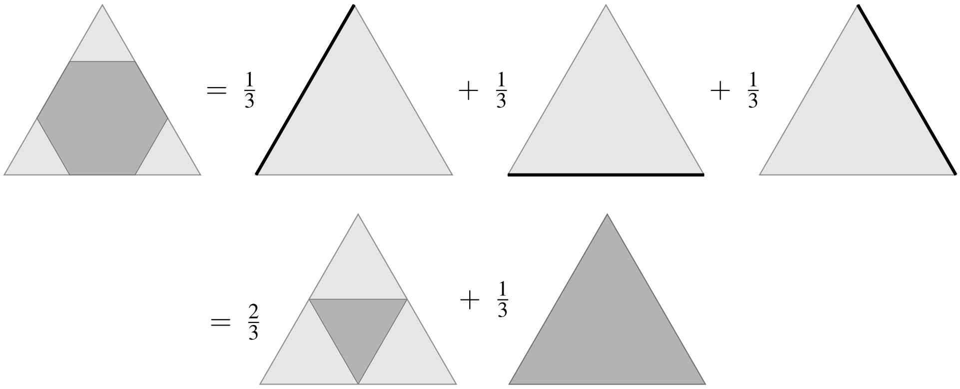

Fig. 5. Two different decompositions of the same credal set ($n = 3$).

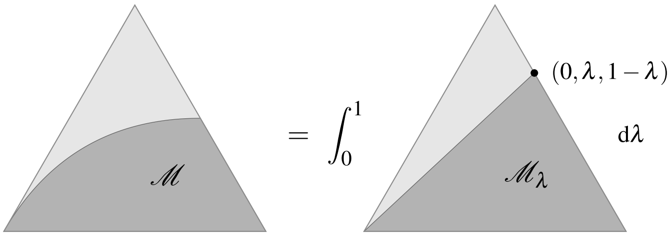

Fig. 6. Decomposition of a fully imprecise credal set into an infinity of extreme ones ($n = 3$).

left-hand side of Fig. 6. The credal set $\mathscr{M}_\lambda$ that is depicted on the right-hand side corresponds to $\underline{P}_\lambda$, with $\lambda = 0.7$; varying λ from 0 to 1 corresponds to moving the black dot upwards along the right edge of the simplex. The coherent lower prevision $\underline{P}$ can be decomposed in the following way[9,10]:

$$\underline{P}(f) = \int_0^1 \underline{P}_\lambda(f)\mathrm{d}\lambda \quad \text{for all } f \in \mathscr{G}(\mathscr{X}). \tag{7}$$

Without the availability of such a decomposition, calculating the lower prevision $\underline{P}(f)$ of some gamble f amounts to solving a convex optimisation problem: minimising $P_p(f)$ while constraining p to be an element of $\mathscr{M}$. This is non-trivial since $\mathscr{M}$ has an infinite number of vertices/constraints. Using the decomposition in Eq. (7), $\underline{P}(f)$ can be obtained either by integration or by Monte Carlo sampling. For the latter approach, it suffices to sample λ from the uniform distribution on the unit interval and keep track of the running average of $\underline{P}_\lambda(f)$. Since $\underline{P}_\lambda(f)$ can be calculated by minimising $P_p(f)$ over the three vertices of $\mathscr{M}_\lambda$, this is particularly easy. We illustrate this approach in Fig. 7.

6.3. General possibility spaces

For $n \geq 4$, finding $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$ becomes even more involved. In fact, we will no longer be able to completely characterise $\operatorname{ext}\underline{\mathbb{M}}(\mathscr{X})$, as we did for $n = 2$ and $n = 3$. It should however be clear that all extreme credal

[9] For more information about expressing lower previsions as such 'non-finitary convex combinations' of extreme lower previsions, we refer to the work by Maaß [13,34], mentioned in the Introduction.

[10] Strictly speaking, according to our definition of an extreme lower prevision, showing that $\underline{P}$ is extreme requires us to write $\underline{P}$ as a convex combination $\lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$ of *two* lower previsions instead of an infinite number of them. Such a decomposition is easily derived from the infinite decomposition in Eq. (7), for example, by considering $\lambda := 1/2$ and defining, for all $f \in \mathscr{G}(\mathscr{X})$, $\underline{P}_1(f) := 2\int_0^{0.5} \underline{P}_\lambda(f)\mathrm{d}\lambda$ and $\underline{P}_2(f) := 2\int_{0.5}^1 \underline{P}_\lambda(f)\mathrm{d}\lambda$.

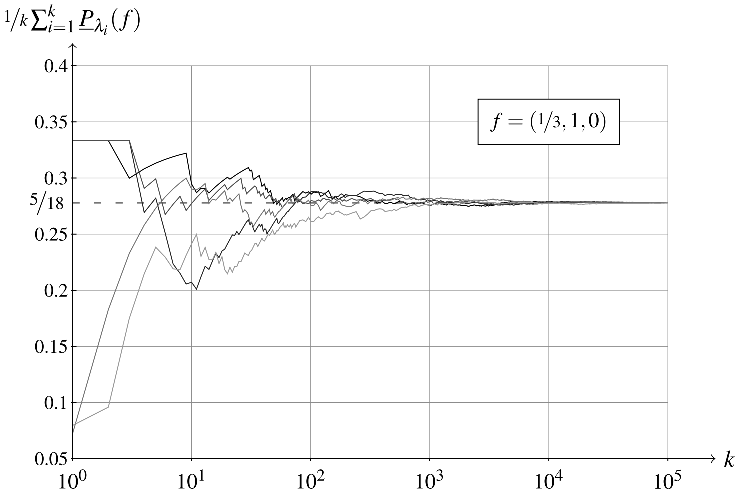

Fig. 7. Monte Carlo sampling with imprecise probabilities ($n = 3$).

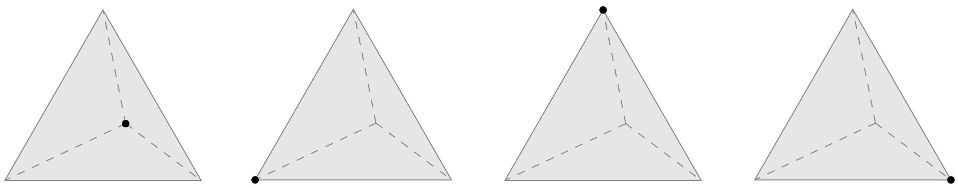

Fig. 8. Extreme credal sets that consist of a single degenerate mass function ($n = 4$).

sets will again be fully imprecise, except for those that consist of a single degenerate mass function. For $n = 4$, these degenerate mass functions are depicted in Fig. 8.[11]

We know from Corollary 20 that fully imprecise extreme credal sets correspond to Minkowski indecomposable non-empty convex compact subsets of $\mathbb{R}^{n-1}$. For $n = 3$, we were dealing with Minkowski indecomposability in the plane, which is completely determined by Theorem 23. In higher dimensions, Minkowski indecomposability is not yet fully resolved in the literature; see Ref. [26, Section 3.2] and the notes therein for a detailed overview. We state some of the most relevant results and explain their implications for extreme credal sets and, by extension, extreme lower previsions.

Most known results deal only with polytopes[12]; Grünbaum [8, Chapter 15] provides a good summary. In our language, polytopes correspond to finitely generated credal sets, which are the convex hulls of finitely many mass functions; also see Section 8 further on. Meyer presents a necessary and sufficient condition for the Minkowski indecomposability of a polytope in terms of the rank of a certain matrix, which is related to the facets of the polytope [16,18]. McMullen obtains the same results by using an arguably simpler approach based on diagrams [14]. Although these results allow Minkowski decomposability to be characterised in an algebraic manner, they are far from intuitive. Therefore, we prefer to focus on a number of easier, intuitive *sufficient* criteria for a polytope to be either Minkowski indecomposable or Minkowski decomposable. We start with some sufficient conditions for Minkowski indecomposability; the introduction to Ref. [21] provides a recent overview.

Theorem 26. *A polytope is Minkowski indecomposable if all of its 2-dimensional faces are triangles [27].*

[11] For $n = 4$, mass functions can be represented by points in the 3-dimensional probability simplex, which is a regular tetrahedron with unit height. The probabilities assigned to the four elements of the possibility space are identified with the perpendicular distances from the four facets of the tetrahedron.

[12] A polytope is the convex hull of a finite number of points [8] and therefore always non-empty, convex and compact.

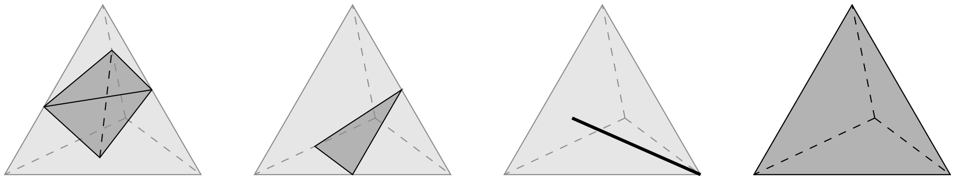

Fig. 9. Examples of credal sets that are extreme due to Corollary 27 ($n = 4$).

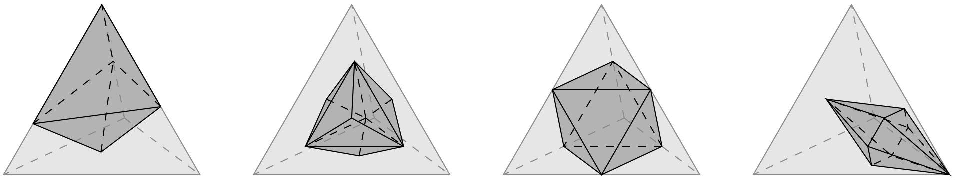

Fig. 10. Examples of extreme credal sets that correspond to simplicial polytopes ($n = 4$).

A first class of polytopes for which the condition in Theorem 26 clearly holds are the simplices. Hence, simplices are always Minkowski indecomposable,[13] which allows us to extend the sufficient part of Corollary 24 to $n \geq 4$.

Corollary 27. *Fully imprecise credal sets that are the convex hull of (at most n) affinely independent mass functions are always extreme.*[14]

Fig. 9 provides some examples of credal sets that are extreme due to Corollary 27.

If we translate Corollary 27 into the language of coherent lower previsions, we obtain the following result.

Corollary 28. *Fully imprecise lower previsions that are the lower envelope of (at most n) affinely independent linear previsions are always extreme.*

Unlike what we found for $n = 3$, the condition in Corollary 27 is not necessary for a fully imprecise credal set to be extreme. For $n \geq 4$, which corresponds to d-dimensional polytopes with $d := n - 1 \geq 3$, another important class of polytopes that are Minkowski indecomposable due to Theorem 26 are those that are simplicial, meaning that all of their facets are simplices. For $d = 3$, a polytope is simplicial if all of its facets are triangles. Fig. 10 depicts some examples of extreme, fully imprecise credal sets whose corresponding polytopes are simplicial. Note that simplices are a special case of simplicial polytopes.

Although Theorem 26 already provides us with a very large class of Minkowski indecomposable polytopes, these are not the only ones. As was first mentioned by Gale, pyramids are Minkowski indecomposable as well [5], which is why the first two credal sets in Fig. 11 are extreme. Clearly, a pyramid can have a 2-dimensional face that is not a triangle. In subsequent papers, a number of authors have established increasingly weaker sufficient conditions for a polytope to be Minkowski decomposable [11,15,21]. Most of these require that "sufficiently many" of its 2-dimensional faces are triangles. In fact, for $d = 3$, any Minkowski indecomposable polytope must have at least four triangular facets [31].[15] The credal set that is depicted on the right-hand side of Fig. 11 corresponds to a Minkowski indecomposable polytope with seven facets, only four of which are triangles [37].

[13] Gale [5] already mentioned this without proof.

[14] A set—in this case, consisting of mass functions—is called affinely independent if none of its elements can be written as an affine combination of the other elements in the set. An affine combination is a linear combination of which the coefficients sum to one.

[15] This does not extend to higher dimensions; Smilansky provides an example of a Minkowski indecomposable 4-dimensional polytope all of whose facets are decomposable [30].

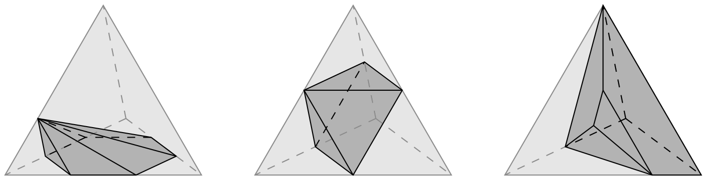

Fig. 11. Additional examples of extreme credal sets ($n = 4$).

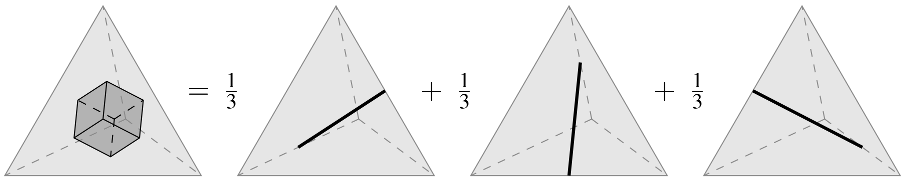

Fig. 12. Decomposition of a credal set that corresponds to a hexahedron ($n = 4$).

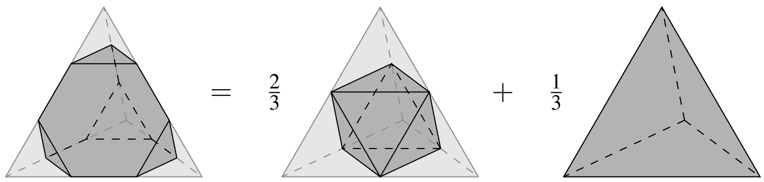

Fig. 13. Decomposition of a credal set that corresponds to a truncated simplex ($n = 4$).

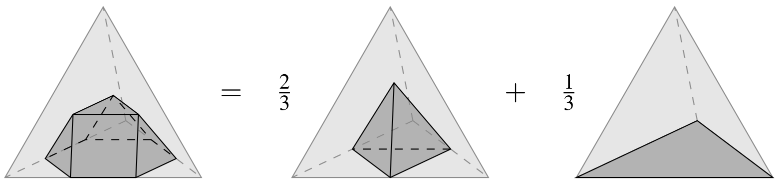

Fig. 14. Decomposition of a credal set whose corresponding polytope is not simple ($n = 4$).

Next, we turn to some sufficient criteria for Minkowski decomposability of a polytope. The following result by Shephard establishes a first, large class of decomposable polytopes.

Theorem 29. *Every simple polytope that is not a simplex is Minkowski decomposable [27].*[16]

A d-dimensional polytope is called *simple* if each of its vertices is contained in exactly d edges or, equivalently, if each of its vertices is contained in exactly d facets. Figs. 12 and 13 provide examples of credal sets whose corresponding polytopes are simple and their decomposition into extreme credal sets.

Other sufficient criteria for Minkowski decomposability of polytopes are available as well. A particularly easy one is that every 3-dimensional polytope that has more vertices than facets is Minkowski decomposable [31]. Fig. 14 provides an example of a credal set whose corresponding polytope is not simple but has more vertices than facets, including its decomposition into two extreme credal sets.

For non-polytopes, very few results are known. There is a conjecture by Gale that all 'sufficiently smooth' sets, including those whose boundaries are twice continuously differentiable, are Minkowski decomposable [5]. However, no further specification of 'sufficiently smooth' was given and, so far, no proof has been published. The most—if not only—important result seems to be due to Sallee [25], who shows that a fairly wide class of non-empty convex compact sets is decomposable. The only condition he imposes is that the boundary of the

[16] This was already mentioned without proof by Gale [5].

set contains a neighbourhood that is (i) rotund—that is, contains no line segments—and (ii) ε-smooth, with $\varepsilon > 0$. Intuitively speaking, a neighbourhood N of the boundary of some convex set C is called ε-smooth if a ball of radius ε can be moved around inside of C to touch every point of N. Since a neighbourhood that is twice continuously differentiable is guaranteed to be ε-smooth for some $\varepsilon > 0$, this bring us fairly close to Gale's conjecture, the main difference being the necessity of condition (i).

We conclude that for polytopes, Minkowski decomposability has been completely characterised on an abstract level, but no intuitive condition exists that is both necessary and sufficient. For general non-empty convex compact subsets of $\mathbb{R}^d$, almost no results are known. As illustrated by our examples, there are surprisingly many fully imprecise extreme credal sets. The next section will show, in a formal way, just how many there really are.

