A behavioural justification for using sensitivity analysis in imprecise multinomial processes

Jasper De Bock[a,*], Gert de Cooman[a]

[a]*Ghent University, SYSTeMS Research Group*
Technologiepark–Zwijnaarde 914, 9052 Zwijnaarde, Belgium.

Abstract

We consider two fundamentally different ways of defining an imprecise-probabilistic multinomial process. On the one hand, we have the so-called sensitivity analysis approach. The corresponding imprecise-probabilistic joint model is defined as the lower envelope of joint models of precise multinomial processes, each of which has a different marginal model taken from some closed and convex set of candidate marginal models. For finite subsets of variables in this process, this corresponds to using Walley's so-called type-2 product of identically distributed variables. On the other hand, we consider a behavioural approach, defining an imprecise multinomial process by imposing exchangeability and either forward irrelevance or epistemic independence. Our main result is that both approaches lead to the exact same imprecise multinomial process. This fairly technical result has an important philosophical consequence as well: it provides the sensitivity analysis approach and the related type-2 product with a behavioural justification. We compare our justification favourably with a previous attempt by Cozman and explain how it fits into the more general problem of providing a behavioural justification for so-called strongly independent models, or equivalently, for using lower envelopes of stochastically independent models.

Keywords: Imprecise multinomial processes, sensitivity analysis, exchangeability, type-2 product, forward irrelevance, epistemic independence

---

1. Introduction

In classical probability theory, a *multinomial process* can be defined as a countable sequence of *stochastically independent* and *identically distributed* (*iid*) random variables $X_1, \dots, X_n, \dots$ assuming values in some finite non-empty set $\mathcal{X}$.[1] Every variable in this

*Corresponding author

Email addresses: `jasper.debock@ugent.be` (Jasper De Bock), `gert.decooman@ugent.be` (Gert de Cooman)

[1]Although this generalisation of a Bernoulli process is very common in statistics, there seems to be no consensus on how to call it. In fact, most textbooks simply describe it, without providing it with a name. Judging from our own experience in trying to choose a name, this is probably because any attempt at fixing a name is bound to lead to confusion. Our current choice of calling it a multinomial process is,

sequence has the same given marginal distribution, which is expressed in terms of a probability mass function p, assigning a probability of occurrence $p(x)$ to every x in $\mathcal{X}$. Due to the assumption of stochastic independence, the probability of any instantiation of a finite subsequence of $X_1, \ldots, X_n, \ldots$ is obtained simply by multiplying the marginal probabilities for the individual variables.

When generalising a multinomial process to allow for imprecision, in which case we call it an *imprecise multinomial process*, the first step is to allow for imprecision in the marginal model for a single variable. This means that it need no longer be represented by a single probability mass function. Instead, we use a set of probability mass functions, which is usually taken to be closed and convex, making it a so-called *credal set* $\mathcal{M}$. An alternative method, which will be important in this paper, is to represent the marginal model by means of a so-called *coherent lower prevision* $\underline{P}$, which can be regarded as a lower expectation functional, or equivalently, a lower envelope of expectation functionals. Coherent lower previsions are mathematically equivalent to credal sets, but have the advantage of having a clear behavioural interpretation in terms of betting behaviour, which we will explain in Section 2.

Once we have such an imprecise marginal model, the next step is to use it to construct an imprecise multinomial process. However, in contrast with the precise-probabilistic case, there is no unique way of doing so, the main reason being that independence has no single unique definition in the context of imprecise probabilities; see Ref. [4] for an influential review of different imprecise-probabilistic notions of independence and Ref. [5, Section 3] for an overview of the relevant literature that clarifies the sometimes confusing terminology that has been used over the past thirty years.

Amongst the many possible ways to construct an imprecise multinomial process, perhaps the most intuitive one is the *sensitivity analysis approach*. It assumes that the countable sequence $X_1, \ldots, X_n, \ldots$ of variables can be modelled by means of a classical, precise multinomial process, implying an assumption of stochastic independence. However, in contrast with the precise-probabilistic case, its marginal probability mass function is not precisely known, but only known to lie within some credal set. This could be due to time-constraints on the elicitation process, because different experts disagree, or for some other reason. The joint model for the resulting imprecise multinomial process is then taken to be the lower envelope of the corresponding set of joints of precise multinomial processes. For finite subsets of the countable sequence, one obtains Walley's so-called *type-2 product* [6, Section 9.3.5].

Although this sensitivity analysis approach is often regarded as intuitive, it suffers from a lack of a clear behavioural interpretation. We will come back to this in Section 6.2, but one of the main problems is that by taking lower envelopes, one implicitly adopts convexity and thereby allows for convex mixtures of stochastically independent models. These need no longer be stochastically independent themselves and are therefore incompatible with the initial assumptions of the sensitivity analysis approach; see Ref. [5] for more information on this incompatibility between stochastic independence and convexity and an overview of the relevant literature. Other than that, the sensitivity analysis approach is also based on an implicit assumption that there is some true—but only par-

---

admittedly, no exception and was based on, amongst others, Refs. [1, Section 2.1.1], [2, page 227] and [3, Section 8.2.7]. Alternative names that have been proposed over the years are 'scheme of repeated trials', 'Bernoulli(an) scheme', 'generalised Bernoulli process' and 'multinomial trials process'.

tially known—probability model that describes our process. As should become clear in Section 2, this is by no means trivial nor required: uncertainty can also be described directly in terms of lower previsions, without any reference to probabilities.

In order to avoid these interpretation problems, one can construct an imprecise multinomial process without making an assumption of stochastic independence, and replace it with a notion of independence that fits better in a behavioural framework that allows for imprecision and partial probability assessments. An example of such an approach, requiring much milder assumptions than sensitivity analysis, is to impose *forward irrelevance* [7]: for every variable X_n, all previous variables $X_1, \ldots, X_{n-1}$ are assumed to be epistemically irrelevant to X_n, where epistemic irrelevance is an asymmetric imprecise-probabilistic notion of independence that will be introduced and discussed in detail in Section 3.4. Loosely speaking, imposing forward irrelevance means that learning the value of 'past' variables should not change our beliefs about the current one. If future variables are also considered to be epistemically irrelevant to the current ones, one obtains *epistemic independence* [8], which can be regarded as a symmetrised version of forward irrelevance. In contrast with sensitivity analysis, forward irrelevance and epistemic independence both have a clear behavioural meaning. They can be stated fully in terms of coherent lower previsions, which provides them with an interpretation in terms of betting behaviour.

The main contribution of this paper is to show that such a behavioural interpretation can also be given to the sensitivity analysis approach by making an additional judgement of *exchangeability* [6, Section 9.5]. Without becoming too technical at this point, this essentially means that the order of the variables in $X_1, \ldots, X_n, \ldots$ is deemed irrelevant. In the precise case, such a judgement is redundant, because we know from de Finetti's representation theorem that a stochastic process is exchangeable if and only if it is a convex mixture of iid processes. Classical multinomial processes are therefore trivially exchangeable since they are themselves iid. When working with imprecise probabilities, however, exchangeability is not necessarily implied by the iid property (under the many different guises independence can have in this wider context). It is therefore natural to wonder what happens if we impose it in addition to some form of the iid property. We will prove that by combining an assessment of exchangeability with forward irrelevance or epistemic independence, the resulting imprecise multinomial process is identical to the one obtained by applying sensitivity analysis. Because forward irrelevance, epistemic independence and exchangeability can all be given a behavioural interpretation, this results in a behavioural justification for the sensitivity analysis approach, and consequently, also for the type-2 product. We compare this justification favourably with a previous attempt by Cozman [5, Section 3.5], argue that the latter provides only a partial justification and provide a simple example that illustrates why.

The paper is organised as follows. We start in Section 2 with a short introduction to Walley's behavioural account of coherent lower previsions [6], as that is one of the main mathematical tools for this paper.

In Section 3, we go on to show how these coherent lower previsions can be used to model beliefs regarding an infinite sequence of variables and how to impose in this framework such behavioural assessments as: a marginal model, exchangeability, epistemic independence and forward irrelevance. We conclude the section by introducing the concept of a least committal, most conservative model.

Sections 4 and 5 constitute the technical core of this paper. We recall in Section 4 how an exchangeable sequence can be represented by means of a coherent lower prevision on the set of all so-called polynomial gambles on a simplex. Section 5 then goes on to introduce a behaviourally justified notion of an imprecise multinomial process, defining it as the most imprecise (least committal, most conservative) one that has a given marginal model and satisfies both exchangeability and forward irrelevance. Using the results in Section 4, we manage to derive simple expressions for this process.