7. How many extreme lower previsions are there?

It should be more than clear by now that extreme lower previsions are either linear or fully imprecise and that the only extreme linear previsions are the degenerate ones, of which there are only a finite number (n). The question we will try and answer in the current section is: how many fully imprecise extreme lower previsions are there? We know from Section 6.1 that for $n = 2$, the vacuous lower prevision is the only one. For $n \geq 3$, answering this question becomes more difficult and will require some topology, which we introduce in Section 7.1. As we will see, one has to distinguish between two distinct cases: for $n = 3$, the 'minority' of the fully imprecise lower previsions are extreme, whereas, for $n \geq 4$, 'most' of them are. For ease of reference, we denote the set consisting of all *fully imprecise extreme* lower previsions by $\mathrm{ext}\,\underline{\underline{\mathbb{P}}}(\mathscr{X}) := \underline{\underline{\mathbb{P}}}(\mathscr{X}) \cap \mathrm{ext}\,\underline{\mathbb{P}}(\mathscr{X})$. Note that, due to Corollary 17, the elements of $\mathrm{ext}\,\underline{\underline{\mathbb{P}}}(\mathscr{X})$ can also be referred to as *extreme fully imprecise* lower previsions: they are extreme points of the set $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ as well as the set $\underline{\mathbb{P}}(\mathscr{X})$. An analogue statement can be made about $\mathrm{ext}\,\underline{\underline{\mathbb{M}}}(\mathscr{X}) := \underline{\underline{\mathbb{M}}}(\mathscr{X}) \cap \mathrm{ext}\,\underline{\mathbb{M}}(\mathscr{X})$; see Proposition 16.

7.1. Topological groundwork

We turn the $\mathscr{X}$-simplex $\Sigma_{\mathscr{X}}$ into a compact metric space by means of the following intuitive distance function between two probability mass functions p and p' in $\Sigma_{\mathscr{X}}$:

$$d(p,p') := \max_{A \subseteq \mathscr{X}} \left|p(A) - p'(A)\right| = \frac{1}{2} \sum_{x \in \mathscr{X}} \left|p(x) - p'(x)\right|,$$

where for all $A \subseteq \mathscr{X}$, $p(A) = \sum_{x \in A} p(x)$ is the probability of the event A.

The corresponding Hausdorff distance between two credal sets $\mathscr{M}$ and $\mathscr{M}'$ is given by

$$d_H(\mathscr{M},\mathscr{M}') := \max\Big\{\max_{p \in \mathscr{M}} \min_{p' \in \mathscr{M}'} d(p,p'), \max_{p' \in \mathscr{M}'} \min_{p \in \mathscr{M}} d(p,p')\Big\}. \tag{8}$$

Since d_H and d coincide on singletons—meaning that $d_H(\{p\},\{p'\}) = d(p,p')$—we will, from now on, denote both of these distances by d. Observe that $0 \leq d(\mathscr{M},\mathscr{M}') \leq 1$ for all credal sets $\mathscr{M}$ and $\mathscr{M}'$. By using d as a metric, $\underline{\mathbb{M}}(\mathscr{X})$ turns into a compact metric space,[17] allowing us to make topological claims such as the following.

Proposition 30. *The credal sets that are not extreme are dense in* $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.[18]

[17] See Appendix A for more information.

[18] A set A is said to be dense in a set B if B is the closure of A. In metric spaces, this is equivalent to the property that for every $b \in B$, there is a sequence of elements of A that converges to b.

In order to allow us to make similar claims for fully imprecise coherent lower previsions, we introduce the following distance between two coherent lower previsions $\underline{P}$ and $\underline{P}'$:

$$\tilde{d}(\underline{P}, \underline{P}') := \max_{f \in \mathscr{G}_1(\mathscr{X})} \left|\underline{P}(f) - \underline{P}'(f)\right|,$$

where $\mathscr{G}_1(\mathscr{X})$ consists of those gambles f on $\mathscr{X}$ for which $0 \leq f(x) \leq 1$ for all $x \in \mathscr{X}$ [32]. Due to the positive homogeneity of coherent lower previsions, $\mathscr{G}_1(\mathscr{X})$ cannot be replaced by $\mathscr{G}(\mathscr{X})$ since this would render $\tilde{d}(\underline{P}, \underline{P}')$ infinitely large for almost every choice of $\underline{P}$ and $\underline{P}'$. Some further intuition about $\tilde{d}$ can be obtained by noticing that

$$\left|\underline{P}(f) - \underline{P}'(f)\right| \leq [\max f - \min f]\tilde{d}(\underline{P}, \underline{P}') \quad \text{for all } f \in \mathscr{G}(\mathscr{X}). \tag{9}$$

What makes $\tilde{d}$ particularly interesting is that, as shown by Škulj and Hable [32, Theorem 2], the distance $\tilde{d}(\underline{P}, \underline{P}')$ between two coherent lower previsions $\underline{P}$ and $\underline{P}'$ is equal to the distance $d(\mathscr{M}_{\underline{P}}, \mathscr{M}_{\underline{P}'})$ between their corresponding credal sets. Therefore, in the sequel, we will denote both of these distances by d as well. A particularly useful consequence of this result is that it turns $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ into isometric metric spaces.[19]

Corollary 31. *The metric spaces $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$, with d as a metric, are isometric. The bijective isometry that yields this result is the one that maps every $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ to its unique corresponding $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

Consequently, every metric or topological statement about $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be easily translated to $\underline{\underline{\mathbb{P}}}(\mathscr{X})$. For example, since $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a compact metric space, $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ is a compact metric space as well. The following result provides another example.

Corollary 32. *The coherent lower previsions that are not extreme are dense in $\underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

Although the topological notions above are sufficient in order to interpret the results further on, some of their proofs require additional topological tools. We have collected these in Appendix A, which also contains further explanations for some of the topological claims that were made above.

For the interested reader, Appendix A contains additional topological results that are of independent interest. As a first example: for probability mass functions and credal sets, the topology that is induced by the metric d is identical to the one that is induced by the Euclidean distance and its associated Hausdorff distance, respectively. Another, perhaps surprising result is that, for finite possibility spaces $\mathscr{X}$, a sequence of coherent lower previsions $\underline{P}_i$ on $\mathscr{G}(\mathscr{X})$ converges with respect to the metric d if and only if it converges pointwise for every gamble $f \in \mathscr{G}(\mathscr{X})$; even stronger: for coherent lower previsions, the topology of pointwise convergence is identical to the topology that is induced by the metric d.

For the remainder of this paper, all topological statements are tacitly understood to refer to the topologies that are induced by the metric d. However, due to the aforementioned results in Appendix A, the results for credal sets can also be interpreted in terms of the Euclidean distance and its associated Hausdorff distance, and the results for lower previsions can be interpreted in terms of the topology of pointwise convergence.

7.2. Possibility spaces with three states

For $n = 3$, we have shown in Section 6.2 that fully imprecise credal sets are extreme if and only if they are the convex hull of only three probability mass functions. Although this is a fairly large class, one gets the

[19] Appendix A provides similar results for other metric spaces, including $\underline{\mathbb{M}}(\mathscr{X})$ and $\underline{\mathbb{P}}(\mathscr{X})$.

intuitive idea that most fully imprecise credal sets are not extreme. The following result shows, in a formal categorical sense, that this is indeed the case.

Theorem 33. *For $n = 3$, $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a nowhere dense closed subset of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.*[20]

Technically, this implies that it is not possible to find an open subset $\mathscr{O}$ of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$, however small, such that every credal set in $\mathscr{O}$ can be approximated arbitrarily closely by an extreme credal set in $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Loosely speaking, nowhere dense sets are 'small' in the intuitive geometric sense of being perforated with holes [20, p. 4]. We conclude that for $n = 3$, the 'minority' of the fully imprecise credal sets are extreme.

By exploiting Corollary 31, this result is easily translated into the language of coherent lower previsions. We find that for $n = 3$, the 'minority' of the fully imprecise coherent lower previsions are extreme.

Corollary 34. *For $n = 3$, $\operatorname{ext} \underline{\underline{\mathbb{P}}}(\mathscr{X})$ is a nowhere dense closed subset of $\underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

The following section will establish that this result does not extend to $n \geq 4$. In fact, the total opposite is proved to hold.

7.3. Possibility spaces with four or more states

For $n = 4$, no intuitive characterisation of extreme credal sets is available. Although we know that the fully imprecise credal sets that are not extreme are dense in $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ (see Proposition 30), the examples in Section 6.3 indicate that there are many fully imprecise extreme credal sets as well. In fact, the following theorem proves that, in a categorical sense, 'most' of the fully imprecise credal sets are extreme.

Theorem 35. *For $n \geq 4$, $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a dense G_δ subset of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.*[21]

This implies that $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be written as (and therefore trivially contains) a countable intersection of dense open subsets of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$, thereby making it a so-called comeagre set, also referred to as a residual set. Since $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a compact metric space and therefore also a Baire space, such a set exhibits properties that can be expected to hold for 'large' sets. For example, in a Baire space, every countable intersection of residual sets is residual. A property that holds on such a residual set is called generic and is considered to be 'typical' for elements of the space [26, p. 119]. Hence for $n \geq 4$, in this specific categorical sense, 'most' of the fully imprecise credal sets are extreme.

Due to Corollary 31, it is particularly simple to obtain an analogous result for coherent lower previsions. We find that for $n \geq 4$, 'most' of the fully imprecise coherent lower previsions are extreme.

Corollary 36. *For $n \geq 4$, $\operatorname{ext} \underline{\underline{\mathbb{P}}}(\mathscr{X})$ is a dense G_δ subset of $\underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

The difference with Corollary 34 is rather striking and indicates that there is a marked difference between the two cases $n = 3$ and $n \geq 4$. For example: for $n = 3$, $\operatorname{ext} \underline{\underline{\mathbb{P}}}(\mathscr{X})$ is a nowhere dense subset of $\underline{\underline{\mathbb{P}}}(\mathscr{X})$, whereas for $n \geq 4$, it is a dense subset. A similar distinction will be observed in the next section, where we study the possibility of approximating coherent lower previsions by finite convex combinations of extreme ones.

[20] A set is said to be nowhere dense if the interior of its closure is empty. A nowhere dense *closed* set is a set whose complement is open and dense.

[21] A set is said to be a G_δ set if it can be written as a countable intersection of open sets.

8. Extending upon the results by Maaß

We started this paper by recalling a result by Maaß: every coherent lower prevision can be written as a 'countably additive convex combination' of extreme ones. In fact, this result is what first inspired us to try and identify the set of all extreme lower previsions for a given finite possibility space $\mathscr{X}$. In the present section, we will show that it is not always necessary to consider 'countably additive convex combinations': within the context of this paper, where $\mathscr{X}$ is considered to be finite, every coherent lower prevision can either be written as, or approximated arbitrarily closely by, a *finite convex combination* of extreme lower previsions.

Exact results can be obtained for finitely generated models [35, Section 4.2.1]. A credal set is called *finitely generated* if it is the convex hull of a finite number of mass functions—see Section 6.3 as well—or, equivalently, if it can be specified by a finite number of linear inequality constraints on the mass functions.[22] The following result shows that every finitely generated credal set can be written as a *finite convex combination* of extreme ones.

Proposition 37. *Every finitely generated credal set can be written as a finite convex combination of finitely generated extreme credal sets.*

The corresponding coherent lower previsions, which are also called finitely generated,[23] are those that are the lower envelope of a finite number of linear previsions. For these finitely generated lower previsions, a similar result holds.

Corollary 38. *Every finitely generated lower prevision can be written as a finite convex combination of finitely generated extreme lower previsions.*

For models that are not finitely generated, the above results hold only approximately. To state this more concisely, we introduce the notion of a universal approximating class [26, p. 162]. We call a subset $\mathscr{A}$ of some set $\mathscr{S}$ a *universal approximating class* for $\mathscr{S}$ if every $S \in \mathscr{S}$ can be approximated arbitrarily closely by finite convex combinations of elements of $\mathscr{A}$. Let $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})^{\mathrm{f}}$ be the set of all finitely generated, fully imprecise, extreme credal sets—the finitely generated credal sets in $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Then our next result establishes that every fully imprecise credal set can be approximated arbitrarily closely by finite convex combinations of elements of $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})^{\mathrm{f}}$. For $n \geq 4$, it is not even necessary to consider convex combinations; a single finitely generated extreme credal set suffices.

Theorem 39. $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})^{\mathrm{f}}$ *is a universal approximating class for* $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ *and, for* $n \geq 4$*, it is even a dense subset.*

It is not hard to see the implications of this result for general—not necessarily fully imprecise—credal sets. By combining Theorem 39 with Propositions 11 and 14, one finds that every $\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})$ can be approximated arbitrarily closely by finite convex combinations of finitely generated extreme ones; for $n \geq 4$, only one of these finitely generated extreme ones needs to be fully imprecise. Furthermore, since a finite convex combination of finitely generated credal sets is again finitely generated, Theorem 39 also implies that every credal set can be approximated arbitrarily closely by a *single* finitely generated—but not necessarily extreme—one; see Lemma 60 in Appendix B as well.

---

[22] In Ref. [2], Cozman refers to such credal sets as *polytopic*.
[23] Troffaes calls them polyhedral lower previsions [33].

As usual, we can apply Corollary 31 to obtain similar results in terms of lower previsions. We let $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})^{\mathrm{f}}$ be the set of all finitely generated, fully imprecise, extreme coherent lower previsions—the finitely generated lower previsions in $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$.

Corollary 40. $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})^{\mathrm{f}}$ *is a universal approximating class for* $\underline{\mathbb{P}}(\mathscr{X})$ *and, for* $n \geq 4$*, it is even a dense subset.*

Again, it is not hard to see the implications of this result for general—not necessarily fully imprecise—coherent lower previsions; it suffices to combine Corollary 40 with Corollaries 12 and 15. We find that every coherent lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ can be approximated arbitrarily closely by finite convex combinations of finitely generated extreme ones and that for $n \geq 4$, only one of these finitely generated extreme ones needs to be fully imprecise. Alternatively, $\underline{P}$ can be approximated arbitrarily closely by a *single* finitely generated one, which is however not guaranteed to be extreme.

Theorem 39 and its corollary are especially powerful in the case $n = 3$, because there, by combining them with Corollaries 24 and 25 respectively, we find that every fully imprecise credal set or lower prevision can be approximated arbitrarily closely by finite convex combinations of very simple building blocks: fully imprecise credal sets that have at most three vertices or fully imprecise lower previsions that are the lower envelope of at most three linear previsions.

For $n \geq 4$, Theorem 39 and its corollary are far less powerful. Of course, mathematically, the result is stronger. However, it is not particularly useful since, for $n \geq 4$, $\operatorname{ext}\underline{\underline{\mathbb{M}}}(\mathscr{X})^{\mathrm{f}}$ and $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})^{\mathrm{f}}$ have no simple characterisation (yet). It would be much more interesting if instead of having to use all of $\operatorname{ext}\underline{\underline{\mathbb{M}}}(\mathscr{X})^{\mathrm{f}}$ or $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})^{\mathrm{f}}$, we could restrict ourselves to smaller, more tractable subsets that can be characterised easily. An obvious first choice would be to use the sets that are the topic of Corollaries 27 and 28, which would result in an intuitive generalisation to $n \geq 4$ of our approximation results for $n = 3$. However, as we are about to show, this is not possible. For $n \geq 4$, every universal approximating class for $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ must contain credal sets with arbitrarily high numbers of extreme points.

Proposition 41. *Let* $\mathscr{X} = \{x_1, \dots, x_n\}$ *with* $n \geq 4$*. Consider any subset* $\mathscr{A}$ *of* $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ *that consists of finitely generated credal sets and for which there is some* $m \in \mathbb{N}$ *such that every* $\mathscr{M} \in \mathscr{A}$ *is the convex hull of at most* m *mass functions. Then* $\mathscr{A}$ *is not a universal approximating class for* $\underline{\underline{\mathbb{M}}}(\mathscr{X})$*.*

By translating this result into the language of coherent lower previsions, we obtain the following rather immediate corollary.

Corollary 42. *Let* $\mathscr{X} = \{x_1, \dots, x_n\}$ *with* $n \geq 4$*. Consider any subset* $\mathscr{A}$ *of* $\underline{\mathbb{P}}(\mathscr{X})$ *that consists of finitely generated lower previsions and for which there is some* $m \in \mathbb{N}$ *such that every* $\underline{P} \in \mathscr{A}$ *is the lower envelope of at most* m *linear previsions. Then* $\mathscr{A}$ *is not a universal approximating class for* $\underline{\mathbb{P}}(\mathscr{X})$*.*

9. Concluding remarks

The main result of this paper is that, when the possibility space $\mathscr{X}$ has a finite number n of elements, the extreme coherent lower previsions on $\mathscr{G}(\mathscr{X})$ are either degenerate linear previsions or fully imprecise and in a one-to-one correspondence with Minkowski indecomposable non-empty convex compact subsets of $\mathbb{R}^{n-1}$. Using this connection, we have constructed the set $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ of all extreme lower previsions whenever possible, investigated—in a categorical sense—how many extreme lower previsions there are, and proved that every coherent lower previsions can be written as—in case it is finitely generated—, or approximated by, a finite convex combination of finitely generated extreme ones.

From a theoretical point of view, the main importance of our results is that they provide a partial answer to an open problem first formulated by Maaß [12,13]: within the set of all coherent lower previsions on

$\mathscr{G}(\mathscr{X})$, which are the extreme ones? For possibility spaces $\mathscr{X}$ with two or three elements, we provide a full answer. For $n = 2$, there are only three extreme lower previsions: two degenerate linear ones and the vacuous lower prevision. For $n = 3$, there are many more extreme lower previsions: three degenerate linear previsions and all the fully imprecise lower previsions that can be written as a lower envelope of at most three linear previsions. For $n \geq 4$, no such simple characterisation is available. However, since we prove that in that case, 'most' of the fully imprecise lower previsions are extreme, such a characterisation does not seem particularly relevant.

From a practical point of view, our most relevant results are those related to the approximation of coherent lower previsions by finite convex combinations of finitely generated extreme ones, thereby going beyond the 'countably additive convex combinations' of Maaß. However, since there are so many extreme lower previsions—at least for $n \geq 4$—, these results are not directly applicable. A possible avenue for future research would therefore be to look for a subclass $\mathscr{A}$ of $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ that (a) is sufficiently small, intuitive and tractable and yet (b) allows one to approximate a 'large enough' class $\mathscr{S}$ of coherent lower previsions by considering finite convex combinations of extreme lower previsions in $\mathscr{A}$. We say 'large enough' because, due to Corollary 42, it does not seem possible to choose $\mathscr{S} = \operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$ while still satisfying condition (a). Inspired by our results for $n = 3$, a reasonable choice for $\mathscr{A}$ seems to be the union of the degenerate linear previsions and the fully imprecise lower previsions that are the lower envelopes of at most n affinely independent linear previsions. It is not clear to us what the corresponding class $\mathscr{S}$ would look like—in other words, which coherent lower previsions could be approximated arbitrarily closely by finite convex combinations of the aforementioned class $\mathscr{A}$. We leave this as an open problem.

The most important conclusion of this paper seems to be that the set of all extreme coherent lower previsions is simply too large and that therefore, future work should restrict attention to specific subsets of coherent lower previsions, and try to identify their extreme points. Two different approaches have been taken in the literature. We introduce them briefly and explain how our present results could help advance research on these topics.

The first approach is to consider some set $\mathscr{S}$ consisting of all coherent lower previsions on $\mathscr{G}(\mathscr{X})$ that satisfy some additional property: 2-monotonicity, k-monotonicity, complete monotonicity, strong invariance and permutation invariance are but a few examples. For some of these sets, the extreme points have already been found; see [34, Chapter 10] for an overview. For example, for $\mathscr{X}$ with finite cardinality, the completely monotone coherent lower previsions are the belief functions and, as is well-known, the extreme belief functions are those that are vacuous over a non-empty subset of $\mathscr{X}$. For other potential properties, these extreme points are not yet known and, since being an element of $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X}) \cap \mathscr{S}$ is clearly a sufficient condition for being an extreme point of $\mathscr{S}$, the results in this paper could serve as a starting point to find them. It would also be interesting to investigate for which properties it is a necessary condition as well—which would make the connection with Minkowski indecomposability even more relevant—and, if it is, whether results similar to those in Sections 7 and 8 could be obtained.

The second approach is to consider lower previsions that are defined on some subset $\mathscr{F}$ of $\mathscr{G}(\mathscr{X})$ rather than on the complete set $\mathscr{G}(\mathscr{X})$. Such a lower prevision is said to be coherent if it is the restriction to $\mathscr{F}$ of a coherent lower prevision on $\mathscr{G}(\mathscr{X})$. We say that a coherent lower prevision $\underline{P}$ on $\mathscr{F}$ is extreme if it cannot be written as a proper convex combination of two other coherent lower previsions on $\mathscr{F}$. For finite $\mathscr{F}$, Ref. [22] investigates the set of all extreme coherent lower previsions on $\mathscr{F}$. By choosing $\mathscr{F}$ as the set of all indicator functions, one obtains coherent lower probabilities as a special case; see Ref. [23] for some work on extreme coherent lower probabilities. Due to the following result, the set of all extreme coherent lower previsions on $\mathscr{F}$ can be identified with a subset of $\operatorname{ext}\underline{\mathbb{P}}(\mathscr{X})$.

Proposition 43. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{F} \subseteq \mathscr{G}(\mathscr{X})$ and let $\underline{E}$ be its natural extension to $\mathscr{G}(\mathscr{X})$.*[24] *Then if $\underline{P}$ is extreme, $\underline{E}$ is extreme as well.*

Besides providing extra evidence that the set $\underline{\mathbb{P}}(\mathscr{X})$ is indeed very large, this result serves as a tool that allows one to use the extreme coherent lower previsions on finite domains in Ref. [22] and the extreme coherent lower probabilities in Ref. [23] to construct additional examples of extreme coherent lower prevision on $\mathscr{G}(\mathscr{X})$; it suffices to consider their natural extension. In order to establish an even tighter connection between our results and the results in Refs. [22,23], we would like for the reverse of the implication in Proposition 43 to hold as well: if $\underline{E}$ is extreme, then $\underline{P}$ is extreme. However, as the following example illustrates, this is not the case in general.

Example 1. Let $\mathscr{X} := \{x_1, x_2, x_3\}$ and consider the coherent lower previsions $\underline{P}$ on $\mathscr{F} := \{-\mathbb{I}_{\{x_1\}}, -\mathbb{I}_{\{x_2\}}, -\mathbb{I}_{\{x_3\}}\}$, as defined for all $f \in \mathscr{F}$ by $\underline{P}(f) := -1/2$. Let $\underline{E}$ be the natural extension of $\underline{P}$ to $\mathscr{G}(\mathscr{X})$. Then $\mathscr{M}_{\underline{E}}$ consists of all probability mass functions $p \in \Sigma_{\mathscr{X}}$ for which, for all $i \in \mathbb{N}_{\leq 3}$, $p(x_i) \leq 1/2$. This credal set is extreme—it is the second credal set on the bottom line of Fig. 3—and therefore, by Corollary 4, $\underline{E}$ is extreme as well. However, $\underline{P}$ is not extreme: $\underline{P} = 1/2(\underline{P}_1 + \underline{P}_2)$, where $\underline{P}_1$ and $\underline{P}_2$ are coherent lower previsions on $\mathscr{F}$ that are defined, for all $f \in \mathscr{F}$, by $\underline{P}_1(f) := -1/3$ and $\underline{P}_2(f) := -2/3$.

Nevertheless, we believe that, under mild conditions on $\mathscr{F}$, Proposition 43 can be strengthened, replacing the implication by an 'if and only if'. We leave this as a possible topic for future research. It would also be interesting to investigate the implications of Proposition 43—or a strengthened version—when it is combined with the results in this paper and, in particular, to see if and how the topological results in Sections 7 and 8 can be extended to the case $\mathscr{F} \neq \mathscr{G}(\mathscr{X})$. That too, we leave as a possible topic for future research.

Acknowledgments

Jasper De Bock is a PhD Fellow of the Research Foundation Flanders (FWO) and he wishes to acknowledge its financial support. Gert de Cooman's research was partially supported through project number 3G012512 of the FWO. Many thanks to Arthur Van Camp for his spontaneous impulse to produce the Monte Carlo graph of Fig. 7 and to Dirk Aeyels for providing us with relevant topology-related literature. The authors also wish to thank the anonymous referees of this paper and a previous conference version [3] for their constructive feedback, which has led us to include a number of additional results.

Appendix A. Additional topological results

We turn $\mathbb{R}^k$ into a complete metric space in the usual way, by means of the Euclidean distance $\delta(v, v') := \|v - v'\|_2 = (\sum_{i=1}^k (v_i - v'_i)^2)^{1/2}$. For $\mathscr{X} = \{x_1, \dots, x_n\}$ with $n = k + 1$, the projected $\mathscr{X}$-simplex $\mathbf{K}_{\mathscr{X}}$ is a convex subset of $\mathbb{R}^k$ that is furthermore closed and bounded and therefore compact with respect to δ. Hence, the metric δ turns $\mathbf{K}_{\mathscr{X}}$ into a compact metric space.

For $\mathscr{X} = \{x_1, \dots, x_n\}$ with $n = k$, the $\mathscr{X}$-simplex $\Sigma_{\mathscr{X}}$ can be regarded as a convex subset of $\mathbb{R}^k$ as well. It is clearly closed and bounded and therefore compact with respect to δ, so the metric δ turns $\Sigma_{\mathscr{X}}$ into a compact metric space. As we already mentioned in Section 7.1, the metric d does this as well. This is due to the following result, which establishes that both of these metrics induce the same topology on $\Sigma_{\mathscr{X}}$.

Proposition 44. $\delta(p, p') \leq 2d(p, p') \leq \sqrt{n}\delta(p, p')$ *for all* $p, p' \in \Sigma_{\mathscr{X}}$, *and therefore the metrics* δ *and* d *induce the same topology on* $\Sigma_{\mathscr{X}}$.

[24] This is the pointwise smallest coherent lower prevision on $\mathscr{G}(\mathscr{X})$ that dominates $\underline{P}$ on $\mathscr{F}$, which always exists. If $\underline{P}$ is coherent—as is the case here—, it furthermore coincides with $\underline{P}$ on $\mathscr{F}$. See Ref. [35, Section 3.1] for more information.

Proof. Using Lemma 45, we find that for all $p, p' \in \Sigma_{\mathscr{X}}$,

$$d(p,p') = \frac{1}{2}\sum_{x=1}^{n}\left|p(x) - p'(x)\right| = \frac{1}{2}\left\|p - p'\right\|_1 \leq \frac{1}{2}\sqrt{n}\left\|p - p'\right\|_2 = \frac{1}{2}\sqrt{n}\delta(p,p')$$

and

$$\delta(p,p') = \left\|p - p'\right\|_2 \leq \left\|p - p'\right\|_1 = 2d(p,p')$$

and therefore $\delta(p,p') \leq 2d(p,p') \leq \sqrt{n}\delta(p,p')$. Hence, any subset of $\Sigma_{\mathscr{X}}$ that is open with respect to any of these two metrics, is also open with respect to the other. Consequently, d and δ induce the same topology on $\Sigma_{\mathscr{X}}$. □

Lemma 45. *For any $v \in \mathbb{R}^k$, with $k \in \mathbb{N}$, let $\|v\|_1 := \sum_{i=1}^{k}|v_i|$ and $\|v\|_2 := (\sum_{i=1}^{k}(v_i)^2)^{1/2}$. Then $\|v\|_2 \leq \|v\|_1$ and $\|v\|_1 \leq \sqrt{k}\|v\|_2$.*

Proof. The first inequality is trivial. The second one follows from the Cauchy–Schwarz inequality; see for example [9, Theorem 6]. □

Proposition 44 is important because it implies that topological statements about (subsets of) $\Sigma_{\mathscr{X}}$ do not depend on whether we use δ or d as a metric. For example, it implies that $\Sigma_{\mathscr{X}}$ is compact, regardless of whether we use δ or d as a metric. Also, a sequence that converges with respect to one metric, say δ, will converge with respect to the other as well, say d.

A similar result holds for the topologies that are induced on $\mathbf{K}_{\mathscr{X}}$ and $\Sigma_{\mathscr{X}}$ by their respective metrics.

Proposition 46. *$\delta(v_p, v_{p'}) \leq \delta(p,p') \leq \sqrt{n}\delta(v_p, v_{p'})$ for all $p, p' \in \Sigma_{\mathscr{X}}$. Consequently, the topological space $\mathbf{K}_{\mathscr{X}}$, as induced by the metric δ, is homeomorphic to the topological space $\Sigma_{\mathscr{X}}$, as induced by either the metric δ or d. The homeomorphism that yields this result is the one that maps every $v \in \mathbf{K}_{\mathscr{X}}$ to its unique corresponding $p_v \in \Sigma_{\mathscr{X}}$.*

Proof. For all $p, p' \in \Sigma_{\mathscr{X}}$, we have that

$$\delta(v_p, v_{p'})^2 = \sum_{i=1}^{n-1}\left[(v_p)_i - (v_{p'})_i\right]^2 = \sum_{i=1}^{n-1}\left[p(x_i) - p'(x_i)\right]^2 \leq \sum_{i=1}^{n}\left[p(x_i) - p'(x_i)\right]^2 = \delta(p,p')^2$$

and

$$\begin{aligned}\delta(p,p')^2 = \sum_{i=1}^{n}\left[p(x_i) - p'(x_i)\right]^2 &= \left(\sum_{i=1}^{n-1}\left[p(x_i) - p'(x_i)\right]^2\right) + \left[p(x_n) - p'(x_n)\right]^2 \\ &= \delta(v_p, v_{p'})^2 + \left(\sum_{i=1}^{n-1}\left[p(x_i) - p'(x_i)\right]\right)^2 \leq n\delta(v_p, v_{p'})^2,\end{aligned}$$

where the final inequality holds because

$$\begin{aligned}\left(\sum_{i=1}^{n-1}\left[(v_p)_i - (v_{p'})_i\right]\right)^2 &\leq \left(\sum_{i=1}^{n-1}\left|p'(x_i) - p(x_i)\right|\right)^2 \\ &= \left(\|v_p - v_{p'}\|_1\right)^2 \leq (n-1)\left(\|v_p - v_{p'}\|_2\right)^2 = (n-1)\delta(v_p, v_{p'})^2,\end{aligned}$$

using Lemma 45 for the second inequality. Hence, we find that for all $p, p' \in \Sigma_{\mathscr{X}}$

$$\delta(v_p, v_{p'}) \leq \delta(p, p') \leq \sqrt{n}\delta(v_p, v_{p'}). \tag{A.1}$$

Now consider the function f that maps every $v \in \mathbf{K}_{\mathscr{X}}$ to its unique corresponding $p_v \in \Sigma_{\mathscr{X}}$. Then due to Eq. (A.1), a sequence $v_i \in \mathbf{K}_{\mathscr{X}}$, $i \in \mathbb{N}$, converges to $v \in \mathbf{K}_{\mathscr{X}}$ if and only if the sequence $p_{v_i} \in \Sigma_{\mathscr{X}}$, $i \in \mathbb{N}$, converges to $p_v \in \Sigma_{\mathscr{X}}$. Hence, both f and its inverse f^{-1} are continuous with respect to the metric δ [6, Chapter 2, Proposition 10]. Since f is clearly a bijective function as well, this means that f is a homeomorphism [36, Definition 7.8] between the topological space $\mathbf{K}_{\mathscr{X}}$, as induced by the metric δ, and the topological space $\Sigma_{\mathscr{X}}$, as induced by the metric δ or, equivalently (due to Proposition 44), the metric d. Hence, these topological spaces are homeomorphic. $\square$

This is the reason why, as was silently taken for granted in Section 3.3, for every credal set $\mathscr{M}$—which is by definition compact—, the corresponding projected credal set $K_{\mathscr{M}}$ is compact as well.

Next, we consider the set $\mathscr{C}^k$ of all non-empty convex compact subsets of $\mathbb{R}^k$. For $C, C' \in \mathscr{C}^k$, the Hausdorff distance between them, with respect to the Euclidean metric δ, is given by

$$\delta_H(C, C') := \max\Big\{\max_{v \in C} \min_{v' \in C'} \delta(v, v'), \max_{v' \in C'} \min_{v \in C} \delta(v, v')\Big\}. \tag{A.2}$$

Since δ_H and δ coincide on singletons—meaning that $\delta_H(\{v\}, \{v'\}) = \delta(v, v')$—we will from now on denote both metrics by δ. The metric δ turns $\mathscr{C}^k$ into a complete metric space [26, Theorems 1.8.2 and 1.8.5].

For $\mathscr{X} = \{x_1, \ldots, x_n\}$ with $n = k + 1$, $\underline{\mathbb{K}}(\mathscr{X})$, $\mathbb{K}(\mathscr{X})$, $\underset{\sim}{\mathbb{K}}(\mathscr{X})$ and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ are subsets of $\mathscr{C}^k$, turning them into metric spaces. $\underline{\mathbb{K}}(\mathscr{X})$, $\mathbb{K}(\mathscr{X})$ and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ are furthermore bounded, closed[25] and therefore also compact with respect to δ [26, Theorem 1.8.3]. Hence, they are compact metric spaces.

For $\mathscr{X} = \{x_1, \ldots, x_n\}$ with $n = k$, $\underline{\mathbb{M}}(\mathscr{X})$, $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be regarded as subsets of $\mathscr{C}^k$ as well, allowing us to use δ to turn them into metric spaces. Alternatively, we can do this by means of the metric d that was introduced in Section 7.1. The following result establishes that it does not matter, since both metrics induce the same topology.

Proposition 47. *It holds for all $\mathscr{M}, \mathscr{M}' \in \underline{\mathbb{M}}(\mathscr{X})$ that*

$$\delta(\mathscr{M}, \mathscr{M}') \leq 2d(\mathscr{M}, \mathscr{M}') \leq \sqrt{n}\delta(\mathscr{M}, \mathscr{M}').$$

Therefore, the metrics δ and d induce the same topology on $\underline{\mathbb{M}}(\mathscr{X})$. An analogous result holds for $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ as well.