After this rather technical part, we introduce the sensitivity analysis approach to imprecise multinomial processes in Section 6. We explain how it suffers from a lack of interpretation and why we believe that the aforementioned justification of Cozman provides only a partial solution to this problem. We then go on to provide our own solution by noting that the imprecise multinomial process that corresponds to sensitivity analysis is identical to the behaviourally justified one constructed in Section 5. This implies that combining exchangeability with forward irrelevance leads to a behavioural justification for using the sensitivity analysis approach to imprecise multinomial processes and the related type-2 product. We also show how a similar justification can be obtained by combining exchangeability with epistemic independence instead of forward irrelevance.

Section 7 shifts the focus away from the sensitivity analysis approach to imprecise multinomial processes and the related type-2 product and towards general so-called *strongly independent* [5] models, of which the former are but special cases. Loosely speaking, a model satisfies strong independence if its joint distribution is a lower envelope of stochastically independent ones. We provide a critical overview of some earlier attempts to justify this notion of strong independence and lay bare some of their conceptual weaknesses.

We summarise the paper in Section 8, which also contains some final remarks and suggestions for future research. The proofs of our main results are gathered in an appendix.

2. Coherent lower previsions

Consider a *variable* X that assumes values in some non-empty set Ω. Its actual value is unknown, but a subject may entertain certain beliefs about this value of X. We will try and model such beliefs by looking at the prices our subject is willing to pay to participate in certain bets, which we call *gambles*. This approach to probability is both subjective and behavioural, following the ideas of de Finetti [9]. In contrast with the more formalist approaches that are usually adopted, it has the advantage of providing simple, behavioural justifications for exchangeability and a number of different imprecise-probabilistic notions of independence, such as epistemic irrelevance and epistemic independence.

A *gamble* f on Ω is defined as a bounded map from Ω to the set $\mathbb{R}$ of real numbers. We denote by $\mathcal{G}(\Omega)$ the set of all gambles on Ω. The reason why we call them gambles is because they can be interpreted as uncertain rewards on the outcome of X. If the actual value of X turns out to be ω, the (possibly negative) reward is $f(\omega)$, expressed in units of some pre-determined linear utility. Receiving a negative reward means giving away its absolute value. For this reason, we will sometimes also use the notation $f(X)$ for f and call it a 'gamble on X'. Stating that you *accept* a gamble $f(X)$ can be interpreted as your willingness to engage in the transaction in which the actual value ω of X is determined and you subsequently receive the associated reward $f(\omega)$.

Bruno de Finetti [9] proposed to model a subject's beliefs by eliciting his fair price, or *prevision*, $P(f)$ for certain gambles f. This $P(f)$ can be defined as the unique real number for which the subject is willing to buy the gamble f for all prices $s < P(f)$ (that is, accept the gamble $f - s$) and sell f for all prices $t > P(f)$ (that is, accept the gamble $t - f$). In other words, $P(f)$ is both his supremum buying price and his infimum selling price for the gamble f. The obvious problem with this approach is that it requires a subject to be able to choose, for (almost) every real r, between buying f for the price r or selling it for that price.

A solution to this problem is given by Walley's theory of lower and upper previsions [6], which goes back to work by Williams [10]. The main idea is to elicit supremum buying prices and infimum selling prices separately, dropping the assumption that they should coincide. For any gamble f, the *lower prevision* $\underline{P}(f)$ is our subject's supremum buying price for f; similarly, our subject's *upper prevision* $\overline{P}(f)$ is his infimum selling price for f. In other words, the subject is willing to buy the gamble f for all prices $s < \underline{P}(f)$ and sell it for all prices $t > \overline{P}(f)$. For prices $\underline{P}(f) \leq r \leq \overline{P}(f)$, he is allowed to remain undecided. If $\underline{P}(f)$ and $\overline{P}(f)$ happen to coincide for a gamble f, then the value $P(f) = \underline{P}(f) = \overline{P}(f)$ is called the subject's (precise) prevision for f and we obtain a fair price for f in de Finetti's sense.

In this paper, we will be working with lower and upper previsions that have the same domain $\mathcal{K}$, which will always be some linear subspace of $\mathcal{G}(\Omega)$. As a consequence, for any gamble f in the domain $\mathcal{K}$, the gamble $-f$ will also be contained in $\mathcal{K}$. Since buying a gamble f for the price s is the same as selling the gamble $-f$ for the price $-s$, the lower and upper previsions are related through the property of conjugacy: $\underline{P}(f) = -\overline{P}(-f)$ for any gamble f in $\mathcal{K}$. It therefore suffices to study one of them, since the other one can be derived from it. We will focus on lower previsions.

In order for a lower prevision to represent a rational subject's beliefs about the value of X, it should satisfy the rationality criterion of *coherence* [6, Chapter 2]. For lower previsions that are defined on a linear space of gambles $\mathcal{K}$, coherence reduces to the following three conditions [6, Section 2.5.5]. For all gambles f and g in $\mathcal{K}$ and any non-negative real λ, it should hold that:

P1. $\underline{P}(f) \geq \inf f$; [accepting sure gains]

P2. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$; [non-negative homogeneity]

P3. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$. [superadditivity]

In order to get a feeling for why it is indeed rational to satisfy requirements P1–P3, one can apply the interpretation of lower previsions that was given earlier: $\underline{P}(f)$ is the supremum buying price for the uncertain reward f. Under this interpretation of a lower prevision, P1–P3 can be reformulated as follows. P1 states that a rational subject should always be willing to buy a gamble for any price that is lower than all the possible rewards associated with the gamble. P2 states that his willingness to buy a gamble for some price should not be affected by the utility units in which the price and rewards of the gamble are measured. And finally, P3 states that he should be willing to pay at least as much for $f + g$ as the sum of what he is willing to pay for f and g separately. See Ref. [6, Chapter 2] for an extensive discussion of the justification of coherence as a rationality requirement.

As a consequence of coherence, one can derive a number of additional useful properties. The following holds for any coherent lower prevision whenever the gambles involved belong to its domain [6, Section 2.6.1]:

P4. $\underline{P}$ is *monotone*, that is, if $f \leq g$, then $\underline{P}(f) \leq \underline{P}(g)$;

P5. $\underline{P}(f + \mu) = \underline{P}(f) + \mu$ for all real μ;

P6. $\inf f \leq \underline{P}(f) \leq \overline{P}(f) \leq \sup f$.

To conclude this introduction to the behavioural theory of lower and upper previsions, let us briefly mention the derived concepts of lower and upper probabilities. For any event $A \subseteq \Omega$, we use $\mathbb{I}_A$ to denote its corresponding *indicator* (gamble) on Ω, defined by $\mathbb{I}_A(\omega) = 1$ if $\omega \in A$ and $\mathbb{I}_A(\omega) = 0$ if $\omega \notin A$. Since these indicators are just zero–one-valued gambles, we can consider their lower prevision. For any event $A \subseteq \Omega$, we call $\underline{P}(A) := \underline{P}(\mathbb{I}_A)$ the lower probability of A. It is the supremum buying price that you are willing to pay for the gamble $\mathbb{I}_A$ that gives you 1 unit if A occurs and nothing otherwise. In other words, $\underline{P}(A)$ is the supremum betting rate at which you are disposed to bet on A. Similarly we define the upper probability of an event $A \in \Omega$ as $\overline{P}(A) := \overline{P}(\mathbb{I}_A)$. Due to coherence property P6, $0 \leq \underline{P}(A) \leq \overline{P}(A) \leq 1$ for all $A \subseteq \Omega$.

3. Behaviourally justified assessments

We are now ready to show how to use the behavioural tools introduced in the previous section to impose a given marginal model and structural assessments such as exchangeability, epistemic independence and forward irrelevance. For each of these properties, we will give a formal definition and explain how to interpret this definition in terms of betting behaviour of a subject. To conclude, we explain how imposing such behavioural assessments leads to a so-called least committal model that is compatible with them. But first, we introduce basic tools and notation for modelling a subject's beliefs about a countable sequence $X_1, \ldots, X_n, \ldots$ of variables assuming values in a *finite* non-empty set $\mathcal{X}$.

3.1. Modelling a countable sequence of variables

We start of by considering a finite sequence $X_1, \ldots, X_n$. In order to easily model beliefs about the values of the variables in such a sequence using the language of lower and upper previsions, we will regard it as a single variable $X_{\downarrow n} := (X_1, \ldots, X_n)$ that assumes values $x_{\downarrow n} := (x_1, \ldots, x_n)$ in the finite non-empty set $\Omega = \mathcal{X}_{\downarrow n} := \mathcal{X}^n$. We will model a subject's beliefs regarding the actual value of $X_{\downarrow n}$ by means of a coherent lower prevision $\underline{P}_{\downarrow n}$ on the set $\mathcal{G}(\mathcal{X}_{\downarrow n})$ of all gambles on $\mathcal{X}_{\downarrow n}$.