Proof. Since we know from Proposition 44 that $\delta(p, p') \leq 2d(p, p') \leq \sqrt{n}\delta(p, p')$ for all $p, p' \in \Sigma_{\mathscr{X}}$, it is easy to infer from Eq. (A.2) that $\delta(\mathscr{M}, \mathscr{M}') \leq 2d(\mathscr{M}, \mathscr{M}') \leq \sqrt{n}\delta(\mathscr{M}, \mathscr{M}')$ for all $\mathscr{M}, \mathscr{M}' \in \underline{\mathbb{M}}(\mathscr{X})$. Hence, any subset of $\underline{\mathbb{M}}(\mathscr{X})$ that is open with respect to any of these two metrics, is also open with respect to the other. Consequently, d and δ induce the same topology on $\underline{\mathbb{M}}(\mathscr{X})$. The proof for $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is identical. $\square$

On top of this, the topological spaces $\underline{\mathbb{K}}(\mathscr{X})$, $\mathbb{K}(\mathscr{X})$, $\underset{\sim}{\mathbb{K}}(\mathscr{X})$ and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ are homeomorphic to $\underline{\mathbb{M}}(\mathscr{X})$, $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$, respectively.

[25] Boundedness follows trivially from Eq. (A.2) and the boundedness of $\mathbf{K}_{\mathscr{X}}$. $\underline{\mathbb{K}}(\mathscr{X})$, $\mathbb{K}(\mathscr{X})$ and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ are closed because (i) they are subsets of $\mathscr{C}^k$, which is complete and therefore closed and (ii) their defining properties are preserved under taking limits with respect to δ; see Eqs. (3), (4) and (5).

Proposition 48. *It holds for all $\mathscr{M}, \mathscr{M}' \in \underline{\mathbb{M}}(\mathscr{X})$ that*

$$\delta(K_{\mathscr{M}}, K_{\mathscr{M}'}) \leq \delta\big(\mathscr{M}, \mathscr{M}'\big) \leq \sqrt{n}\delta(K_{\mathscr{M}}, K_{\mathscr{M}'}).$$

Consequently, the topological space $\underline{\mathbb{K}}(\mathscr{X})$, as induced by the metric δ, is homeomorphic to the topological space $\underline{\mathbb{M}}(\mathscr{X})$, as induced by either the metric δ or d. The homeomorphism that yields this result is the one that maps every $K \in \underline{\mathbb{K}}(\mathscr{X})$ to its unique corresponding $\mathscr{M}_K \in \underline{\mathbb{M}}(\mathscr{X})$. An analogous result holds for the topological spaces $\mathbb{K}(\mathscr{X})$ and $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{K}}(\mathscr{X})$ and $\underset{\sim}{\mathbb{M}}(\mathscr{X})$, and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.

Proof. Since we know from Proposition 46 that $\delta(v_p, v_{p'}) \leq \delta(p, p') \leq \sqrt{n}\delta(v_p, v_{p'})$ for all $p, p' \in \Sigma_{\mathscr{X}}$, it is easy to infer from Eqs. (8) and (A.2) that, for all $\mathscr{M}, \mathscr{M}' \in \underline{\mathbb{M}}(\mathscr{X})$, $\delta(K_{\mathscr{M}}, K_{\mathscr{M}'}) \leq \delta(\mathscr{M}, \mathscr{M}') \leq \sqrt{n}\delta(K_{\mathscr{M}}, K_{\mathscr{M}'})$.

Now consider the function f that maps every $K \in \underline{\mathbb{K}}(\mathscr{X})$ to its unique corresponding $\mathscr{M}_K \in \underline{\mathbb{M}}(\mathscr{X})$. Then due to the inequalities above, a sequence $K_i \in \underline{\mathbb{K}}(\mathscr{X})$, $i \in \mathbb{N}$, converges to $K \in \underline{\mathbb{K}}(\mathscr{X})$ if and only if the sequence $\mathscr{M}_{K_i} \in \underline{\mathbb{M}}(\mathscr{X})$, $i \in \mathbb{N}$, converges to $\mathscr{M}_K \in \underline{\mathbb{M}}(\mathscr{X})$. Hence, both f and its inverse f^{-1} are continuous with respect to the metric δ [6, Chapter 2, Proposition 10]. Since f is clearly a bijective function as well, this means that f is a homeomorphism [36, Definition 7.8] between the topological space $\underline{\mathbb{K}}(\mathscr{X})$, as induced by the metric δ, and the topological space $\underline{\mathbb{M}}(\mathscr{X})$, as induced by the metric δ or, equivalently (due to Proposition 47), the metric d. Hence, these topological spaces are homeomorphic. The proof for the topological spaces $\mathbb{K}(\mathscr{X})$ and $\mathbb{M}(\mathscr{X})$, $\underset{\sim}{\mathbb{K}}(\mathscr{X})$ and $\underset{\sim}{\mathbb{M}}(\mathscr{X})$, and $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is completely analogous. □

As a direct consequence, we find that $\underline{\mathbb{M}}(\mathscr{X})$, $\mathbb{M}(\mathscr{X})$ and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ are compact metric spaces.

Finally, as explained in Section 7.1, the metric d can also be used to turn sets of coherent lower previsions on $\mathscr{G}(\mathscr{X})$, with $\mathscr{X}$ finite, into metric spaces. Important examples are the sets $\underline{\mathbb{P}}(\mathscr{X})$, $\mathbb{P}(\mathscr{X})$, $\underset{\sim}{\mathbb{P}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$. The following result generalises Corollary 31 and implies, as an immediate consequence, that $\underline{\mathbb{P}}(\mathscr{X})$, $\mathbb{P}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ are compact metric spaces.

Proposition 49. *The metric spaces $\underline{\mathbb{M}}(\mathscr{X})$ and $\underline{\mathbb{P}}(\mathscr{X})$, with d as a metric, are isometric. The bijective isometry that yields this result is the one that maps every $\mathscr{M} \in \underline{\mathbb{M}}(\mathscr{X})$ to its unique corresponding $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$. Analogous results hold for the metric spaces $\mathbb{M}(\mathscr{X})$ and $\mathbb{P}(\mathscr{X})$, $\underset{\sim}{\mathbb{M}}(\mathscr{X})$ and $\underset{\sim}{\mathbb{P}}(\mathscr{X})$, and $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$.*

Proof. As we have seen in Section 3, there is a one-to-one correspondence between credal sets and coherent lower previsions. By combining this with the fact that the distance $d(\underline{P}, \underline{P}')$ between two coherent lower previsions $\underline{P}$ and $\underline{P}'$ is equal to the distance $d(\mathscr{M}_{\underline{P}}, \mathscr{M}_{\underline{P}'})$ between their corresponding credal sets [32, Theorem 2], we find that the metric spaces $\underline{\mathbb{M}}(\mathscr{X})$ and $\underline{\mathbb{P}}(\mathscr{X})$, with d as their metric, are isometric. The proof for the other three couples of metric spaces is analogous. □

Interestingly, and perhaps rather surprisingly, for finite $\mathscr{X}$, convergence of a sequence of coherent lower previsions with respect to d turns out to be equivalent to pointwise convergence.

Proposition 50. *Consider a finite possibility space $\mathscr{X}$ and any coherent lower prevision $\underline{P}$ on $\mathscr{G}(\mathscr{X})$. A sequence of coherent lower previsions $\underline{P}_i \in \underline{\mathbb{P}}(\mathscr{X})$, $i \in \mathbb{N}$, converges to $\underline{P}$ with respect to the metric d if and only if the sequence $\underline{P}_i(f)$ converges to $\underline{P}(f)$ for all $f \in \mathscr{G}(\mathscr{X})$.*

Proof. Since the direct implication follows trivially from Eq. (9), we only need to prove the converse implication. So consider any $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ and any sequence of coherent lower previsions $\underline{P}_i \in \underline{\mathbb{P}}(\mathscr{X})$, $i \in \mathbb{N}$, such that $\underline{P}_i(f)$ converges to $\underline{P}(f)$ for all $f \in \mathscr{G}(\mathscr{X})$. Assume *ex absurdo* that the sequence $\underline{P}_i$, $i \in \mathbb{N}$, does not converge to $\underline{P}$ with respect to the metric d. By definition of convergence, this implies that there is some $\varepsilon > 0$

for which there is an infinite number of indices i in $\mathbb{N}$ such that $d(\underline{P}, \underline{P}_i) > \varepsilon$. Hence, we can assume without loss of generality that, for all $i \in \mathbb{N}$, $d(\underline{P}, \underline{P}_i) > \varepsilon$. By applying the definition of the metric d, this implies that there is a sequence of gambles $f_i \in \mathscr{G}_1(\mathscr{X})$, $i \in \mathbb{N}$, for which $|\underline{P}(f_i) - \underline{P}_i(f_i)| > \varepsilon$ for all $i \in \mathbb{N}$. Since $\mathscr{G}_1(\mathscr{X})$ is a bounded subset of $\mathbb{R}^n$, we can apply the Bolzano–Weierstraß theorem to find that this sequence has a convergent subsequence of gambles f_{i_j}, $j \in \mathbb{N}$, for which, of course, also $|\underline{P}(f_{i_j}) - \underline{P}_{i_j}(f_{i_j})| > \varepsilon$ for all $j \in \mathbb{N}$. Let us denote the gamble to which this subsequence converges by f. For all $j \in \mathbb{N}$, we find that

$$\begin{aligned}\left|\underline{P}(f_{i_j}) - \underline{P}_{i_j}(f_{i_j})\right| &\leq \left|\underline{P}(f_{i_j}) - \underline{P}(f)\right| + \left|\underline{P}(f) - \underline{P}_{i_j}(f)\right| + \left|\underline{P}_{i_j}(f) - \underline{P}_{i_j}(f_{i_j})\right| \\ &\leq \overline{P}\left(|f_{i_j} - f|\right) + \left|\underline{P}(f) - \underline{P}_{i_j}(f)\right| + \overline{P}_{i_j}\left(|f - f_{i_j}|\right) \leq 2\sup|f_{i_j} - f| + \left|\underline{P}(f) - \underline{P}_{i_j}(f)\right|,\end{aligned}$$

where the second and third inequality follow from coherence [35, Section 2.6.1]. The first term on the right-hand side converges to zero because f_{i_j} converges to f, and the second term on the right-hand side converges to zero because, by assumption, the sequence $\underline{P}_{i_j}(f)$ converges to $\underline{P}(f)$. Hence, the left-hand side converges to zero as well, a contradiction. □

Inspired by this equivalence, we are led to think that the topology that is induced on $\underline{\mathbb{P}}(\mathscr{X})$ by the metric d is identical to the so-called topology of pointwise convergence. As we are about to show, this is indeed the case. However, before we do so, let us recall some properties of this topology. First of all, the sets

$$B(\underline{P}, f, \varepsilon) := \left\{\underline{P}' \in \underline{\mathbb{P}}(\mathscr{X})\colon \left|\underline{P}(f) - \underline{P}'(f)\right| < \varepsilon\right\}, \quad \text{with } \underline{P} \in \underline{\mathbb{P}}(\mathscr{X}),\ f \in \mathscr{G}(\mathscr{X}) \text{ and } \varepsilon > 0,$$

constitute a sub-base for the topology of pointwise convergence [34, Section 11.1] and therefore, by definition of a sub-base [36, Definition 5.5],

$$\mathscr{B} := \left\{\bigcap_{i=1}^{m} B(\underline{P}_i, f_i, \varepsilon_i)\colon\ m \in \mathbb{N},\ \underline{P}_i \in \underline{\mathbb{P}}(\mathscr{X}),\ f_i \in \mathscr{G}(\mathscr{X}) \text{ and } \varepsilon_i > 0\right\},$$

is a base. Furthermore, this topology is the smallest—the weakest—topology that makes all evaluation functionals continuous [34, Section 11.1]. Hence, its relativisation to $\mathbb{P}(\mathscr{X})$ is identical to the weak* topology [35, Appendix D]. The following proposition establishes an important technical property of the topology of pointwise convergence.

Proposition 51. *Let $\mathscr{X}$ be finite. Then the set $\underline{\mathbb{P}}(\mathscr{X})$, endowed with the topology of pointwise convergence, is a first-countable space.*

Proof. By definition of a first-countable space [36, Definition 10.3], we have to show that every $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ has a countable neighbourhood base. So let us fix $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ and consider the set

$$\mathscr{B}_{\underline{P}}^{\mathrm{rat}} := \left\{\bigcap_{i=1}^{m} B(\underline{P}, f_i, \varepsilon_i)\colon\ m \in \mathbb{N},\ f_i \in \mathbb{Q}^{\mathscr{X}} \text{ and } \varepsilon_i \in \mathbb{Q}_{>0}\right\},$$

where $\mathbb{Q}^{\mathscr{X}} \subset \mathscr{G}(\mathscr{X})$ is the set of all rational-valued gambles on $\mathscr{X}$. Since $\mathscr{X}$ is finite, $\mathbb{Q}^{\mathscr{X}}$ is countable and therefore, $\mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ is countable as well. Furthermore, $\mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ is a subset of $\mathscr{B}$, consisting of neighbourhoods of $\underline{P}$. Hence, in order to prove that $\mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ is a countable neighbourhood base for $\underline{P}$, we are left to show that, for every neighbourhood N of $\underline{P}$, there is some $B_{\underline{P}}^{\mathrm{rat}} \in \mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ such that $B_{\underline{P}}^{\mathrm{rat}} \subseteq N$. So consider any neighbourhood N of $\underline{P}$. Then by definition of a neighbourhood, there is some open set $O \in \mathscr{B}$ such that $\underline{P} \in O \subseteq N$. Consequently, it suffices to prove that there is some $B_{\underline{P}}^{\mathrm{rat}} \in \mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ such that $B_{\underline{P}}^{\mathrm{rat}} \subseteq O$.

Since $O \in \mathscr{B}$, we know that $O = \bigcap_{i=1}^{m} B(\underline{P}_i, f_i, \varepsilon_i)$, with $m \in \mathbb{N}$ and, for all $i \in \{1, \ldots, m\}$, $\underline{P}_i \in \underline{\mathbb{P}}(\mathscr{X})$, $f_i \in \mathscr{G}(\mathscr{X})$ and $\varepsilon_i > 0$. Consider now any $i \in \{1, \ldots, m\}$. Then since $\underline{P} \in O \subseteq B(\underline{P}_i, f_i, \varepsilon_i)$, we have that

$c_i := |\underline{P}_i(f_i) - \underline{P}(f_i)| < \varepsilon_i$. Choose $\varepsilon_i^* \in \mathbb{Q}_{>0}$ and $\delta_i > 0$ such that $c_i + \varepsilon_i^* + 2\delta_i < \varepsilon_i$ and choose $f_i^* \in \mathbb{Q}^{\mathscr{X}}$ such that $\sup|f_i^* - f_i| < \delta_i$ (this is always possible). Consider any $\underline{P}' \in B(\underline{P}, f_i^*, \varepsilon_i^*)$. Then $|\underline{P}(f_i^*) - \underline{P}'(f_i^*)| < \varepsilon_i^*$ and therefore

$$\begin{aligned}\left|\underline{P}_i(f_i) - \underline{P}'(f_i)\right| &\leq \left|\underline{P}_i(f_i) - \underline{P}(f_i)\right| + \left|\underline{P}(f_i) - \underline{P}(f_i^*)\right| + \left|\underline{P}(f_i^*) - \underline{P}'(f_i^*)\right| + \left|\underline{P}'(f_i^*) - \underline{P}'(f_i)\right| \\ &< c_i + \overline{P}\left(\left|f_i - f_i^*\right|\right) + \varepsilon_i^* + \overline{P}'\left(\left|f_i - f_i^*\right|\right) \leq c_i + \varepsilon_i^* + 2\sup\left|f_i - f_i^*\right| < c_i + \varepsilon_i^* + 2\delta_i < \varepsilon_i,\end{aligned}$$

where the second and third inequalities follow from coherence [35, Section 2.6.1]. Hence, we find that $\underline{P}' \in B(\underline{P}_i, f_i, \varepsilon_i)$ and therefore also, since this holds for all $\underline{P}' \in B(\underline{P}, f_i^*, \varepsilon_i^*)$, that $B(\underline{P}, f_i^*, \varepsilon_i^*) \subseteq B(\underline{P}_i, f_i, \varepsilon_i)$. If we now define $B_{\underline{P}}^{\mathrm{rat}} := \bigcap_{i=1}^m B(\underline{P}, f_i^*, \varepsilon_i^*)$, then as required: $B_{\underline{P}}^{\mathrm{rat}} \in \mathscr{B}_{\underline{P}}^{\mathrm{rat}}$ and $B_{\underline{P}}^{\mathrm{rat}} \subseteq O$. □

By combining Propositions 50 and 51, the aforementioned claim is now easily proved.