So how can we use this to model our subject's beliefs about a countable sequence $X_1, \ldots, X_n, \ldots$? Our approach will be to provide, for each $n \in \mathbb{N}$, a coherent lower prevision $\underline{P}_{\downarrow n}$ that serves as a model for the first n variables in the countable sequence. We will show further on that this enables us to derive a model, in the form of a coherent lower prevision, for any finite subset of the variables in the sequence $X_1, \ldots, X_n, \ldots$ We will not try to construct a single coherent lower prevision describing the complete countable sequence at once (except later on in Section 4, under the assumption of exchangeability), but the interested reader may take a look at Refs. [11, Section 5] and [12, Section 3.2] to see how it could be derived from our sequence of finite models $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$.

Before we go further, we need to introduce some more notation. For any non-empty finite subset R of $\mathbb{N}$, we will denote by X_R the tuple of variables (with one component for each $r \in R$) that takes values in $\mathcal{X}_R := \mathcal{X}^{|R|}$, in which $|R|$ is the number of elements in R. Elements of $\mathcal{X}_R$ will usually be denoted by x_R. To give an example: with $R = \{2, 3\}$, the variable $X_R = X_{\{2,3\}} = (X_2, X_3)$ takes values $x_R = x_{\{2,3\}} = (x_2, x_3)$ in $\mathcal{X}_R = \mathcal{X}_{\{2,3\}} = \mathcal{X}^2$. For any $n \in \mathbb{N}$, both $X_{\downarrow n}$ and the individual variable X_n can be seen as a special case. It suffices to choose $R = \{1, \dots, n\}$ and $R = \{n\}$ respectively.

For two non-empty and finite subsets R and N of $\mathbb{N}$, with $R \subseteq N$, one can derive a coherent lower prevision $\underline{P}_R$ on $\mathcal{G}(\mathcal{X}_R)$ by applying marginalisation to an existing coherent lower prevision $\underline{P}_N$ on $\mathcal{G}(\mathcal{X}_N)$. The marginal model $\underline{P}_R$ is defined by

$$\underline{P}_R(f_R) := \underline{P}_N(f_R) \text{ for all } f_R \in \mathcal{G}(\mathcal{X}_R). \tag{1}$$

In this expression, we have identified the gamble f_R on $\mathcal{X}_R$ with its *cylindrical extension* f_N on $\mathcal{X}_N$. It is given by $f_N(x_N) := f_R(x_R)$ for all $x_N \in \mathcal{X}_N$, where x_R is the restriction (i.e., the projection) of x_N to $\mathcal{X}_R$. To give an example: with $R = \{2\}$, $N = \{2, 3\}$ and some given gamble f_R on $\mathcal{X}_R$, cylindrical extension yields $f_N(x_2, x_3) = f_R(x_2)$ for all x_2 and x_3 in $\mathcal{X}$. As you can see from this example, f_N is a gamble that theoretically depends on the value of $X_N = (X_2, X_3)$, but in practice only depends on the value $x_R = x_2$ that $X_R = X_2$ takes in $\mathcal{X}_R = \mathcal{X}$ and on that domain, coincides with f_R. It is therefore a rational requirement that a subject's supremum buying prices for the gambles f_R and f_N should be identical, motivating the definition of the marginal model $\underline{P}_R$.

At this point, it should be clear that we can not just use any sequence of models $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$, to describe our subject's beliefs about a countable sequence $X_1, \dots, X_n, \dots$ of variables. If we choose r and n in $\mathbb{N}$ with $r \leq n$ and let $R = \{1, \dots, r\}$ and $N = \{1, \dots, n\}$, then $\underline{P}_N = \underline{P}_{\downarrow n}$ and $\underline{P}_R = \underline{P}_{\downarrow r}$. Therefore, by the argument given above, $\underline{P}_{\downarrow n}$ and $\underline{P}_{\downarrow r}$ should be related through marginalisation. We will refer to this property as *time-consistency*. It can be formally stated as follows: for all k and n in $\mathbb{N}$, with $k \leq n$, it should hold that

$$\underline{P}_{\downarrow k}(f) = \underline{P}_{\downarrow n}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow k}). \tag{2}$$

It is now easy to see that from a sequence of models $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$, we can derive a model, in the form of a coherent lower prevision, for any non-empty finite subset of our countable sequence $X_1, \dots, X_n, \dots$ of variables. In order to do so, consider any non-empty finite subset R of $\mathbb{N}$ and let k be the biggest element of R. If we pick $n \geq k$ and choose $N = \{1, \dots, n\}$, then $\underline{P}_N = \underline{P}_{\downarrow n}$ and we can derive a lower prevision $\underline{P}_R$ from it by marginalising it as described above. Due to the time-consistency requirement on our models $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$, and the transitivity of marginalisation, the obtained model $\underline{P}_R$ will always be the same, whatever $n \geq k$ we use to construct it. In this way, for any non-empty finite subset R of $\mathbb{N}$, the sequence $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$, provides us with a unique coherent lower prevision $\underline{P}_R$ that describes our subject's beliefs about the value that X_R will assume in $\mathcal{X}_R = \mathcal{X}^{|R|}$.

3.2. Exchangeability and permutability

A sequence of variables is called *exchangeable* [6, Section 9.5] if, simply put, the order of the variables is deemed irrelevant to inferences about them. In order to define it more formally, consider any permutation π of the set of indices $\{1,\ldots,n\}$ and denote the set of all such permutations as $\mathcal{P}_n$. With any permutation $\pi \in \mathcal{P}_n$, we associate a permutation of $\mathcal{X}_{\downarrow n}$, defined by $\pi x_{\downarrow n} = \pi(x_1,\ldots,x_n) := (x_{\pi(1)},\ldots,x_{\pi(n)})$. Similarly, for any gamble f in $\mathcal{G}(\mathcal{X}_{\downarrow n})$, we define the permuted gamble $\pi f = f \circ \pi$, so $(\pi f)(x_{\downarrow n}) = f(\pi x_{\downarrow n})$ for all $x_{\downarrow n}$ in $\mathcal{X}_{\downarrow n}$.

A finite sequence of imprecisely modelled variables $X_1,\ldots,X_n$ is called *exchangeable* if a subject is willing to exchange any gamble $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ for the permuted gamble πf, for any permutation $\pi \in \mathcal{P}_n$. In our framework, this translates to demanding that $\underline{P}_{\downarrow n}(\pi f - f) \geq 0$ for any $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\pi \in \mathcal{P}_n$. A countable sequence $X_1,\ldots,X_n,\ldots$ of variables is called exchangeable if any finite subsequence is exchangeable, or equivalently, if the finite sequence $X_1,\ldots,X_n$ is exchangeable for all $n \in \mathbb{N}$:

$$\underline{P}_{\downarrow n}(\pi f - f) \geq 0 \text{ for any } n \in \mathbb{N},\ f \in \mathcal{G}(\mathcal{X}_{\downarrow n}) \text{ and } \pi \in \mathcal{P}_n. \tag{3}$$

Because this property has to hold for any gamble f, it also holds for $-f$, and therefore $\underline{P}_{\downarrow n}(f - \pi f) = \underline{P}_{\downarrow n}(\pi(-f) - (-f)) \geq 0$, which is equivalent to $\overline{P}_{\downarrow n}(\pi f - f) \leq 0$ by conjugacy. Since we always have that $\overline{P}_{\downarrow n}(\pi f - f) \geq \underline{P}_{\downarrow n}(\pi f - f)$, we find that $\overline{P}_{\downarrow n}(\pi f - f) = \underline{P}_{\downarrow n}(\pi f - f) = 0$, for any $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\pi \in \mathcal{P}_n$. In other words, our subject has a precise prevision $P_{\downarrow n}(\pi f - f) = 0$ for the gamble $\pi f - f$. This illustrates that exchangeability is a remarkably strong property.

Because a precise prevision for a gamble can be interpreted as a subject's fair price for that gamble in de Finetti's sense, an assessment of exchangeability has a very clear behavioural interpretation. Exchanging a gamble f for its permuted version πf—which is the same as receiving the gamble $\pi f - f$—is equivalent to a zero pay-off to our subject, meaning that he is completely indifferent between πf and f.