Corollary 52. *For any finite possibility space $\mathscr{X}$, the topology that is induced on $\underline{\mathbb{P}}(\mathscr{X})$ by the metric d is identical to the topology of pointwise convergence. Analogous results hold for $\mathbb{P}(\mathscr{X})$, $\underset{\sim}{\mathbb{P}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ as well.*

Proof. Since a first-countable space is uniquely determined by its notion of convergence [36, Corollary 10.5(a)], this is a direct consequence of Propositions 50 and 51. The proof for $\mathbb{P}(\mathscr{X})$, $\underset{\sim}{\mathbb{P}}(\mathscr{X})$ and $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ is now immediate, since these are subspaces of $\underline{\mathbb{P}}(\mathscr{X})$. □

Appendix B. Proofs of results in the main text

Proof of Proposition 1. We show that a linear prevision can never be fully imprecise. So consider any linear prevision $P \in \mathbb{P}(\mathscr{X})$. We have shown in Section 2.2 that it can be uniquely characterised by a mass function $p \in \Sigma_{\mathscr{X}}$, defined by $p(x_i) := P(\mathbb{I}_{\{x_i\}})$, $i \in \mathbb{N}_{\leq n}$. Now assume *ex absurdo* that P is fully imprecise. This would mean for all $i \in \mathbb{N}_{\leq n}$ that $P(\mathbb{I}_{\{x_i\}}) = 0$, implying that $\sum_{i=1}^n p(x_i) = 0$, a contradiction. □

Proof of Corollary 2. The three sets are clearly disjoint and they cover $\underline{\mathbb{P}}(\mathscr{X})$. So it suffices to prove that none of these sets are empty. $\underline{\underline{\mathbb{P}}}(\mathscr{X})$ contains the vacuous lower prevision $\underline{P}_V := \min$, $\mathbb{P}(\mathscr{X})$ contains all so-called degenerate linear previsions P_i°, $i \in \mathbb{N}_{\leq n}$, with probability mass functions $p_i^\circ = \mathbb{I}_{\{x_i\}}$, and $\underset{\sim}{\mathbb{P}}(\mathscr{X})$ contains, for instance, the lower prevision $\frac{1}{2}\underline{P}_V + \frac{1}{2}\sum_{i=1}^n \frac{1}{n} P_i^\circ$. □

Proof of Proposition 3. First assume that $\mathscr{M}_{\underline{P}} = \lambda \mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}$. Consequently, we find for all $f \in \mathscr{G}(\mathscr{X})$ that

$$\begin{aligned}\underline{P}(f) = \min\left\{P_p(f)\colon\ p \in \mathscr{M}_{\underline{P}}\right\} &= \min\left\{P_p(f)\colon\ p = \lambda p_1 + (1-\lambda)p_2 \text{ with } p_1 \in \mathscr{M}_{\underline{P}_1} \text{ and } p_2 \in \mathscr{M}_{\underline{P}_2}\right\} \\ &= \min\left\{\lambda P_{p_1}(f) + (1-\lambda)P_{p_2}(f)\colon\ p_1 \in \mathscr{M}_{\underline{P}_1} \text{ and } p_2 \in \mathscr{M}_{\underline{P}_2}\right\} \\ &= \lambda \min\left\{P_{p_1}(f)\colon\ p_1 \in \mathscr{M}_{\underline{P}_1}\right\} + (1-\lambda)\min\left\{P_{p_2}(f)\colon\ p_2 \in \mathscr{M}_{\underline{P}_2}\right\} \\ &= \lambda \underline{P}_1(f) + (1-\lambda)\underline{P}_2(f).\end{aligned}$$

To prove the converse implication, assume that $\underline{P} = \lambda \underline{P}_1 + (1-\lambda)\underline{P}_2$ and consider the credal set $\mathscr{M}^* := \lambda \mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}$. Due to the first part of this proof, we find for all $f \in \mathscr{G}(\mathscr{X})$ that $\underline{P}_{\mathscr{M}^*}(f) = \lambda \underline{P}_1(f) + (1-\lambda)\underline{P}_2(f)$, which by assumption means that $\underline{P}_{\mathscr{M}^*}(f) = \underline{P}(f)$. Because the credal set that corresponds with a given lower prevision is unique, this means that $\mathscr{M}_{\underline{P}} = \mathscr{M}^* = \lambda \mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}$. □

Proof of Proposition 6. First of all, notice that for all p, p_1 and p_2 in $\Sigma_{\mathscr{X}}$ and $\lambda \in [0,1]$: $p = \lambda p_1 + (1-\lambda)p_2$ if and only if $v_p = \lambda v_{p_1} + (1-\lambda)v_{p_2}$. For the direct implication, assume that $\mathscr{M} = \lambda \mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$. Then

$$
\begin{aligned}
K_{\mathscr{M}} = \{v_p\colon\ p \in \mathscr{M}\} &= \bigl\{v_p\colon\ p = \lambda p_1 + (1-\lambda)p_2 \text{ with } p_1 \in \mathscr{M}_1 \text{ and } p_2 \in \mathscr{M}_2\bigr\} \\
&= \bigl\{\lambda v_{p_1} + (1-\lambda)v_{p_2}\colon\ p_1 \in \mathscr{M}_1 \text{ and } p_2 \in \mathscr{M}_2\bigr\} \\
&= \lambda\{v_{p_1}\colon\ p_1 \in \mathscr{M}_1\} + (1-\lambda)\{v_{p_2}\colon\ p_2 \in \mathscr{M}_2\} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}.
\end{aligned}
$$

For the converse implication, assume that $K_{\mathscr{M}} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}$ and introduce the credal set $\mathscr{M}^* = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$. Due to the first part of this proof, we know that $K_{\mathscr{M}^*} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}$, which by assumption means that $K_{\mathscr{M}} = K_{\mathscr{M}^*}$, implying that $\mathscr{M} = \mathscr{M}^* = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$. □

Proof of Proposition 8. Since $\underline{\underline{A}}$ is clearly non-empty, convex and compact, it suffices to show that $\min\{v_i\colon\ v \in \underline{\underline{A}}\} = 0$ for all $i \in \mathbb{N}_{<n}$ and that $\max\{\sum_{i=1}^{n-1} v_i\colon\ v \in \underline{\underline{A}}\} = 1$. Since A is not a singleton, we know that $\mu(A) > 0$. Therefore

$$
\begin{aligned}
\max\Biggl\{\sum_{i=1}^{n-1} v_i\colon\ v \in \underline{\underline{A}}\Biggr\} &= \max\Biggl\{\sum_{i=1}^{n-1} \frac{1}{\mu(A)}\bigl[v_i - m_i(A)\bigr]\colon\ v \in A\Biggr\} \\
&= \frac{1}{\mu(A)}\Biggl(\max\Biggl\{\sum_{i=1}^{n-1} v_i\colon\ v \in A\Biggr\} - \sum_{i=1}^{n-1} m_i(A)\Biggr) = 1
\end{aligned}
$$

and for any $i \in \mathbb{N}_{<n}$

$$
\min\{v_i\colon\ v \in \underline{\underline{A}}\} = \min\Biggl\{\frac{1}{\mu(A)}\bigl[v_i - m_i(A)\bigr]\colon\ v \in A\Biggr\} = \frac{1}{\mu(A)}\bigl[\min\{v_i\colon\ v \in A\} - m_i(A)\bigr] = 0. \qquad \square
$$

Proof of Proposition 9. Consider any non-empty convex compact subset A of $\mathbb{R}^{n-1}$ that is not a singleton, implying that $\mu(A) > 0$.

Let us first assume that A is Minkowski decomposable, implying that $A = A_1 + A_2$, with A_1 and A_2 non-empty convex compact subsets of $\mathbb{R}^{n-1}$ that are neither homothetic to A nor singletons. We now define $A_1' := \frac{1}{\mu(A)}[A_1 - m(A)]$ and $A_2' := \frac{1}{\mu(A)}A_2$, which are both non-empty convex compact subsets of $\mathbb{R}^{n-1}$. Neither of them is a singleton or homothetic to $\underline{\underline{A}}$, because that would contradict A_1 and A_2 not being singletons or homothetic to A. Therefore, since $\underline{\underline{A}} = A_1' + A_2'$, $\underline{\underline{A}}$ is Minkowski decomposable.

Conversely, assume that $\underline{\underline{A}}$ is Minkowski decomposable and can therefore be written as a sum $A_1' + A_2'$, with A_1' and A_2' non-empty convex compact subsets of $\mathbb{R}^{n-1}$ that are neither homothetic to $\underline{\underline{A}}$ nor singletons. Let $A_1 := \mu(A)A_1' + m(A)$ and $A_2 := \mu(A)A_2'$, which are both non-empty convex compact subsets of $\mathbb{R}^{n-1}$. Neither of them is a singleton or homothetic to A, because that would contradict with A_1' and A_2' not being singletons or homothetic to $\underline{\underline{A}}$. Therefore, since $A = A_1 + A_2$, A is Minkowski decomposable. □

Proof of Theorem 10. Consider any non-empty convex compact subset A of $\mathbb{R}^{n-1}$ that is not a singleton. First assume that A is Minkowski decomposable, implying that $A = A_1 + A_2$, with A_1 and A_2 non-empty convex compact subsets of $\mathbb{R}^{n-1}$ that are neither homothetic to A nor singletons. Since $A = A_1 + A_2$ implies that $m(A) = m(A_1) + m(A_2)$ and $\mu(A) = \mu(A_1) + \mu(A_2)$, we find that

$$
\begin{aligned}
\underline{\underline{A}} = \frac{1}{\mu(A)}\bigl[A - m(A)\bigr] &= \frac{1}{\mu(A_1) + \mu(A_2)}\bigl[A_1 + A_2 - m(A_1) - m(A_2)\bigr] \\
&= \frac{\mu(A_1)}{\mu(A_1) + \mu(A_2)}\frac{1}{\mu(A_1)}\bigl[A_1 - m(A_1)\bigr] + \frac{\mu(A_2)}{\mu(A_1) + \mu(A_2)}\frac{1}{\mu(A_2)}\bigl[A_2 - m(A_2)\bigr] \\
&= \frac{\mu(A_1)}{\mu(A_1) + \mu(A_2)}\underline{\underline{A}}_1 + \frac{\mu(A_2)}{\mu(A_1) + \mu(A_2)}\underline{\underline{A}}_2.
\end{aligned}
$$

If we now define $\lambda := \frac{\mu(A_1)}{\mu(A_1)+\mu(A_2)}$ and choose $K_1 = \underline{\underline{A}}_1$ and $K_2 = \underline{\underline{A}}_2$, we find that $\underline{\underline{A}} = \lambda K_1 + (1-\lambda)K_2$. Since $\mu(A_1) > 0$ and $\mu(A_2) > 0$, we know that $0 < \lambda < 1$. Also, K_1 and K_2 are elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ by Proposition 8. It is therefore enough to prove that $K_1 \neq K_2$. Assume *ex absurdo* that $K_1 = K_2$, implying that $\underline{\underline{A}} = K_1 = \underline{\underline{A}}_1$. This means that A_1 is homothetic to A, a contradiction.

To prove the converse implication, assume that $\underline{\underline{A}}$ can be written as a non-trivial convex combination $\lambda K_1 + (1-\lambda)K_2$, with K_1 and K_2 both elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$, $K_1 \neq K_2$ and $0 < \lambda < 1$. For the non-empty convex compact sets $A_1 := \lambda K_1$ and $A_2 := (1-\lambda)K_2$ we have $\underline{\underline{A}} = A_1 + A_2$. We are therefore done if we can show that A_1 and A_2 are neither homothetic to $\underline{\underline{A}}$, nor singletons, because this would mean that $\underline{\underline{A}}$ is Minkowski decomposable, which in turn would imply that A is Minkowski decomposable, due to Proposition 9.

We only provide the proof for A_1, since the one for A_2 is similar. First of all, A_1 cannot be a singleton because then so would be K_1, thereby contradicting $K_1 \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$ because of Corollary 7. So assume *ex absurdo* that A_1 is homothetic to $\underline{\underline{A}}$, or equivalently that K_1 is homothetic to $\underline{\underline{A}}$, which would mean that $K_1 = v + \lambda' \underline{\underline{A}}$ for some $v \in \mathbb{R}^{n-1}$ and $\lambda' > 0$. Since K_1 and $\underline{\underline{A}}$ are both elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ (for $\underline{\underline{A}}$, this is due to Proposition 8), this is only possible if $v = 0$ and $\lambda' = 1$, and therefore $K_1 = \underline{\underline{A}}$. Hence $\underline{\underline{A}} = \lambda\underline{\underline{A}} + (1-\lambda)K_2$ and therefore, by Lemma 53, $K_2 = \underline{\underline{A}} = K_1$, a contradiction. □

Lemma 53. *Consider non-empty convex compact subsets A and B of $\mathbb{R}^{n-1}$ and $0 < \lambda < 1$. Then $A = \lambda A + (1-\lambda)B$ if and only if $B = A$.*

Proof. If $B = A$, then $A = \lambda A + (1-\lambda)B$ holds trivially because A is convex.

For the converse implication, we start by proving that $B \subseteq A$. Consider any $b \in B$, then we prove that $b \in A$. Indeed, fix any $a \in A$, then by assumption also $\lambda a + (1-\lambda)b \in A$, and therefore also $\lambda^2 a + (1-\lambda^2)b = \lambda[\lambda a + (1-\lambda)b] + (1-\lambda)b \in A$. Continuing in the same vein, we find that A contains every element of the sequence $\lambda^k a + (1-\lambda^k)b$, $k \in \mathbb{N}$, and since A is compact and therefore closed (since $\mathbb{R}^{n-1}$ is Hausdorff) and since $0 < \lambda < 1$, the limit b of this sequence also belongs to A.

Next, we prove that $A \subseteq B$. Consider any extreme point a_{ext} of A: any point in A that cannot be written as a convex combination of two other points in A. Since $a_{\mathrm{ext}} \in A$ and $A = \lambda A + (1-\lambda)B$, we infer that there are $a \in A$ and $b \in B \subseteq A$ such that $a_{\mathrm{ext}} = \lambda a + (1-\lambda)b$. Since a_{ext} is an extreme point of A and $0 < \lambda < 1$, the only way for this to hold is if $a_{\mathrm{ext}} = a = b$ and therefore $a_{\mathrm{ext}} \in B$. Hence all extreme points of A belong to B, and since by Minkowski's extreme point theorem [24, Corollary 18.5.1], the compact and convex set A is the convex hull of its extreme points, we see that A is included in the convex hull of B, which in turn is equal to B since B is convex. □

Proof of Proposition 11. Consider any partially imprecise credal set $\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X})$. First of all, suppose that it can indeed be written as a convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ of a credal set $\mathscr{M}_1 \in \mathbb{M}(\mathscr{X})$ that contains only a single mass function $p_1 \in \Sigma_{\mathscr{X}}$ and a fully imprecise credal set $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. It will then hold for all $i \in \mathbb{N}_{\leq n}$ that

$$\begin{aligned}\min\{p(x_i)\colon\ p \in \mathscr{M}\} &= \min\{\lambda p_1(x_i) + (1-\lambda)p_2(x_i)\colon\ p_2 \in \mathscr{M}_2\}\\ &= \lambda p_1(x_i) + (1-\lambda)\min\{p_2(x_i)\colon\ p_2 \in \mathscr{M}_2\} = \lambda p_1(x_i).\end{aligned}$$

Since p_1 is a mass function, this implies that

$$\sum_{i=1}^{n} \min\{p(x_i)\colon\ p \in \mathscr{M}\} = \sum_{i=1}^{n} \lambda p(x_i) = \lambda \sum_{i=1}^{n} p(x_i) = \lambda. \tag{B.1}$$

We already know that $0 \leq \lambda \leq 1$ because it is the coefficient of a convex combination but we can also show that $\lambda \neq 0$ and $\lambda \neq 1$, implying that $0 < \lambda < 1$. Indeed, $\lambda = 0$ would mean that $\mathscr{M} = \mathscr{M}_2$ is fully imprecise and $\lambda = 1$ would mean that $\mathscr{M} = \mathscr{M}_1$ contains only a single mass function, both contradicting $\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X})$, due to Corollary 5. Since $\lambda > 0$, we find for all $i \in \mathbb{N}_{\leq n}$ that

$$p_1(x_i) = \frac{1}{\lambda}\min\{p(x_i)\colon p \in \mathscr{M}\}, \tag{B.2}$$

implying that the set $\mathscr{M}_1 = \{p_1\}$ in the convex combination above is indeed unique. Also, we derive from $\lambda < 1$ and $\mathscr{M} = \lambda\{p_1\} + (1-\lambda)\mathscr{M}_2$ that

$$\mathscr{M}_2 = \frac{1}{1-\lambda}\mathscr{M} - \frac{\lambda}{1-\lambda}\{p_1\} = \left\{\frac{1}{1-\lambda}p - \frac{\lambda}{1-\lambda}p_1\colon p \in \mathscr{M}\right\}, \tag{B.3}$$

implying that the set $\mathscr{M}_2$ in the convex combination above is unique as well.