Due to the superadditivity of coherent lower previsions, exchangeability implies that $\underline{P}_{\downarrow n}(\pi f) = \underline{P}_{\downarrow n}((\pi f - f) + f) \geq \underline{P}_{\downarrow n}(\pi f - f) + \underline{P}_{\downarrow n}(f) = \underline{P}_{\downarrow n}(f)$. Similarly, we find that $\underline{P}_{\downarrow n}(f) \geq \underline{P}_{\downarrow n}(\pi f)$. We therefore see that the exchangeability of the variables $X_1,\ldots,X_n$ implies that they are also *permutable* [6, Section 9.4], meaning that $\underline{P}_{\downarrow n}(\pi f) = \underline{P}_{\downarrow n}(f)$, for any $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\pi \in \mathcal{P}_n$. For precise probability models, permutability and exchangeability are equivalent, but for the more general imprecise probability models, permutability is a weaker property than exchangeability. Similarly to our definition of exchangeability, we call a countable sequence $X_1,\ldots,X_n,\ldots$ permutable if every finite subsequence is. It should be clear that any countable sequence that is exchangeable, is also permutable. Such sequences satisfy the following intuitive, but very useful property.

Proposition 1. *Consider a countable sequence $X_1,\ldots,X_n,\ldots$ that is permutable. Then it holds for any finite subset R of $\mathbb{N}$ that has $r = |R|$ elements that*

$$\underline{P}_R(f) = \underline{P}_{\downarrow r}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}^r).$$

What this means is that the coherent lower prevision that describes our subject's beliefs about a subset of $r \in \mathbb{N}$ variables out of the countable sequence $X_1,\ldots,X_n,\ldots$ does not depend on the particular subset or the order of its elements and is uniquely given by the coherent lower prevision $\underline{P}_{\downarrow r}$.

3.3. Marginalisation to a given local model

The requirement that every individual variable should be described by the same given marginal model is crucial to any definition of an imprecise multinomial process.

In Walley's framework, this marginal model is given in the form of a coherent lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$. The requirement can then be formally stated as

$$\underline{P}_{\{k\}}(f) = \underline{P}(f) \text{ for all } k \in \mathbb{N} \text{ and all } f \in \mathcal{G}(\mathcal{X}). \tag{4}$$

The behavioural interpretation of this requirement is the following: whether or not a subject is willing to accept a given bet on the outcome of some variable X_k in the sequence $X_1, \ldots, X_n, \ldots$ does not depend on the value of $k \in \mathbb{N}$.

Under an additional assumption of permutability (or exchangeability, since that implies permutability), Eq. (4) can be reformulated very elegantly. Due to Proposition 1, it suffices to require that $\underline{P}$ and $\underline{P}_{\downarrow 1}$ are identical:

$$\underline{P}_{\downarrow 1}(f) = \underline{P}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}). \tag{5}$$

We should also mention that we have assumed that $\underline{P}$ is defined on all gambles in $\mathcal{G}(\mathcal{X})$. In practice however, it might occur that we are only given a lower prevision $\underline{P}(f)$ for a limited number of gambles $f \in \mathcal{G}(\mathcal{X})$. This does not impose any real restrictions. As long as such a finite number of assessments of lower previsions is consistent (*avoids sure loss*), we can always use *natural extension* to extend them to a unique least committal (most conservative) coherent lower prevision that is defined on all gambles in $\mathcal{G}(\mathcal{X})$; see Ref. [6, Chapter 3] for more information on avoiding sure loss and natural extension.

3.4. Epistemic independence and forward irrelevance

Of all the properties that we promised to elaborate on, epistemic independence [8] and forward irrelevance [7] are the only ones that we still need to explain in more detail. We start by defining them for finite sequences of variables. Consider a finite sequence of variables $X_1, \ldots, X_n$ for which our subject's beliefs are described by a coherent lower prevision $\underline{P}_{\downarrow n}$. When do we call such a sequence epistemically independent? And how do we define forward irrelevance?

To explain this, we need to introduce conditional lower previsions and the concept of epistemic irrelevance. We use N as an alternative notation for the set $\{1, \ldots, n\}$ and consider two disjoint subsets I and O of N and their corresponding sets of variables X_I and X_O. Previously, we introduced the coherent lower prevision $\underline{P}_O$ as a model for a subject's (unconditional) beliefs about the value that X_O assumes in $\mathcal{X}_O$. Now what happens if the subject's receives the information that $X_I = x_I$? His beliefs about the value of X_O need not remain the same and we will model his new, conditional beliefs by means of a conditional lower prevision $\underline{P}_O(\cdot|x_I)$. If his beliefs do remain the same, whatever the value that X_I assumes, we say that the subject regards X_I as *epistemically irrelevant* for X_O. Formally, X_I is irrelevant to X_O if $\underline{P}_O(f) = \underline{P}_O(f|x_I)$ for all $f \in \mathcal{G}(\mathcal{X}_O)$ and $x_I \in \mathcal{X}_I$. The behavioural interpretation of such an assessment is that the subject's willingness to accept a given bet on the outcome of X_O does not change if he receives the information that X_I attains some value x_I in $\mathcal{X}_I$.

We can use assessments of irrelevance to define each of the different imprecise-probabilistic notions of independence mentioned above. The variables $X_1, \ldots, X_n$ are called *epistemically many-to-many independent* [8] if X_I is irrelevant to X_O for any two disjoint subsets I and O of N. By restricting the possible choices of I and O, weaker concepts of independence can be obtained. We call the variables $X_1, \ldots, X_n$ *epistemically* *many-to-one independent* [8] if X_I is irrelevant to X_o for any $o \in N$ and $I \subseteq N \setminus \{o\}$. Finally, we say that the variables $X_1, \ldots, X_n$ are *forward irrelevant* [7] if $X_{\downarrow k} = X_{\{1,\ldots,k\}}$ is irrelevant to X_{k+1} for all $1 \leq k \leq n-1$.

These different kinds of independence are defined by equating conditional lower previsions with unconditional ones. In order to translate these definitions into conditions on the joint model $\underline{P}_{\downarrow n}$, we need to require that the conditional lower previsions that are obtained by independence, are compatible with one another and with $\underline{P}_{\downarrow n}$. This property is called *joint coherence* (also referred to as coherence or strong coherence). In its general form, it is a rather complicated requirement that we will not explain in detail; we refer to [6, Section 7.1.4] for an extensive discussion that justifies it as a rationality requirement. In order to get some intuitive notion of joint coherence, it is useful to know that it is similar to (but more general than) the precise-probabilistic requirement that conditional and unconditional probabilities should be related through Bayes's rule. Similarly to what happens in this precise case, difficulties arise and definitions of joint coherence get more complicated whenever (lower) probabilities of conditioning events are equal to zero.

Due to the rather complicated nature of joint coherence, it might seem difficult to impose the imprecise-probabilistic notions of independence that were introduced in this section. However, for the case of epistemic many-to-one independence, we can use the following result.

Proposition 2 ([8, Corollary 14(ii)])**.** *A finite sequence $X_1, \ldots, X_n$ of variables, modelled by means of a coherent lower prevision $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, with for all $1 \leq o \leq n$ a corresponding marginal coherent lower previsions $\underline{P}_{\{o\}}$ on $\mathcal{G}(\mathcal{X}_o)$, is epistemically many-to-one independent if and only if for every $1 \leq o \leq n$ and $I \subseteq \{1, \ldots, n\} \setminus \{o\}$*

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_I\}}[f - \underline{P}_{\{o\}}(f)]) = 0 \quad \text{for all } f \in \mathcal{G}(\mathcal{X}_o) \text{ and } x_I \in \mathcal{X}_I. \tag{6}$$

We can do something similar for forward irrelevance. By trivially adapting the proof of Proposition 2 [8, Corollary 14(ii)], we obtain the following unique characterisation.

Proposition 3. *A finite sequence $X_1, \ldots, X_n$ of variables, modelled by means of a coherent lower prevision $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, with for all $1 \leq k \leq n$ a corresponding marginal coherent lower prevision $\underline{P}_{\{k\}}$ on $\mathcal{G}(\mathcal{X}_k)$, is forward irrelevant if and only if for every $1 \leq k \leq n-1$*

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}_{\{k+1\}}(f)]) = 0 \quad \text{for all } f \in \mathcal{G}(\mathcal{X}_{k+1}) \text{ and } x_{\downarrow k} \in \mathcal{X}_{\downarrow k}. \tag{7}$$

So far, we have defined epistemic independence and forward irrelevance for finite numbers of variables only. In order to apply these notions to imprecise multinomial processes, we will need to define them for countable sequences as well. Therefore, we extend their definitions as follows. A countable sequence $X_1, \ldots, X_n, \ldots$ of variables is called epistemically independent—either many-to-many or many-to-one—if and only if every finite subsequence of variables is. Since both kinds of epistemic independence are preserved under marginalisation [8, Proposition 9], this is equivalent to requiring that, for every $n \in \mathbb{N}$, the variables $X_1, \ldots, X_n$ should be epistemically independent—either many-to-many or many-to-one. Similarly, a countable sequence $X_1, \ldots, X_n, \ldots$ of variables is called forward irrelevant if and only if, for all $n \in \mathbb{N}$, the variables $X_1, \ldots, X_n$ are.