To conclude the proof, we now only need to show that λ, p_1 and $\mathscr{M}_2$, as given by Eqs. (B.1), (B.2) and (B.3), satisfy the following properties: $0 < \lambda < 1$, $p_1 \in \Sigma_{\mathscr{X}}$, $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ and $\mathscr{M} = \lambda\{p_1\} + (1-\lambda)\mathscr{M}_2$.

It is obvious that $0 \leq \lambda \leq 1$, so assume *ex absurdo* that $\lambda = 0$. Eq. (B.1) then implies that $\min\{p(x_i)\colon p \in \mathscr{M}\} = 0$ for all $i \in \mathbb{N}_{\leq n}$, making $\mathscr{M}$ fully imprecise and contradicting $\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X})$, due to Corollary 5. Next, assume *ex absurdo* that $\lambda = 1$. Eq. (B.2) then implies for all $p \in \mathscr{M}$ and all $i \in \mathbb{N}_{\leq n}$ that $p(x_i) = \min\{p(x_i)\colon p \in \mathscr{M}\}$, making $\mathscr{M}$ a singleton and contradicting $\mathscr{M} \in \underset{\sim}{\mathbb{M}}(\mathscr{X})$, due to Corollary 5. That $p_1 \in \Sigma_{\mathscr{X}}$ is now an immediate consequence of Eqs. (B.1) and (B.2), and $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ holds because we have for all $i \in \mathbb{N}_{\leq n}$ that

$$\begin{aligned}\min\{p_2(x_i)\colon p_2 \in \mathscr{M}_2\} &= \min\left\{\frac{1}{1-\lambda}p(x_i) - \frac{\lambda}{1-\lambda}p_1(x_i)\colon p \in \mathscr{M}\right\} \\ &= \frac{1}{1-\lambda}\min\{p(x_i)\colon p \in \mathscr{M}\} - \frac{\lambda}{1-\lambda}p_1(x_i) = 0,\end{aligned}$$

where the first equality follows from Eq. (B.3) and the last one from Eq. (B.2). Finally, we infer from Eq. (B.3) that

$$\begin{aligned}\lambda\{p_1\} + (1-\lambda)\mathscr{M}_2 = \{\lambda p_1 + (1-\lambda)p_2\colon p_2 \in \mathscr{M}_2\} &= \left\{\lambda p_1 + (1-\lambda)\left(\frac{1}{1-\lambda}p - \frac{\lambda}{1-\lambda}p_1\right)\colon p \in \mathscr{M}\right\} \\ &= \{p\colon p \in \mathscr{M}\} = \mathscr{M}. \qquad \square\end{aligned}$$

Proof of Corollary 12. Since $\underline{P}$ is partially imprecise, so is $\mathscr{M}_{\underline{P}}$, and we can therefore use Proposition 11 to see that $\mathscr{M}_{\underline{P}}$ can be uniquely written as a convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ of a credal set $\mathscr{M}_1 \in \mathbb{M}(\mathscr{X})$ that contains only a single mass function $p_1 \in \Sigma_{\mathscr{X}}$ and a fully imprecise credal set $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. In this expression, we have that $0 < \lambda = \sum_{i=1}^n \min\{p(x_i)\colon p \in \mathscr{M}\} < 1$, which turns into $0 < \lambda = \sum_{i=1}^n \underline{P}(\mathbb{I}_{\{x_i\}}) < 1$ since $\underline{P}(\mathbb{I}_{\{x_i\}}) = \min\{p(x_i)\colon p \in \mathscr{M}\}$ for all $i \in \mathbb{N}_{\leq n}$. The credal set $\mathscr{M}_1 = \{p_1\}$, where the mass function p_1 is given by $p_1(x_i) = \frac{1}{\lambda}\min\{p(x_i)\colon p \in \mathscr{M}\}$ for all $i \in \mathbb{N}_{\leq n}$, has a corresponding linear prevision P_1. For all $f \in \mathscr{G}(\mathscr{X})$, we have that

$$P_1(f) := P_{\mathscr{M}_1}(f) = P_{p_1}(f) = \sum_{i=1}^n f(x_i)p_1(x_i) = \frac{1}{\lambda}\sum_{i=1}^n f(x_i)\underline{P}(\mathbb{I}_{\{x_i\}}).$$

Similarly, the fully imprecise credal set $\mathscr{M}_2$ has a corresponding fully imprecise lower prevision $\underline{P}_2 := \underline{P}_{\mathscr{M}_2} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$. By Proposition 3, $\mathscr{M}_{\underline{P}} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ implies that $\underline{P} = \lambda P_1 + (1-\lambda)\underline{P}_2$, so we find that

$$\underline{P}_2(f) = \frac{1}{1-\lambda}\underline{P}(f) - \frac{\lambda}{1-\lambda}P_1(f) \quad \text{for all } f \in \mathscr{G}(\mathscr{X}).$$

Now to prove that this decomposition is unique, consider any $\lambda' \in [0,1]$, $P'_1 \in \mathbb{P}(\mathscr{X})$ and $\underline{P}'_2 \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$ such that $\underline{P} = \lambda' P'_1 + (1-\lambda')\underline{P}'_2$. By Proposition 3, this implies that $\mathscr{M}_{\underline{P}} = \lambda' \mathscr{M}_{P'_1} + (1-\lambda')\mathscr{M}_{\underline{P}'_2}$, where $\mathscr{M}_{P'_1}$ is a singleton and $\mathscr{M}_{\underline{P}'_2}$ is fully imprecise. Proposition 11 then tells us that $\lambda' = \lambda$, $\mathscr{M}_{P'_1} = \mathscr{M}_1$ and $\mathscr{M}_{\underline{P}'_2} = \mathscr{M}_2$, and therefore, due to the one-to-one correspondence between credal sets and coherent lower previsions, $P_1 = P'_1$ and $\underline{P}'_2 = \underline{P}_2$. This shows that the decomposition is indeed unique. □

Proof of Corollary 13. This follows trivially from Proposition 11 and Corollary 12. □

Proof of Proposition 14. We start by proving that a credal set that consists of a single degenerate mass function is an extreme credal set. Consider $i \in \mathbb{N}_{\leq n}$ and assume *ex absurdo* that the credal set $\mathscr{M} = \{p_i^\circ\}$ is not extreme. By Lemma 54, this implies the existence of two mass functions $p_1, p_2 \in \Sigma_{\mathscr{X}}$ and $0 < \lambda < 1$ such that $p_1 \neq p_2$ and $p_i^\circ = \lambda p_1 + (1-\lambda)p_2$, implying that $1 = p_i^\circ(x_i) = \lambda p_1(x_i) + (1-\lambda)p_2(x_i)$. Since the probabilities $p_1(x_i)$ and $p_2(x_i)$ cannot exceed one, this implies that $p_1(x_i) = p_2(x_i) = 1$, which in turn implies that $p_1 = p_2 = p_i^\circ$, a contradiction. Hence, $\{p_i^\circ\}$ is an extreme credal set.

Any mass function $p \in \Sigma_{\mathscr{X}}$ can be written as a convex combination of degenerate ones: we have that $p = \sum_{i=1}^n p(x_i)p_i^\circ$ and therefore also that $\{p\} = \sum_{i=1}^n p(x_i)\{p_i^\circ\}$. If p is not degenerate then there are at least two $i \neq j$ such that $p(x_i) > 0$ and $p(x_j) > 0$, so p is a non-trivial convex combination of at least two mass functions and therefore $\{p\}$ is not an extreme credal set. □

Lemma 54. *If a credal set $\mathscr{M} = \{p\} \in \mathbb{M}(\mathscr{X})$, $p \in \Sigma_{\mathscr{X}}$, can be written as a non-trivial convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, where $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\mathbb{M}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$, then $\mathscr{M}_1$ and $\mathscr{M}_2$ are elements of $\mathbb{M}(\mathscr{X})$, and so there are $p_1, p_2 \in \Sigma_{\mathscr{X}}$ such that $\mathscr{M}_1 = \{p_1\}$ and $\mathscr{M}_2 = \{p_2\}$, and $p_1 \neq p_2$ and $p = \lambda p_1 + (1-\lambda)p_2$.*

Proof. Assume *ex absurdo* that at least one of the two credal sets $\mathscr{M}_1$ and $\mathscr{M}_2$ is not an element of $\mathbb{M}(\mathscr{X})$, say $\mathscr{M}_1$. Consequently, $\mathscr{M}_1$ has at least two elements $q_a, q_b \in \Sigma_{\mathscr{X}}$, with $q_a \neq q_b$. If we denote by p_2 an arbitrary element of $\mathscr{M}_2$, then $q_1 := \lambda q_a + (1-\lambda)p_2$ and $q_2 := \lambda q_b + (1-\lambda)p_2$ by definition both belong to $\mathscr{M} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$. But since $\lambda > 0$ and $q_a \neq q_b$, we find that $q_1 \neq q_2$, contradicting $\mathscr{M} \in \mathbb{M}(\mathscr{X})$. Therefore, $\mathscr{M}_1$ and $\mathscr{M}_2$ are indeed both elements of $\mathbb{M}(\mathscr{X})$, meaning that there are $p_1, p_2 \in \Sigma_{\mathscr{X}}$ such that $\mathscr{M}_1 = \{p_1\}$ and $\mathscr{M}_2 = \{p_2\}$. It follows from $\mathscr{M}_1 \neq \mathscr{M}_2$ that $p_1 \neq p_2$ and from $\mathscr{M} = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ that $p = \lambda p_1 + (1-\lambda)p_2$. □

Proof of Corollary 15. Due to the one-to-one correspondence between credal sets in $\mathbb{M}(\mathscr{X})$ and linear previsions in $\mathbb{P}(\mathscr{X})$, this is a trivial consequence of Propositions 3 and 14 and Corollary 4. □

Proof of Proposition 16. Consider any fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ that can be written as a non-trivial convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, with $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\mathbb{M}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$. Suppose *ex absurdo* that at least one of the credal sets $\mathscr{M}_1$ or $\mathscr{M}_2$ is not an element of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$, say $\mathscr{M}_1 \notin \underline{\underline{\mathbb{M}}}(\mathscr{X})$. This implies that there is some $i \in \mathbb{N}_{\leq n}$ such that $\min\{p_1(x_i)\colon p_1 \in \mathscr{M}_1\} > 0$. Consequently, we find that

$$\begin{aligned}\min\{p(x_i)\colon p \in \mathscr{M}\} &= \min\{\lambda p_1(x_i) + (1-\lambda)p_2(x_i)\colon p_1 \in \mathscr{M}_1 \text{ and } p_2 \in \mathscr{M}_2\}\\ &= \lambda \min\{p_1(x_i)\colon p_1 \in \mathscr{M}_1\} + (1-\lambda)\min\{p_2(x_i)\colon p_2 \in \mathscr{M}_2\} > 0,\end{aligned}$$

contradicting that $\mathscr{M}$ is fully imprecise. □

Proof of Corollary 17. Consider any fully imprecise lower prevision $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$ that can be written as a non-trivial convex combination $\lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$, with $\underline{P}_1, \underline{P}_2 \in \underline{\mathbb{P}}(\mathscr{X})$, $\underline{P}_1 \neq \underline{P}_2$ and $0 < \lambda < 1$. Due to Proposition 3 and the one-to-one correspondence between credal sets and coherent lower previsions, this implies that $\mathscr{M}_{\underline{P}} = \lambda\mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}$, with $\mathscr{M}_{\underline{P}_1} \neq \mathscr{M}_{\underline{P}_2}$. Since $\mathscr{M}_{\underline{P}} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ by definition, we can apply

Proposition 16 to find that $\mathscr{M}_{\underline{P}_1}$ and $\mathscr{M}_{\underline{P}_2}$ are both fully imprecise. Therefore, their corresponding lower previsions $\underline{P}_1$ and $\underline{P}_2$ are also fully imprecise and thus elements of $\underline{\underline{\mathbb{P}}}(\mathscr{X})$. □

Proof of Theorem 18. Consider any fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. To prove the direct implication, suppose that $\mathscr{M}$ can be written as a non-trivial convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, with $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$. We now need to prove that the projected credal set $K_{\mathscr{M}}$ is Minkowski decomposable. To do so, we first apply Proposition 16 to find that, besides $\mathscr{M}$, $\mathscr{M}_1$ and $\mathscr{M}_2$ are also elements of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$. Therefore, the projected credal sets $K_{\mathscr{M}}$, $K_{\mathscr{M}_1}$ and $K_{\mathscr{M}_2}$ are elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$. It follows from $\mathscr{M}_1 \neq \mathscr{M}_2$ that $K_{\mathscr{M}_1} \neq K_{\mathscr{M}_2}$ and since $K_{\mathscr{M}} \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, it follows by definition that $\underline{\underline{K}}_{\mathscr{M}} = K_{\mathscr{M}}$. Due to Proposition 6, we have that $K_{\mathscr{M}} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}$ and therefore that $\underline{\underline{K}}_{\mathscr{M}} = \lambda K_{\mathscr{M}_1} + (1-\lambda)K_{\mathscr{M}_2}$, which by Theorem 10 means that $K_{\mathscr{M}}$ is Minkowski decomposable, since the non-empty $K_{\mathscr{M}}$ is compact and convex by definition and not a singleton because $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ and $\mathbb{K}(\mathscr{X})$ are disjoint due to Corollary 7.

To prove the converse implication, suppose that $K_{\mathscr{M}}$ is Minkowski decomposable. We now need to show that $\mathscr{M}$ can be written as a non-trivial convex combination $\lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$, with $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$. Since $K_{\mathscr{M}} \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, it is by definition not a singleton, compact and convex and, furthermore, $\underline{\underline{K}}_{\mathscr{M}} = K_{\mathscr{M}}$. Therefore, it follows from Theorem 10 that $K_{\mathscr{M}}$ can be written as a non-trivial convex combination $\lambda K_1 + (1-\lambda)K_2$, with K_1 and K_2 both elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$, $K_1 \neq K_2$ and $0 < \lambda < 1$. Due to Proposition 6 and the one-to-one correspondence between credal sets and projected credal sets this implies that $\mathscr{M} = \lambda\mathscr{M}_{K_1} + (1-\lambda)\mathscr{M}_{K_2}$, with $\mathscr{M}_{K_1} \neq \mathscr{M}_{K_2}$. □

Proof of Corollary 19. Consider any fully imprecise lower prevision $\underline{P} \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$. Due to Proposition 3 and the one-to-one correspondence between credal sets and coherent lower previsions, $\underline{P}$ can be written as a non-trivial convex combination $\lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$, with $\underline{P}_1, \underline{P}_2 \in \underline{\underline{\mathbb{P}}}(\mathscr{X})$, $\underline{P}_1 \neq \underline{P}_2$ and $0 < \lambda < 1$ if and only if $\mathscr{M}_{\underline{P}} = \lambda\mathscr{M}_{\underline{P}_1} + (1-\lambda)\mathscr{M}_{\underline{P}_2}$, with $\mathscr{M}_{\underline{P}_1} \neq \mathscr{M}_{\underline{P}_2}$. Due to Theorem 18, this in turn is the case if and only if $K_{\underline{P}} := K_{\mathscr{M}_{\underline{P}}}$ is Minkowski decomposable, which concludes the proof. □

Proof of Proposition 21. Assume *ex absurdo* that the vacuous credal set is not extreme, meaning that $\mathscr{M}_V = \lambda\mathscr{M}_1 + (1-\lambda)\mathscr{M}_2$ for some $\mathscr{M}_1, \mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$, $\mathscr{M}_1 \neq \mathscr{M}_2$ and $0 < \lambda < 1$. It now suffices to show for all $i \in \mathbb{N}_{\leq n}$ that $p_i^\circ \in \mathscr{M}_1$ and $p_i^\circ \in \mathscr{M}_2$. Using the convexity of $\mathscr{M}_1$ and $\mathscr{M}_2$ and the second part of Proposition 14, we will then find that $\mathscr{M}_1 = \mathscr{M}_V = \mathscr{M}_2$, a contradiction.