For the remainder of this paper, it will be useful to have a simple way of imposing forward irrelevance on a countable sequence $X_1, \dots, X_n, \dots$ of variables that are known to be identically distributed. The following simple result is a fairly direct consequence of Proposition 3.

Corollary 4. *A countable sequence $X_1, \dots, X_n, \dots$ of variables, modelled by means of a time-consistent sequence of coherent lower previsions $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n \in \mathbb{N}$, and identically distributed with a given marginal model $\underline{P}$ on $\mathcal{G}(\mathcal{X})$, is forward irrelevant if and only if for all $k \in \mathbb{N}$*

$$\underline{P}_{\downarrow k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]) = 0 \textit{ for all } f \in \mathcal{G}(\mathcal{X}_{k+1}) \textit{ and } x_{\downarrow k} \in \mathcal{X}_{\downarrow k}. \tag{8}$$

3.5. Least committal models

Let Γ be a (possibly infinite) index set such that for all $\gamma \in \Gamma$, we have a time-consistent sequence of coherent lower previsions $\{\underline{P}^{\gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, describing a subject's beliefs about a countable sequence $X_1, \dots, X_n, \dots$ of variables.

Consider now two such sequences, corresponding to $\gamma_1 \in \Gamma$ and $\gamma_2 \in \Gamma$ respectively. We then say that the sequence $\{\underline{P}^{\gamma_1}_{\downarrow n}\}_{n\in\mathbb{N}}$, is *less committal*, or *more conservative*, than $\{\underline{P}^{\gamma_2}_{\downarrow n}\}_{n\in\mathbb{N}}$, if for all $n \in \mathbb{N}$ the lower prevision $\underline{P}^{\gamma_1}_{\downarrow n}$ is *point-wise dominated* by the corresponding lower prevision $\underline{P}^{\gamma_2}_{\downarrow n}$: $\underline{P}^{\gamma_1}_{\downarrow n}(f) \leq \underline{P}^{\gamma_2}_{\downarrow n}(f)$ for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$. The reason for this terminology should be clear: a subject using the sequence $\{\underline{P}^{\gamma_2}_{\downarrow n}\}_{n\in\mathbb{N}}$, will be buying gambles f on $\mathcal{X}_{\downarrow n}$ at supremum prices $\underline{P}^{\gamma_2}_{\downarrow n}(f)$ that are at least as high as the supremum prices $\underline{P}^{\gamma_1}_{\downarrow n}(f)$ of a subject that uses the sequence $\{\underline{P}^{\gamma_1}_{\downarrow n}\}_{n\in\mathbb{N}}$.

Next, consider the sequence of lower envelopes $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, given for every $n \in \mathbb{N}$ by

$$\underline{P}^{\Gamma}_{\downarrow n}(f) := \inf\big\{\underline{P}^{\gamma}_{\downarrow n}(f) : \gamma \in \Gamma\big\} \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}).$$

Then $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, will be a sequence of coherent lower previsions [6, Section 2.6.3] that is clearly time-consistent as well. Furthermore, we find that for all $\gamma \in \Gamma$, the sequence $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, is less committal than $\{\underline{P}^{\gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$.

Suppose now that for all $\gamma \in \Gamma$, the sequence $\{\underline{P}^{\gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, satisfies one—the same for each γ—of the behavioural properties introduced in the previous sections. Will the sequence of lower envelopes $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, then satisfy this property as well? For exchangeability, this is clearly the case because if $\underline{P}^{\gamma}_{\downarrow n}(\pi f - f) \geq 0$ for all $\gamma \in \Gamma$, then obviously $\inf\{\underline{P}^{\gamma}_{\downarrow n}(\pi f - f) : \gamma \in \Gamma\} \geq 0$. Similarly, one can easily show that permutability and being identically distributed with a given marginal model $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ are properties that are preserved under taking lower envelopes. For forward irrelevance and—many-to-many or many-to-one—epistemic independence, this is not as trivial. However, since joint coherence is preserved under taking lower envelopes [6, Section 7.1.6], these three different notions of independence are preserved as well; see Ref. [8, Section 4.3] for an explicit argument for the case of many-to-one epistemic independence, which can be trivially adapted to many-to-many epistemic independence and forward irrelevance.

With this in mind, let Γ no longer be an arbitrary index set, but the one that corresponds to all time-consistent sequences $\{\underline{P}^{\gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, that satisfy a given subset of the behavioural assessments that were introduced in the previous sections. For the current paper, an important example of such a set of assessments would be to impose an identical, given marginal model $\underline{P}$ and make structural assessments of exchangeability and forward irrelevance. As we have just explained, the sequence of lower envelopes $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$ will then satisfy that same set of assessments, implying the existence of some $\gamma_* \in \Gamma$ such that $\underline{P}^{\gamma_*}_{\downarrow n} = \underline{P}^{\Gamma}_{\downarrow n}$ for all $n \in \mathbb{N}$. This means that amongst all the time-consistent sequences $\{\underline{P}^{\gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, that are compatible with some chosen set of behavioural assessments, there is a unique sequence that is at most as committal, or at least as conservative, as all the others. It is equal to the sequence of lower envelopes $\{\underline{P}^{\Gamma}_{\downarrow n}\}_{n\in\mathbb{N}}$, and we call it the *least committal*, or *most conservative*, sequence of coherent lower previsions that is compatible with the assessments. If no additional information is available, then this most conservative model is the only one that is implied by a given set of behavioural assessments. Other models might be compatible with, but will never be a consequence of the assessments only, because using these models would mean adding commitments (dispositions to buy gambles) that are not implied by the assessments.

4. Representation in terms of polynomials

In the precise case, as mentioned in the Introduction, a stochastic process is exchangeable if and only if it is a convex mixture of iid processes. This well-known representation theorem by de Finetti also has an imprecise generalisation [13], which we will make extensive use of in the remainder of this paper. Stating it requires the introduction of some more notation though. We begin by explaining the concept of a polynomial gamble on the simplex.

4.1. Polynomial gambles on the simplex

So far, all of the gambles in this paper were defined on finite spaces of the form $\Omega = \mathcal{X}^n$, with $\mathcal{X}$ being a non-empty finite set, representing the values that the individual variables in the sequence $X_1, \dots, X_n, \dots$ can assume. We now introduce gambles that are defined on the so-called $\mathcal{X}$-simplex

$$\Sigma := \left\{\theta \in \mathbb{R}^{\mathcal{X}} : \sum_{x\in\mathcal{X}} \theta_x = 1 \text{ and } \theta_x \geq 0 \text{ for all } x \in \mathcal{X}\right\},$$

which is clearly not a finite set. A gamble f on Σ is a bounded map from Σ to $\mathbb{R}$, associating a value $f(\theta)$ with every $\theta \in \Sigma$. The set of all gambles on Σ is denoted by $\mathcal{G}(\Sigma)$.

As a special case, we consider the linear subspace of the so-called *polynomial gambles* on Σ, which are the restrictions to Σ of polynomials on $\mathbb{R}^{\mathcal{X}}$. The set of all polynomial gambles on Σ is denoted by $\mathcal{V}(\Sigma)$ and is related to the set $\mathcal{V}(\mathbb{R}^{\mathcal{X}})$ of all polynomials on $\mathbb{R}^{\mathcal{X}}$ in the following way. For all $h \in \mathcal{G}(\Sigma)$, we have that

$$h \in \mathcal{V}(\Sigma) \Leftrightarrow (\exists\, p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}}))(\forall \theta \in \Sigma)\ h(\theta) = p(\theta). \tag{9}$$

To see why a polynomial gamble is indeed a gamble on Σ, it suffices to see that polynomials are continuous and therefore bounded on the compact set Σ. It then follows from Eq. (9), that polynomial gambles are bounded maps from Σ to $\mathbb{R}$ and therefore indeed gambles on Σ.

We will use p^* to denote the restriction to Σ of a polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$. In this way, a gamble h on Σ is a polynomial gamble if and only if there is some $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$ for which $h = p^*$. We call p a representing polynomial of h. For any $h \in \mathcal{V}(\Sigma)$, the set of all polynomials that represent h is given by $\mathcal{V}(h) := \{p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}}) : h = p^*\}$. We define the degree of a polynomial gamble $h \in \mathcal{V}(\Sigma)$ as $\deg(h) := \min\{\deg(p) : p \in \mathcal{V}(h)\}$, which is the smallest degree of all polynomials that represent h. If we denote, for every $n \in \mathbb{N}$, the set of all polynomials of degree up to n as $\mathcal{V}_n(\mathbb{R}^{\mathcal{X}})$, then the set

$$\mathcal{V}_n(\Sigma) := \{h \in \mathcal{V}(\Sigma) : \deg(h) \leq n\} = \{p^* : p \in \mathcal{V}_n(\mathbb{R}^{\mathcal{X}})\}$$

is the subset of $\mathcal{V}(\Sigma)$ that contains all polynomial gambles of degree up to n. It will become clear in the following section that $\mathcal{V}_n(\Sigma)$ is closely related to the set $\mathcal{G}(\mathcal{X}_{\downarrow n})$ of all gambles on $\mathcal{X}_{\downarrow n}$.