So fix any $i \in \mathbb{N}_{\leq n}$. Since $p_i^\circ \in \mathscr{M}_V$, we can infer from our assumption that $p_i^\circ = \lambda p_1 + (1-\lambda)p_2$ for some $p_1 \in \mathscr{M}_1$ and $p_2 \in \mathscr{M}_2$. By definition of p_i°, we have $p_i^\circ(x_i) = 1$ and thus $\lambda p_1(x_i) + (1-\lambda)p_2(x_i) = 1$. Since, $0 < \lambda < 1$ and $p_1, p_2 \in \Sigma_{\mathscr{X}}$, the only way to achieve this is if $p_1 = p_2 = p_i^\circ$, implying that p_i° is indeed an element of both $\mathscr{M}_1$ and $\mathscr{M}_2$. □

Proof of Corollary 22. Trivially from Proposition 21 and Corollary 4. □

Proof of Corollary 24. Fix a ternary possibility space $\mathscr{X} = \{x_1, x_2, x_3\}$ and a fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. We know from Corollary 20 that $\mathscr{M}$ is extreme if and only if its projected credal set $K_{\mathscr{M}}$ is Minkowski indecomposable, or by Theorem 23, if $K_{\mathscr{M}}$ is a triangle or a line segment. Since line segments are degenerate triangles, we are left to prove that $K_{\mathscr{M}}$ is a triangle if and only if we can find $p_1, p_2, p_3 \in \Sigma_{\mathscr{X}}$ such that

$$\mathscr{M} = \left\{ \sum_{i=1}^{3} \lambda_i p_i \colon (\lambda_1, \lambda_2, \lambda_3) \in \Sigma_{\mathscr{X}} \right\},$$

which is trivial because projecting credal sets on $\mathbf{K}_{\mathscr{X}}$ preserves convex combinations; see for example Proposition 6. □

Proof of Corollary 25. Fix a ternary possibility space $\mathscr{X} = \{x_1, x_2, x_3\}$ and a fully imprecise lower prevision $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$. We know from Corollary 4 that $\underline{P}$ is extreme if and only if its corresponding credal set $\mathscr{M}_{\underline{P}}$ is, or by Corollary 24, if we can find $p_1, p_2, p_3 \in \Sigma_{\mathscr{X}}$ such that $\mathscr{M}_{\underline{P}} = \{\sum_{i=1}^{3} \lambda_i p_i\colon (\lambda_1, \lambda_2, \lambda_3) \in \Sigma_{\mathscr{X}}\}$. By letting P_1, P_2 and P_3 be the linear previsions that correspond to the probability mass functions p_1, p_2 and p_3 respectively, we infer from Eq. (1) that for all $f \in \mathscr{G}(\mathscr{X})$

$$\underline{P}(f) = \min\{P_p(f)\colon p \in \mathscr{M}_{\underline{P}}\} = \min\left\{\sum_{i=1}^{3} \lambda_i P_i(f)\colon (\lambda_1, \lambda_2, \lambda_3) \in \Sigma_{\mathscr{X}}\right\} = \min_{i \in \mathbb{N}_{\leq 3}} P_i(f). \qquad \square$$

Proof of Corollary 27. Consider any $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ that is the convex hull of m affinely independent mass functions p_i, $i = 1, \ldots, m$. Note that this is only possible if $m \leq n$. Then clearly, $K_{\mathscr{M}}$ is the convex hull of the points v_{p_i}, $i = 1, \ldots, m$, which are affinely independent as well. Hence, $K_{\mathscr{M}}$ is a simplex in $\mathbb{R}^{n-1}$ and therefore, due to Theorem 26, Minkowski indecomposable. By applying Corollary 20, we find that $\mathscr{M}$ is an extreme credal set. $\square$

Proof of Corollary 28. Consider any $\underline{P} \in \underline{\mathbb{P}}(\mathscr{X})$ that is the lower envelope of m affinely independent linear previsions P_i, $i = 1, \ldots, m$. Let $p_i \in \Sigma_{\mathscr{X}}$, $i = 1, \ldots, m$, be the unique mass functions for which $P_i = P_{p_i}$. Then $\mathscr{M}_{\underline{P}}$ is the convex hull of the mass functions p_i, $i = 1, \ldots, m$, which are affinely independent because the linear previsions P_i are, implying that $m \leq n$. Due to Corollary 27, $\mathscr{M}_{\underline{P}}$ is extreme and therefore, by applying Corollary 4, we find that $\underline{P}$ is extreme as well. $\square$

Proof of Proposition 30. Since $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a metric space, it suffices to show that for every $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$, there is a sequence of decomposable credal sets $\mathscr{M}_i$ that converges to $\mathscr{M}$. Choose $0 < \varepsilon < 1$ and $\mathscr{M}^* \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ such that $\mathscr{M}^* \neq \mathscr{M}$ and let, for all $i \in \mathbb{N}$, $\mathscr{M}_i := \varepsilon^i \mathscr{M}^* + (1 - \varepsilon^i)\mathscr{M}$, so the credal set $\mathscr{M}_i$ is clearly decomposable. We then find that

$$\begin{aligned}
\max_{p \in \mathscr{M}} \min_{p_i \in \mathscr{M}_i} d(p, p_i) = \max_{p \in \mathscr{M}} \min_{\substack{p^* \in \mathscr{M}^* \\ p' \in \mathscr{M}}} d\big(p, \varepsilon^i p^* + \big(1 - \varepsilon^i\big)p'\big) &\leq \max_{p \in \mathscr{M}} \min_{p^* \in \mathscr{M}^*} d\big(p, \varepsilon^i p^* + \big(1 - \varepsilon^i\big)p\big) \\
&= \max_{p \in \mathscr{M}} \min_{p^* \in \mathscr{M}^*} \varepsilon^i d\big(p, p^*\big) \\
&= \varepsilon^i \max_{p \in \mathscr{M}} \min_{p^* \in \mathscr{M}^*} d\big(p, p^*\big) \leq \varepsilon^i,
\end{aligned}$$

and

$$\begin{aligned}
\max_{p_i \in \mathscr{M}_i} \min_{p \in \mathscr{M}} d(p, p_i) = \max_{\substack{p^* \in \mathscr{M}^* \\ p' \in \mathscr{M}}} \min_{p \in \mathscr{M}} d\big(p, \varepsilon^i p^* + \big(1 - \varepsilon^i\big)p'\big) &\leq \max_{\substack{p^* \in \mathscr{M}^* \\ p' \in \mathscr{M}}} d\big(p', \varepsilon^i p^* + \big(1 - \varepsilon^i\big)p'\big) \\
&= \max_{\substack{p^* \in \mathscr{M}^* \\ p' \in \mathscr{M}}} \varepsilon^i d\big(p', p^*\big) = \varepsilon^i \max_{\substack{p^* \in \mathscr{M}^* \\ p' \in \mathscr{M}}} d\big(p', p^*\big) \leq \varepsilon^i.
\end{aligned}$$

Hence, $d(\mathscr{M}, \mathscr{M}_i) \leq \varepsilon^i$, implying that $\mathscr{M}_i$ indeed converges to $\mathscr{M}$. $\square$

Proof of Corollary 31. This is a special case of Proposition 49. $\square$

Proof of Corollary 32. Follows directly from Proposition 30 and Corollaries 4 and 31. $\square$

Proof of Theorem 33. Since $n = 3$, and due to Corollary 24, $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ can be identified with $\underline{\underline{\mathbb{M}}}(\mathscr{X}) \cap \mathscr{K}_3^3$, where $\mathscr{K}_3^3$ is the set consisting of all polytopes in $\mathscr{C}^3$ that have at most 3 vertices. We know from Appendix A that $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ is closed and, by applying Lemma 55 with $k = m = 3$, we find that $\mathscr{K}_3^3$ is closed, so $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$

is closed as well. Therefore, $\underline{\underline{\mathbb{M}}}(\mathscr{X}) \setminus \operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is an open subset of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ that, by Proposition 30, is dense in $\underline{\underline{\mathbb{M}}}(\mathscr{X})$. By definition, this means that $\operatorname{ext} \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is a nowhere dense closed subset of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$. $\square$

Lemma 55. *Consider any $m \in \mathbb{N}$, then the set $\mathscr{K}_m^k$ of all polytopes in $\mathscr{C}^k$ that have at most m vertices is a closed subset of $\mathscr{C}^k$.*

Proof. Consider any sequence of polytopes K_i, $i \in \mathbb{N}$, each of which is an element of $\mathscr{K}_m^k$, such that this sequence converges to some element of $\mathscr{C}^k$, say C. We will prove that C is an element of $\mathscr{K}_m^k$ as well.

For any $i \in \mathbb{N}$, since K_i has at most m vertices, it is possible to find m points $v_{i,1}, \dots, v_{i,m}$ in $\mathbb{R}^k$ such that K_i is their convex hull. Since the sequence K_i, $i \in \mathbb{N}$, converges to the compact (and therefore bounded) set C, the sequence $v_{i,1}$, $i \in \mathbb{N}$, is bounded, which implies that we can apply the Bolzano–Weierstraß theorem to find a convergent subsequence $v_{i_r,1}$, $r \in \mathbb{N}$, that converges to some $v_1 \in \mathbb{R}^k$. Now let K_{i_r}, $r \in \mathbb{N}$, be the corresponding subsequence of polytopes. Then since K_i, $i \in \mathbb{N}$, converges to C, we have that K_{i_r}, $r \in \mathbb{N}$, converges to C as well. Hence, we may assume without loss of generality that the sequence $v_{i,1}$, $i \in \mathbb{N}$, converges to v_1 (simply replace the original sequence K_i, $i \in \mathbb{N}$, by the subsequence K_{i_r}, $r \in \mathbb{N}$). Next, we consider the sequence $v_{i,2}$, $i \in \mathbb{N}$. Again, in much the same way, we find that there is a convergent subsequence $v_{i_\ell,2}$, $\ell \in \mathbb{N}$, that converges to some $v_2 \in \mathbb{R}^k$. The corresponding subsequences K_{i_ℓ}, $\ell \in \mathbb{N}$, and $v_{i_\ell,1}$, $\ell \in \mathbb{N}$, converge to C and v_1, respectively. Hence, we may assume without loss of generality that the sequences $v_{i,1}$, $i \in \mathbb{N}$, and $v_{i,2}$, $i \in \mathbb{N}$, converge to v_1 and v_2, respectively (simply replace the sequence K_i, $i \in \mathbb{N}$, by the subsequence K_{i_ℓ}, $\ell \in \mathbb{N}$). By repeating this argument for every $j \in \{1, \dots, m\}$, we find that we can assume without loss of generality that, for all $j \in \{1, \dots, m\}$, the sequence $v_{i,j}$, $i \in \mathbb{N}$, converges to some $v_j \in \mathbb{R}^k$.

Now let K be the convex hull of the points v_j, $j \in \{1, \dots, m\}$. Then since, for all $j \in \{1, \dots, m\}$, the sequence $v_{i,j}$, $i \in \mathbb{N}$, converges to v_j, it is not hard to infer that the sequence K_i, $i \in \mathbb{N}$, converges to K or, equivalently, that $C = K$. Indeed, if we denote by Σ_m the simplex of all probability mass functions $(\lambda_1, \dots, \lambda_m)$ on $\{1, \dots, m\}$, we get for any $i \in \mathbb{N}$ that

$$\begin{aligned}
\max_{v \in K} \min_{v' \in K_i} \delta\big(v, v'\big) &= \max_{(\lambda_1,\dots,\lambda_m) \in \Sigma_m} \min_{(\lambda'_1,\dots,\lambda'_m) \in \Sigma_m} \delta\left(\sum_{j=1}^m \lambda_j v_j, \sum_{j=1}^m \lambda'_j v_{i,j}\right) \\
&\leq \max_{(\lambda_1,\dots,\lambda_m) \in \Sigma_m} \delta\left(\sum_{j=1}^m \lambda_j v_j, \sum_{j=1}^m \lambda_j v_{i,j}\right) \\
&\leq \max_{(\lambda_1,\dots,\lambda_m) \in \Sigma_m} \sum_{j=1}^m \lambda_j \delta(v_j, v_{i,j}) \leq \max_{j=1}^m \delta(v_j, v_{i,j}),
\end{aligned}$$

and similarly, we also find that $\max_{v' \in K_i} \min_{v \in K} \delta(v, v') \leq \max_{j=1}^m \delta(v_j, v_{i,j})$, leading us to conclude that $\delta(K, K_i) \leq \max_{j=1}^m \delta(v_j, v_{i,j}) \to 0$. Hence, since K is clearly an element of $\mathscr{K}_m^k$, so is C. $\square$

Proof of Corollary 34. Follows directly from Theorem 33 and Corollaries 4 and 31. $\square$

Proof of Theorem 35. Immediate consequence of Lemma 56, Corollary 20 and Proposition 48. $\square$

Lemma 56. *For $n \geq 4$, the Minkowski indecomposable elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ constitute a dense G_δ subset.*

Proof. The starting point for this proof is that for $k \geq 3$, the Minkowski indecomposable non-empty convex compact subsets of $\mathbb{R}^k$ are a dense G_δ subset of $\mathscr{C}^k$ [26, Theorem 3.2.14]. For ease of reference, we denote this subset by $\operatorname{ind} \mathscr{C}^k$. Since for $n \geq 4$ and $k = n - 1$, the set of all Minkowski indecomposable elements of

$\underline{\underline{\mathbb{K}}}(\mathscr{X})$ is equal to $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$, we are left to prove that, for $k \geq 3$ and $n = k+1$, $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ is a dense G_δ subset of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$.

We first prove that $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ is G_δ. Since $\operatorname{ind}\mathscr{C}^k$ is a G_δ subset of $\mathscr{C}^k$, it can be written as a countable intersection $\bigcap_{i=1}^{\infty}\mathscr{C}_i^k$ of open subsets $\mathscr{C}_i^k$ of $\mathscr{C}^k$, and therefore $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k = \underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \bigcap_{i=1}^{\infty}\mathscr{C}_i^k = \bigcap_{i=1}^{\infty}[\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \mathscr{C}_i^k]$ is a countable intersection of open subsets $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \mathscr{C}_i^k$ of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$. Hence $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ is a G_δ subset of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$.

Next, we prove that $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ is dense in $\underline{\underline{\mathbb{K}}}(\mathscr{X})$. Consider any $C \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$ and therefore also $C \in \mathscr{C}^k$. Then we have to prove that there is some sequence of elements of $\underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ that converges to C. Since $\operatorname{ind}\mathscr{C}^k$ is a dense subset of $\mathscr{C}^k$, there is a sequence of $C_i \in \operatorname{ind}\mathscr{C}^k$ that converges to C. Since C is not a singleton, we can assume without loss of generality that, for all $i \in \mathbb{N}$, C_i is not a singleton. Hence, after applying Lemma 57, we find that the sequence $\underline{\underline{C}}_i$ converges to C as well. This concludes the proof since, for all $i \in \mathbb{N}$, $C_i \in \underline{\underline{\mathbb{K}}}(\mathscr{X}) \cap \operatorname{ind}\mathscr{C}^k$ because of Propositions 8 and 9. □

Lemma 57. *Consider any $K \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, with $n = k+1$, and any sequence of non-singleton $C_i \in \mathscr{C}^k$ that converges to K. Then $\underline{\underline{C}}_i$ converges to K as well.*

Proof. By assumption, $\lim_{i\to\infty}\delta(K, C_i) = 0$. Since $m(\cdot)$ and $\mu(\cdot)$ are continuous operators, and since $m(K) = 0$ and $\mu(K) = 1$, this implies that $\lim_{i\to\infty} m(C_i) = 0$ and $\lim_{i\to\infty}\mu(C_i) = 1$. Also, for all $i \in \mathbb{N}$, $\mu(C_i) > 0$ because by assumption, C_i is not a singleton. Hence, we can define $\lambda_i := {}^1/_{\mu(C_i)} > 0$ for all $i \in \mathbb{N}$. Clearly also $\lim_{i\to\infty}\lambda_i = 1$. By definition of $\underline{\underline{C}}_i$ and δ, and also recalling that δ is a metric, we find that

$$\begin{aligned}\delta(K, \underline{\underline{C}}_i) = \delta\big(K, \lambda_i\big[C_i - m(C_i)\big]\big) &\leq \delta(K, \lambda_i K) + \delta\big(\lambda_i K, \lambda_i\big[C_i - m(C_i)\big]\big)\\ &\leq (1-\lambda_i)\max_{v\in K}\|v\| + \lambda_i\delta\big(K, C_i - m(C_i)\big)\\ &\leq (1-\lambda_i)\max_{v\in K}\|v\| + \lambda_i\delta(K, C_i) + \lambda_i\delta\big(C_i, C_i - m(C_i)\big)\\ &\leq (1-\lambda_i)\max_{v\in K}\|v\| + \lambda_i\delta(K, C_i) + \lambda_i\big\|m(C_i)\big\|,\end{aligned}$$

implying that $\lim_{i\to\infty}\delta(K, \underline{\underline{C}}_i) = 0$, so the sequence $\underline{\underline{C}}_i$ converges to K. □

Proof of Corollary 36. Follows directly from Theorem 35 and Corollaries 4 and 31. □

Proof of Proposition 37. Consider any finitely generated credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Then due to Proposition 11, it can be written as a (possibly degenerate) convex combination of a credal set $\mathscr{M}_1 \in \mathbb{M}(\mathscr{X})$ that is a singleton and a fully imprecise credal set $\mathscr{M}_2 \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Clearly, $\mathscr{M}_1$ and $\mathscr{M}_2$ are both finitely generated. By Proposition 14 and Lemma 58, $\mathscr{M}_1$ and $\mathscr{M}_2$ can both be written as a finite convex combination of finitely generated extreme credal sets. Hence, $\mathscr{M}$ can be written as a finite convex combination of finitely generated extreme credal sets. □

Lemma 58. *Every finitely generated fully imprecise credal set can be written as a finite convex combination of finitely generated fully imprecise extreme credal sets.*