4.2. Bernstein basis gambles

For any $x_{\downarrow n}$ in $\mathcal{X}_{\downarrow n}$ and any x in $\mathcal{X}$, we define

$$T_x(x_{\downarrow n}) = T_x(x_1, \ldots, x_n) := |\{k \in \{1, \ldots, n\} : x_k = x\}| \tag{10}$$

to be the number of elements in the tuple $x_{\downarrow n}$ that are equal to x. We use $T(x_{\downarrow n})$ to denote the vector whose components are the $T_x(x_{\downarrow n})$, $x \in \mathcal{X}$. One can easily check that $T(x_{\downarrow n})$ assumes values in the set of *count vectors*

$$\mathcal{N}^n := \left\{ m \in \mathbb{N}_0^{\mathcal{X}} : \sum_{x \in \mathcal{X}} m_x = n \right\}.$$

In this expression, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ is the set of all non-negative integers (including zero). Different sequences $x_{\downarrow n} \in \mathcal{X}_{\downarrow n}$ can lead to the same count vector $m \in \mathcal{N}^n$. We use $[m] := \{x_{\downarrow n} \in \mathcal{X}_{\downarrow n} : T(x_{\downarrow n}) = m\}$ to denote the set of all of them. The total number of sequences that lead to the same $m \in \mathcal{N}^n$ is given by

$$\nu(m) := |[m]| = \frac{n!}{\prod_{x \in \mathcal{X}} m_x!}.$$

Similarly to what we have done for $\mathcal{X}^n$ and Σ, we can also consider gambles on $\mathcal{N}^n$. The set of all of them will be denoted by $\mathcal{G}(\mathcal{N}^n)$. For any gamble $b \in \mathcal{G}(\mathcal{N}^n)$, its value in $m \in \mathcal{N}^n$ is denoted by $b(m)$. Count vectors can be used to create a special kind of polynomials and their corresponding polynomial gambles. For every count vector $m \in \mathcal{N}^n$, the unique corresponding *Bernstein basis polynomial* B_m is defined by

$$B_m(\theta) := \nu(m) \prod_{x \in \mathcal{X}} \theta_x^{m_x} \text{ for any } \theta \in \mathbb{R}^{\mathcal{X}}.$$

We will call its restriction to the $\mathcal{X}$-simplex Σ a *Bernstein basis gamble* and denote it as B_m^*. We use $\mathcal{B}_n := \{B_m^* : m \in \mathcal{N}^n\}$ to denote the set of all Bernstein basis gambles of degree n and $\mathcal{B} := \cup_{n \in \mathbb{N}_0} \mathcal{B}_n$ is the set of all Bernstein basis gambles of arbitrary degree. For $\mathcal{B}_n$, we have the following important property.[2]

[2]This result was already mentioned and used in Refs. [13] and [14], both of which referred to Ref. [15] for a proof. While it is indeed possible to derive Proposition 5 from the results in Ref. [15], the link between them is however not immediate. Ref. [15] establishes a result for Bernstein (basis) *polynomials* on $\mathbb{R}^n$, which are defined by means of barycentric coordinates with respect to a simplex that has $n+1$ vertices, whereas we prove it for Bernstein basis *gambles* on Σ, a simplex in $\mathbb{R}^n$ that has $n = |\mathcal{X}|$ vertices, which are defined by means of standard Euclidean coordinates. We therefore prefer to provide an independent proof of our own in the Appendix.

Proposition 5. *The set $\mathcal{B}_n$ of all Bernstein basis gambles of degree n is (i) a partition of unity and (ii) a basis for the linear space $\mathcal{V}_n(\Sigma)$ of all polynomial gambles on Σ whose degree is at most n. In other words, for every $\theta \in \Sigma$ we find that $\sum_{m\in\mathcal{N}^n} B^*_m(\theta) = 1$ and for every $h \in \mathcal{V}_n(\Sigma)$ there is a unique $b \in \mathcal{G}(\mathcal{N}^n)$ for which*

$$h(\theta) = \sum_{m\in\mathcal{N}^n} b(m)B^*_m(\theta) \text{ for all } \theta \in \Sigma.$$

As a direct consequence, we can derive the following corollaries.

Corollary 6. *The set $\mathcal{B}$ of all Bernstein basis gambles of arbitrary degree spans the linear space $\mathcal{V}(\Sigma)$ of all polynomial gambles on Σ.*

Corollary 7. *Every polynomial gamble $h \in \mathcal{V}(\Sigma)$ has a unique corresponding homogeneous polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$ of degree $\deg(p) = \deg(h)$, for which $h = p^*$. In other words, every polynomial gamble has a unique homogeneous polynomial of the same degree that represents it.*

To conclude this section, we will justify our earlier statement that $\mathcal{V}_n(\Sigma)$ is closely related to the set $\mathcal{G}(\mathcal{X}_{\downarrow n})$ of all gambles on $\mathcal{X}_{\downarrow n}$. We start from a gamble $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and construct an associated gamble $b_f \in \mathcal{G}(\mathcal{N}^n)$ by defining $b_f(m)$ to be the uniform average of f over $[m]$:

$$b_f(m) := \frac{1}{\nu(m)} \sum_{x_{\downarrow n}\in[m]} f(x_{\downarrow n}) \text{ for all } m \in \mathcal{N}^n.$$

Next, we use b_f to construct a polynomial gamble $\mathrm{Mn}_n(f)$ by defining for all $\theta \in \Sigma$:

$$\mathrm{Mn}_n(f)(\theta) := \sum_{m\in\mathcal{N}^n} b_f(m)B^*_m(\theta) = \sum_{x_{\downarrow n}\in\mathcal{X}_{\downarrow n}} f(x_{\downarrow n}) \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow n})}, \tag{11}$$

which is the so-called multinomial expectation of the gamble f associated with the mass function θ. In this way, Mn_n becomes a linear map from $\mathcal{G}(\mathcal{X}_{\downarrow n})$ to $\mathcal{V}_n(\Sigma)$. It should be obvious that you can always find, for any $b \in \mathcal{G}(\mathcal{N}^n)$, a gamble $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ for which $b = b_f$. If we combine this with Proposition 5, it follows that $\mathcal{V}_n(\Sigma)$ is equal to $\{\mathrm{Mn}_n(f) : f \in \mathcal{G}(\mathcal{X}_{\downarrow n})\}$. In other words, the operator Mn_n maps $\mathcal{G}(\mathcal{X}_{\downarrow n})$ onto $\mathcal{V}_n(\Sigma)$.

4.3. Representation theorem for exchangeable sequences

The reason why polynomial gambles are useful to us, is because we can define a coherent lower prevision $\underline{R}$ on them that allows us to easily represent an exchangeable sequence $X_1, \dots, X_n, \dots$ of variables assuming values in some finite non-empty set $\mathcal{X}$. Since the set $\mathcal{V}(\Sigma)$ of all polynomial gambles on Σ is a linear subspace of the space of all gambles on Σ, coherence of a lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ can be expressed by imposing requirements P1–P3. The following rather immediate result shows how every such coherent lower prevision on $\mathcal{V}(\Sigma)$ determines a corresponding time-consistent sequence of exchangeable lower previsions.

Proposition 8. *Consider a coherent lower prevision $\underline{R}$ on the linear space $\mathcal{V}(\Sigma)$ of all polynomial gambles on the $\mathcal{X}$-simplex and, for all $n \in \mathbb{N}$, let $\underline{P}_{\downarrow n} : \mathcal{G}(\mathcal{X}_{\downarrow n}) \to \mathbb{R}$ be defined by*

$$\underline{P}_{\downarrow n}(f) := \underline{R}(\mathrm{Mn}_n(f)) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}).$$

Then $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$ is a time-consistent sequence of exchangeable coherent lower previsions.

What is far from immediate, however, is that the converse is true as well.