Proof. The starting point for this proof is that every polytope $K \in \mathscr{C}^k$ that is not a singleton can be written as a finite sum of Minkowski indecomposable polytopes $K_i \in \mathscr{C}^k$, $i \in \{1, \ldots, r\}$ (each of which is by definition not a singleton) [18, Theorem 4]. Consider now any finitely generated fully imprecise credal set $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Then for $k = n-1$, $K_{\mathscr{M}}$ is a polytope in $\mathscr{C}^k$ that is not a singleton. By combining the result mentioned in the beginning of this proof with Lemma 59, we find that $K_{\mathscr{M}}$ can be written as a finite convex combination of $\underline{\underline{K}}_i$, $i \in \{1, \ldots, r\}$, in which every K_i is a Minkowski indecomposable polytope and therefore, by Propositions 8 and 9 and Eq. (6), every $\underline{\underline{K}}_i$ is a Minkowski indecomposable polytope in $\underline{\underline{\mathbb{K}}}(\mathscr{X})$. Applying Proposition 6 multiple times, we find that $\mathscr{M} = \mathscr{M}_{K_{\mathscr{M}}}$ can be written as a finite convex combination of the credal sets $\mathscr{M}_{\underline{\underline{K}}_i}$. This concludes the proof since, for all $i \in \{1, \dots, r\}$, the fact that $\underline{\underline{K}}_i$ is an element of $\underline{\underline{\mathbb{K}}}(\mathscr{X})$ that is furthermore a Minkowski indecomposable polytope implies that $\mathscr{M}_{\underline{\underline{K}}_i}$ is an element of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ that is furthermore extreme (Corollary 19) and finitely generated. □

Lemma 59. *Fix $k \in \mathbb{N}$ and consider any $C \in \mathscr{C}^k$ that is not a singleton. If C can be written as a finite Minkowski sum of $C_i \in \mathscr{C}^k$, $i \in \{1, \dots, r\}$, each of which is not a singleton, then $\underline{\underline{C}}$ can be written as a finite convex combination of $\underline{\underline{C}}_i$, $i \in \{1, \dots, r\}$.*

Proof. Since $C = \sum_{i=1}^r C_i$, we find that $m(C) = \sum_{i=1}^r m(C_i)$ and $\mu(C) = \sum_{i=1}^r \mu(C_i)$. Furthermore, for all $i \in \{1, \dots, r\}$, $\mu(C_i) > 0$ because C_i is by assumption not a singleton. Let $\lambda_i := {}^{\mu(C_i)}/_{\mu(C)} > 0$ for all $i \in \{1, \dots, r\}$, then $\sum_{i=1}^r \lambda_i = 1$. Hence, $\sum_{i=1}^r \lambda_i \underline{\underline{C}}_i$ is a convex combination of the $\underline{\underline{C}}_i$, $i \in \{1, \dots, r\}$. This concludes the proof since

$$\sum_{i=1}^r \lambda_i \underline{\underline{C}}_i = \frac{1}{\mu(C)} \sum_{i=1}^r [C_i - m(C_i)] = \frac{1}{\mu(C)} \left(\sum_{i=1}^r C_i - \sum_{i=1}^r m(C_i) \right) = \frac{1}{\mu(C)} [C - m(C)] = \underline{\underline{C}}. \qquad \square$$

Proof of Corollary 38. Since finitely generated credal sets correspond to finitely generated lower previsions, and extreme credal sets to extreme lower previsions (see Corollary 4), this follows rather directly from Proposition 37 by applying Proposition 3 multiple times. □

Proof of Theorem 39. The first part of this theorem is an immediate consequence of Lemmas 60 and 58. The second part follows from Lemma 61. □

Lemma 60. *Every fully imprecise credal set can be approximated arbitrarily closely by a finitely generated fully imprecise credal set.*

Proof. Consider any $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Then for $k = n-1$, $K_{\mathscr{M}} \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$ is not a singleton. Due to, for instance, Ref. [4, Chapter IV, Theorem 2.8(d)], there is a sequence of non-singleton polytopes $C_i \in \mathscr{C}^k$, $i \in \mathbb{N}$, that converges to $K_{\mathscr{M}}$. By applying Lemma 57, we find that the sequence $\underline{\underline{C}}_i \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, $i \in \mathbb{N}$, converges to $K_{\mathscr{M}}$ as well. Hence, by Proposition 48, the sequence $\mathscr{M}_{\underline{\underline{C}}_i}$, $i \in \mathbb{N}$ converges to $\mathscr{M}_{K_{\mathscr{M}}} = \mathscr{M}$. This concludes the proof since, for all $i \in \mathbb{N}$, $\underline{\underline{C}}_i$ is a polytope because C_i is, and therefore $\mathscr{M}_{\underline{\underline{C}}_i}$ is finitely generated. □

Lemma 61. *For $n \geq 4$, every fully imprecise credal set can be approximated arbitrarily closely by a finitely generated fully imprecise extreme credal set.*

Proof. Consider any $\mathscr{M} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Then for $k = n - 1$, $K_{\mathscr{M}} \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$ is not a singleton. Due to, for instance, Ref. [4, Chapter IV, Theorem 2.8(a)], there is a sequence of *simplicial* polytopes $C_i \in \mathscr{C}^k$, $i \in \mathbb{N}$, that converges to $K_{\mathscr{M}}$. Clearly, without loss of generalisation, we can assume that none of these C_i are singletons (because $K_{\mathscr{M}}$ is not). By applying Lemma 57, we find that the sequence $\underline{\underline{C}}_i \in \underline{\underline{\mathbb{K}}}(\mathscr{X})$, $i \in \mathbb{N}$, converges to $K_{\mathscr{M}}$ as well. Hence, by Proposition 48, the sequence $\mathscr{M}_{\underline{\underline{C}}_i}$, $i \in \mathbb{N}$ converges to $\mathscr{M}_{K_{\mathscr{M}}} = \mathscr{M}$. This concludes the proof because, for all $i \in \mathbb{N}$, since C_i is a simplicial polytope, $\underline{\underline{C}}_i$ is also a simplicial and therefore Minkowski indecomposable (because $k \geq 3$; see the text below Corollary 28 in Section 6.3) polytope, implying that $\mathscr{M}_{\underline{\underline{C}}_i} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$ is extreme (by Corollary 20) and finitely generated. □

Proof of Corollary 40. Immediate consequence of Theorem 39, Proposition 3 and Corollaries 4 and 31. □

Proof of Proposition 41. Consider any subset $\mathscr{A}$ of $\underline{\underline{\mathbb{M}}}(\mathscr{X})$ that consists of finitely generated credal sets and for which there exists some $m \in \mathbb{N}$ such that every $\mathscr{M} \in \mathscr{A}$ is the convex hull of at most m mass functions. Assume *ex absurdo* that $\mathscr{A}$ is a universal approximating class for $\underline{\underline{\mathbb{M}}}(\mathscr{X})$.

Now let $k = n - 1$, implying that $k \geq 3$, and let C be any Minkowski indecomposable polytope in $\mathscr{C}^k$ that has more than m vertices. It is always possible to construct such a polytope: for any $v > k$, one can for example construct a k-dimensional cyclic polytope with v vertices, which is always simplicial—see for example Ref. [8, p. 61–62]—and therefore, since $k \geq 3$, Minkowski indecomposable because of Theorem 26. By Propositions 8 and 9, $\underline{\underline{C}}$ is a Minkowski indecomposable polytope in $\underline{\underline{\mathbb{K}}}(\mathscr{X})$. Clearly, since C has more than m vertices, $\underline{\underline{C}}$ has more than m vertices as well.

Now consider the credal set $\mathscr{M}_{\underline{\underline{C}}} \in \underline{\underline{\mathbb{M}}}(\mathscr{X})$. Since by assumption, $\mathscr{A}$ is a universal approximating class for $\underline{\underline{\mathbb{M}}}(\mathscr{X})$, there is a sequence of credal sets $\mathscr{M}_i$, $i \in \mathbb{N}$, that converges to $\mathscr{M}_{\underline{\underline{C}}}$ such that, for all $i \in \mathbb{N}$, $\mathscr{M}_i$ is a finite convex combination of credal sets in $\mathscr{A}$. By Proposition 48, the corresponding sequence $K_{\mathscr{M}_i}$, $i \in \mathbb{N}$, converges to $\underline{\underline{C}}$. By Proposition 6, each element $K_{\mathscr{M}_i}$ of this sequence is a finite convex combination of elements of $\mathscr{K}_m^k$, which is the set consisting of all polytopes in $\mathscr{C}^k$ that have at most m vertices. Since $\underline{\underline{C}}$ is Minkowski indecomposable, this implies that the closure of $\mathscr{K}_m^k$ contains a homothetic copy C' of $\underline{\underline{C}}$ [26, Theorem 3.3.3]. By combining this with Lemma 55, we find that $C' \in \mathscr{K}_m^k$. However, since $\underline{\underline{C}}$ has more than m vertices, C' has more than m vertices as well, a contradiction. □

Proof of Corollary 42. Since by Proposition 3, convex combinations are preserved when going from credal sets to coherent lower previsions, this follows easily from Proposition 41 and Corollary 31. □

Proof of Proposition 43. We provide a proof by contraposition. Assume that $\underline{E}$ is not extreme or, equivalently, that we can find $\underline{P}_1^*, \underline{P}_2^* \in \underline{\mathbb{P}}(\mathscr{X})$ such that $\underline{P}_1^* \neq \underline{P}_2^*$ and $\lambda\underline{P}_1^* + (1-\lambda)\underline{P}_2^* = \underline{E}$, with $0 < \lambda < 1$. Let $\underline{P}_1$ and $\underline{P}_2$ be the restrictions to $\mathscr{F}$ of $\underline{P}_1^*$ and $\underline{P}_2^*$, respectively. Then $\underline{P}_1$ and $\underline{P}_2$ are coherent. Furthermore, since $\underline{P}$ is coherent, $\underline{E}$ coincides with $\underline{P}$ on $\mathscr{F}$ and therefore, we have that $\underline{P} = \lambda\underline{P}_1 + (1-\lambda)\underline{P}_2$. In order to prove that $\underline{P}$ is not extreme, it suffices to show that $\underline{P}_1 \neq \underline{P}_2$.

Now let $\underline{E}_1$ and $\underline{E}_2$ be the natural extensions of $\underline{P}_1$ and $\underline{P}_2$ to $\mathscr{G}(\mathscr{X})$, respectively. Then since $\underline{P}_1$ and $\underline{P}_2$ are coherent, $\underline{E}_1$ and $\underline{E}_2$ are by definition the pointwise smallest coherent lower previsions on $\mathscr{G}(\mathscr{X})$ that coincide with $\underline{P}_1$ and $\underline{P}_2$ on $\mathscr{F}$, respectively. Similarly, $\underline{E}$ is the pointwise smallest coherent lower prevision on $\mathscr{G}(\mathscr{X})$ that coincides with $\underline{P}$ on $\mathscr{F}$. Assume *ex absurdo* that $\underline{P}_1^* \neq \underline{E}_1$. Then we infer that, for at least one $f_1 \in \mathscr{G}(\mathscr{X})$, $\underline{E}_1(f_1) < \underline{P}_1^*(f_1)$. Consider now $\underline{P}^* := \lambda\underline{E}_1 + (1-\lambda)\underline{P}_2^*$. Then $\underline{P}^*$ is a coherent lower prevision on $\mathscr{G}(\mathscr{X})$. Furthermore, since $\underline{E}_1$ and $\underline{P}_1^*$ both coincide with $\underline{P}_1$ on $\mathscr{F}$, we find that $\underline{P}^*$ and $\underline{E}$ coincide with $\underline{P}$ on $\mathscr{F}$. Therefore, by definition of $\underline{E}$, for all $f \in \mathscr{G}(\mathscr{X})$, $\underline{E}(f) \leq \underline{P}^*(f)$. However, we also have that

$$\underline{P}^*(f_1) = \lambda\underline{E}_1(f_1) + (1-\lambda)\underline{P}_2^*(f_1) < \lambda\underline{P}_1^*(f_1) + (1-\lambda)\underline{P}_2^*(f_1) = \underline{E}(f_1),$$

a contradiction. Hence, we infer that $\underline{P}_1^* = \underline{E}_1$. Similarly, by symmetry, we find that $\underline{P}_2^* = \underline{E}_2$. Assume *ex absurdo* that $\underline{P}_1 = \underline{P}_2$. Then $\underline{E}_1 = \underline{E}_2$ and therefore also $\underline{P}_1^* = \underline{P}_2^*$, a contradiction. Hence, we find that $\underline{P}_1 \neq \underline{P}_2$ and therefore that $\underline{P}$ is an extreme coherent lower prevision on $\mathscr{F}$. □

References

[1] T. Augustin, F.P.A. Coolen, G. de Cooman, M.C.M. Troffaes (Eds.), Introduction to Imprecise Probabilities, John Wiley & Sons, 2014.

[2] F.G. Cozman, Robustness analysis of Bayesian networks with local convex sets of distributions, in: Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence, 1997, pp. 108–115.

[3] J. De Bock, G. de Cooman, Extreme lower previsions and Minkowski indecomposability, in: Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 2013, pp. 157–168.

[4] G. Ewald, Combinatorial Convexity and Algebraic Geometry, Springer, New York, 1996.

[5] D. Gale, Irreducible convex sets, in: Proceedings of the International Congress of Mathematicians, vol. 2, 1954, pp. 217–218.

[6] M.C. Gemignani, Elementary Topology, second edition, Dover Publications, New York, 1990.
[7] P.K. Ghosh, K. Deguchi, Mathematics of Shape Description: A Morphological Approach to Image Processing and Computer Graphics, Wiley, Singapore, 2008.
[8] B. Grünbaum, Convex Polytopes, second edition, Springer, 2003.
[9] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, second edition, Cambridge University Press, 1952.
[10] P.J. Huber, Robust Statistics, John Wiley & Sons, New York, 1981.
[11] M. Kallay, Indecomposable polytopes, Israel J. Math. 41 (1982) 235–243.
[12] S. Maaß, Continuous linear representations of coherent lower previsions, in: Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, 2003, pp. 372–382.
[13] S. Maaß, Exact functionals, functionals preserving linear inequalities, Lévy's metric, PhD thesis, University of Bremen, 2003.
[14] P. McMullen, Representation of polytopes and polyhedral sets, Geom. Dedicata 2 (1973) 83–99.
[15] P. McMullen, Indecomposable convex polytopes, Israel J. Math. 58 (1987) 321–323.
[16] W.J. Meyer, Minkowski addition of convex sets, PhD thesis, University of Wisconsin, Madison, 1969.
[17] W.J. Meyer, Decomposing plane convex bodies, Arch. Math. 23 (1) (1972) 534–536.
[18] W.J. Meyer, Indecomposable polytopes, Trans. Amer. Math. Soc. 190 (1974) 77–86.
[19] E. Miranda, A survey of the theory of coherent lower previsions, Internat. J. Approx. Reason. 48 (2) (2008) 628–658.
[20] J.C. Oxtoby, Measure and Category: A Survey of the Analogies Between Topological and Measure Spaces, Springer-Verlag, New York, 1980.
[21] K. Przeslawski, D. Yost, Decomposability of polytopes, Discrete Comput. Geom. 39 (1–3) (2008) 460–468.
[22] E. Quaeghebeur, Characterizing the set of coherent lower previsions with a finite number of constraints or vertices, in: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, 2010, pp. 466–473.
[23] E. Quaeghebeur, G. de Cooman, Extreme lower probabilities, Fuzzy Sets and Systems 159 (16) (2008) 2163–2175.
[24] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[25] G.T. Sallee, Minkowski decomposition of convex sets, Israel J. Math. 12 (1972) 266–276.
[26] R. Schneider, Convex Bodies: The Brunn–Minkowski Theory, Cambridge University Press, Cambridge, 1993.
[27] G.C. Shephard, Decomposable convex polytopes, Mathematika 10 (1963) 89–95.
[28] R. Silverman, Decomposition of plane convex sets, PhD thesis, University of Washington, 1970.
[29] R. Silverman, Decomposition of plane convex sets, part I, Pacific J. Math. 47 (1973) 521–530.
[30] Z. Smilansky, An indecomposable polytope all of whose facets are decomposable, Mathematika 33 (2) (1986) 192–196.
[31] Z. Smilansky, Decomposability of polytopes and polyhedra, Geom. Dedicata 24 (1987) 29–49.
[32] D. Škulj, R. Hable, Coefficients of ergodicity for Markov chains with uncertain parameters, Metrika 76 (2013) 107–133.
[33] M.C.M. Troffaes, Online documentation of improb (a Python module for working with imprecise probabilities), http://pythonhosted.org/improb/lowprev/lowpoly.html, 2013, online, accessed 5 December 2013.
[34] M.C.M. Troffaes, G. de Cooman, Lower Previsions, Wiley, 2014.
[35] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, London, 1991.
[36] S. Willard, General Topology, Addison–Wesley, 1970.
[37] D. Yost, Some indecomposable polyhedra, Optimization 56 (5–6) (2007) 715–724.