Theorem 9 (Representation theorem [13, Theorem 5])**.** *Given a time-consistent sequence of exchangeable coherent lower previsions $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n \in \mathbb{N}$, there is a unique coherent lower prevision $\underline{R}$ on the linear space $\mathcal{V}(\Sigma)$ of all polynomial gambles on the $\mathcal{X}$-simplex such that for all $n \in \mathbb{N}$:*

$$\underline{P}_{\downarrow n}(f) = \underline{R}(\mathrm{Mn}_n(f)) \quad \textit{for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}). \tag{12}$$

In other words, each of the coherent lower previsions $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$ is completely determined by the restriction to $\mathcal{V}_n(\Sigma)$ of a coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$.

At first, this might seem like a strange result. $\mathcal{V}_n(\Sigma)$ is a much lower-dimensional space than $\mathcal{G}(\mathcal{X}_{\downarrow n})$, and one would therefore expect to lose information when transforming a gamble $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ to a polynomial gamble $\mathrm{Mn}_n(f) \in \mathcal{V}_n(\Sigma)$. In order to get an intuitive understanding of why this theorem is not that surprising, recall from Section 3.2 that the exchangeability of the variables $X_1, \dots, X_n$ implies that they are also *permutable* [6, Section 9.4], meaning that $\underline{P}_{\downarrow n}(\pi f) = \underline{P}_{\downarrow n}(f)$, for any $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\pi \in \mathcal{P}_n$. This serves as a nice illustration that $\underline{P}_{\downarrow n}$ contains redundant information. Furthermore, it also suggests that it should be possible to replace $\underline{P}_{\downarrow n}$ with an operator on a lower-dimensional space that does not distinguish between f and πf. Since Mn_n is invariant under taking permutations, meaning that the polynomial gambles $\mathrm{Mn}_n(f)$ and $\mathrm{Mn}_n(\pi f)$ are identical, using an operator on polynomial gambles seems like a suitable choice. Proposition 8 and Theorem 9 establish that this is indeed the case. Loosely speaking, the symmetry imposed by exchangeability renders it possible to represent $\{\underline{P}_{\downarrow n}\}_{n \in \mathbb{N}}$ in the lower-dimensional space of polynomial gambles on the simplex Σ.

Finally, for readers that are not closely familiar with lower previsions, it may not be immediate that Theorem 9 generalises de Finetti's representation theorem. That this is indeed the case can be seen by noting that for a precise stochastic process, $\underline{P}_{\downarrow n}$ and $\underline{R}$ become linear previsions that can both be interpreted as expectation operators. The expectation operator that corresponds to $\underline{R}$ can then be regarded as an integral with respect to a prior on θ. Now recall that $\mathrm{Mn}_n(f)(\theta)$ is the multinomial expectation of f for the mass function θ. It then follows that in the precise case, Equation (12) states that the expectation of f is a convex mixture of multinomial expectations.

5. A behavioural approach to multinomial processes

With the tools of the previous two sections in hand, we are now ready to start constructing a behaviourally justified notion of an imprecise multinomial process. We do so by means of three defining properties. The countable sequence $X_1, \dots, X_n, \dots$ of variables, modelled by means of a time-consistent sequence of coherent lower previsions $\underline{P}^{\mathrm{beh}}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n \in \mathbb{N}$, should (i) be exchangeable, (ii) have a given lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ as its marginal model and (iii) be forward irrelevant. As was explained in Section 3, each of these requirements is a behavioural assessment, meaning that it can be expressed purely in terms of the behaviour of a subject: supremum buying prices for gambles. Therefore, using these requirements as defining properties has a behavioural meaning.

However, these assessments do not necessarily lead to a unique multinomial process. There might be multiple models that satisfy requirements (i), (ii) and (iii). Therefore, we introduce a fourth requirement: (iv) our behavioural imprecise multinomial process should be the least committal, most conservative model to satisfy the three requirements above. As explained in Section 3.5, this least committal model does indeed exist and, if no additional judgements are made, it is the only one that is implied by our assessments alone.

In order to arrive at this least committal model, which we will eventually do in Section 5.5, it will prove useful to investigate the consequences of requirements (i), (ii), (iii), when they are imposed on an arbitrary time-consistent sequence of coherent lower previsions $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$.

5.1. Consequences of exchangeability

We start by imposing requirement (i). Such an assessment of exchangeability allows us to use Theorem 9 to represent the exchangeable sequence $X_1,\dots,X_n,\dots$ by means of a coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ instead of a time consistent sequence of exchangeable models $\underline{P}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n\in\mathbb{N}$.

The hard step is now to impose requirements (ii) and (iii) on $\underline{R}$ and to use these to determine the coherent lower prevision $\underline{R}^{\mathrm{beh}}$ on $\mathcal{V}(\Sigma)$ that corresponds to the sequence $\{\underline{P}^{\mathrm{beh}}_{\downarrow n}\}_{n\in\mathbb{N}}$, which is the unique sequence that satisfies requirement (iv) as well. This is the most technical part of this paper and constitutes the next four subsections.

5.2. Consequences of the marginal model

For exchangeable sequences, the requirement (ii) of identically distributed variables with a given marginal model $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ can be expressed very elegantly. As follows from Eq. (5), it suffices to require that $\underline{P}_{\downarrow 1}$ is equal to $\underline{P}$. If we translate this into a property of the representing coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$, we obtain that

$$\underline{R}(\mathrm{Mn}_1(f))=\underline{P}(f)\text{ for all }f\in\mathcal{G}(\mathcal{X}).\tag{13}$$

In this expression, the polynomial gamble $\mathrm{Mn}_1(f)$ is very simple. It follows from Eq. (11) that

$$\mathrm{Mn}_1(f)(\theta)=\sum_{x\in\mathcal{X}}f(x)\,\theta_x\text{ for all }f\in\mathcal{G}(\mathcal{X})\text{ and }\theta\in\Sigma.\tag{14}$$

We see that it is a linear polynomial gamble that has the components of the gamble f on $\mathcal{X}$ as its coefficients. By Proposition 5, this representation is unique: a linear polynomial gamble $h\in\mathcal{V}_1(\Sigma)$ has a unique gamble $f=(\mathrm{Mn}_1)^{-1}(h)\in\mathcal{G}(\mathcal{X})$ for which $h=\mathrm{Mn}_1(f)$. As a consequence, we also find that

$$\mathcal{V}_1(\Sigma)=\{\mathrm{Mn}_1(f):f\in\mathcal{G}(\mathcal{X})\},\tag{15}$$

which is a special case of a result that was already mentioned at the end of Section 4.2.

Next, we introduce an alternative representation for the local model. With any given coherent lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ there corresponds a unique closed and convex subset

$$\begin{aligned}\mathcal{M}:=&\big\{\theta\in\Sigma:(\forall f\in\mathcal{G}(\mathcal{X}))\,\mathrm{Mn}_1(f)(\theta)\geq\underline{P}(f)\big\}\\=&\Big\{\theta\in\Sigma:(\forall f\in\mathcal{G}(\mathcal{X}))\sum_{x\in\mathcal{X}}f(x)\,\theta_x\geq\underline{P}(f)\Big\}\end{aligned}$$

of Σ that is called its *credal set*. It is unique [6, Section 3.3] in the sense that it is the only closed and convex subset of the $\mathcal{X}$-simplex Σ that can reproduce the original coherent lower prevision $\underline{P}$ by defining

$$\underline{P}(f) = \min\{\mathrm{Mn}_1(f)(\theta) : \theta \in \mathcal{M}\} \tag{16}$$

$$= \min\left\{\sum_{x\in\mathcal{X}} f(x)\,\theta_x : \theta \in \mathcal{M}\right\} \text{ for all } f \in \mathcal{G}(\mathcal{X}). \tag{17}$$

An important subset of $\mathcal{V}_1(\Sigma)$ that can be derived from the credal set $\mathcal{M}$ is

$$\begin{aligned}\mathcal{H}_1 :=& \{h \in \mathcal{V}_1(\Sigma) : h(\theta) \geq 0 \text{ for all } \theta \in \mathcal{M}\}\\ =& \{\mathrm{Mn}_1(f) : f \in \mathcal{G}(\mathcal{X}) \text{ and } \underline{P}(f) \geq 0\},\end{aligned} \tag{18}$$

where the second equality is a direct consequence of Eqs. (15) and (16). This set contains all linear polynomial gambles that are non-negative over $\mathcal{M}$. By combining Eq. (13) with Eqs. (15) and (18), we find that for all $h \in \mathcal{V}_1(\Sigma)$

$$h \in \mathcal{H}_1 \Leftrightarrow \underline{R}(h) \geq 0. \tag{19}$$

5.3. Consequences of forward irrelevance

Due to Corollary 4, we can impose forward irrelevance (requirement (iii)) by demanding that the sequence of coherent lower previsions $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$ should satisfy Eq. (8). In order to translate this into a property of the representing coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$, consider the following chain of equivalences. For all $k \in \mathbb{N}$, $f \in \mathcal{G}(\mathcal{X}_{k+1})$ and $x_{\downarrow k} \in \mathcal{X}_{\downarrow k}$ it holds that

$$\begin{aligned}\underline{P}_{\downarrow k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]) = 0 &\Leftrightarrow \underline{R}(\mathrm{Mn}_{k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)])) = 0\\ &\Leftrightarrow \nu(T(x_{\downarrow k}))\underline{R}(\mathrm{Mn}_{k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)])) = 0\\ &\Leftrightarrow \underline{R}(\nu(T(x_{\downarrow k}))\,\mathrm{Mn}_{k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)])) = 0,\end{aligned}$$

where the first equivalence is a consequence of Eq. (12) and the third is due to coherence (P2). In its current form, the argument of $\underline{R}$ is not very intuitive, but we can use Eq. (11) to show (see Appendix A) that for all $k \in \mathbb{N}$, $x_{\downarrow k} \in \mathcal{X}_{\downarrow k}$ and $f \in \mathcal{G}(\mathcal{X}_{k+1})$

$$\nu(T(x_{\downarrow k}))\,\mathrm{Mn}_{k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]) = B^*_{T(x_{\downarrow k})}[\mathrm{Mn}_1(f) - \underline{P}(f)]. \tag{20}$$

Therefore, when stated directly in terms of the coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$, Eq. (8) becomes equivalent to demanding that for any $k \in \mathbb{N}$

$$\underline{R}(B^*_{T(x_{\downarrow k})}[\mathrm{Mn}_1(f) - \underline{P}(f)]) = 0 \text{ for all } x_{\downarrow k} \in \mathcal{X}^k \text{ and } f \in \mathcal{G}(\mathcal{X}). \tag{21}$$

We can use this property to derive that for any $k \in \mathbb{N}$, $m \in \mathcal{N}^k$ and $h \in \mathcal{H}_1$

$$\underline{R}(B^*_m h) = \underline{R}(B^*_{T(x_{\downarrow k})}\,\mathrm{Mn}_1(f)) \geq \underline{R}(B^*_{T(x_{\downarrow k})}[\mathrm{Mn}_1(f) - \underline{P}(f)]) = 0. \tag{22}$$

For the first equality, choose any $x_{\downarrow k} \in [m]$ and apply Eq. (18) to see that $h = \mathrm{Mn}_1(f)$ for some $f \in \mathcal{G}(\mathcal{X})$ such that $\underline{P}(f) \geq 0$. The inequality then follows from coherence property P4 and the last equality is due to Eq. (21).

In summary, we have found that requirements (ii) and (iii) imply that for all $k \in \mathbb{N}_0$, $m \in \mathcal{N}^k$ and $h \in \mathcal{H}_1$

$$\underline{R}(B_m^* h) \geq 0, \tag{23}$$

where the case $k \in \mathbb{N}$ corresponds to Eq. (22) and the special case $k = 0$ follows from Eq. (19). This equation is deceivingly simple, but it has great consequences, since it allows us to prove the following result, where we use $\mathcal{V}_{>0}(\Sigma)$ to denote all polynomial gambles that are strictly positive over the simplex Σ.

Proposition 10. *Consider any coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ that satisfies Eq. (23). Then it holds for all $h \in \mathcal{H}_1$ and $g \in \mathcal{V}_{>0}(\Sigma)$ that $\underline{R}(hg) \geq 0$.*

5.4. Putinar's Positivstellensatz and its consequences

Next, we want to recall some results on the positivity of polynomials over compact convex subsets of $\mathbb{R}^{\mathcal{X}}$. Let us start with *polyhedral convex sets*, which are the subsets of $\mathbb{R}^{\mathcal{X}}$ that can be expressed as the intersection of some finite collection of closed half-spaces. In other words, a subset P of $\mathbb{R}^{\mathcal{X}}$ is polyhedral if there is some finite subset $S = \{q_1, \ldots, q_s\}$ of $\mathcal{V}_1(\mathbb{R}^{\mathcal{X}})$ for which $P = K_S$, with

$$K_S := \left\{\theta \in \mathbb{R}^{\mathcal{X}} : q_i(\theta) \geq 0 \text{ for all } i \in \{1, \ldots, s\}\right\}.$$

We denote the set of all finite sums of squares of polynomials on $\mathbb{R}^{\mathcal{X}}$ as $\sum \mathcal{V}^2(\mathbb{R}^{\mathcal{X}})$ and use it to define the so-called quadratic module of S as

$$M_S := \left\{\sigma_0 + q_1\sigma_1 + \ldots + q_s\sigma_s : \sigma_i \in \sum \mathcal{V}^2(\mathbb{R}^{\mathcal{X}}) \text{ for all } i \in \{0, \ldots, s\}\right\}.$$

It satisfies the following very strong property, which is a special case of Putinar's Positivstellensatz.

Theorem 11 ([16, Theorem 5.6.1 and 7.1.3])**.** *Consider any finite subset $S = \{q_1, \ldots, q_s\}$ of linear polynomials in $\mathcal{V}_1(\mathbb{R}^{\mathcal{X}})$. If the corresponding polyhedral convex set K_S is compact, then for every polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$*

$$p > 0 \text{ on } K_S \Rightarrow p \in M_S.$$

By exploiting the uniform continuity of polynomials over compact sets, we can use Theorem 11 to prove a similar result that holds for general compact convex sets, and not only for the special subset of those that are also polyhedral.

Proposition 12. *Consider any non-empty convex and compact subset C of $\mathbb{R}^{\mathcal{X}}$. For every polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$ that is strictly positive over C there is a finite subset $S = \{q_1, \ldots, q_s\}$ of linear polynomials in $\mathcal{V}_1(\mathbb{R}^{\mathcal{X}})$ for which p is an element of M_S and K_S is a compact superset of C.*

As a corollary, we can derive a useful result for polynomial gambles.

Corollary 13. *Consider any non-empty closed and convex subset $\mathcal{M}$ of Σ. Every polynomial gamble $h \in \mathcal{V}(\Sigma)$ that is strictly positive over $\mathcal{M}$ can be written as a finite sum*

$$h = \sum_{i=0}^{s} h_i g_i, \text{ with } h_i \in \mathcal{H}_1 \text{ and } g_i \in \mathcal{V}_{>0}(\Sigma) \text{ for all } i \in \{0, \ldots, s\}.$$

This corollary enables us to strengthen the consequences of requirement (ii) and (iii) even further. By combining it with Proposition 10 and the properties that follow from coherence, we can derive the following theorem, which will be of crucial importance to prove our main result further on.

Theorem 14. *Consider any coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ that satisfies Eq.* (23)*. Then for every polynomial gamble $h \in \mathcal{V}(\Sigma)$ it holds that*

$$h \geq 0 \text{ on } \mathcal{M} \Rightarrow \underline{R}(h) \geq 0.$$

5.5. Constructing the final model

We are now in the possession of all tools necessary to construct the unique time-consistent sequence of coherent lower previsions $\underline{P}^{\text{beh}}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n \in \mathbb{N}$, that (i) is exchangeable, (ii) has a given lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ as its marginal model, (iii) is forward irrelevant and (iv) is the point-wise smallest of all sequences that satisfy the previous three requirements. For every $n \in \mathbb{N}$ and all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$, it is given by

$$\underline{P}^{\text{beh}}_{\downarrow n}(f) := \min\{\text{Mn}_n(f)(\theta) : \theta \in \mathcal{M}\}. \tag{24}$$

Because of exchangeability and Proposition 1, $\underline{P}^{\text{beh}}_{\downarrow n}$ provides a model for any subset of n variables out of the countable sequence $X_1, \ldots, X_n, \ldots$, and not only for the specific subset $X_1, \ldots, X_n$. The rest of this section will be devoted to the actual proof that the model that is given by Eq. (24) indeed satisfies all of the properties that were stated above.

We start with the requirement that $\{\underline{P}^{\text{beh}}_{\downarrow n}\}_{n\in\mathbb{N}}$ should be a time-consistent exchangeable sequence of coherent lower previsions. That this is indeed the case follows immediately from Proposition 8, using the trivial coherent lower prevision $\underline{R}^{\text{beh}}$ on $\mathcal{V}(\Sigma)$, defined by

$$\underline{R}^{\text{beh}}(h) := \min\{h(\theta) : \theta \in \mathcal{M}\} \text{ for all } h \in \mathcal{V}(\Sigma). \tag{25}$$

In order for the models $\underline{P}^{\text{beh}}_{\downarrow n}$, $n \in \mathbb{N}$, to additionally marginalise to a given lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ and be forward irrelevant, it suffices for $\underline{R}^{\text{beh}}$ to satisfy Eqs. (13) and (21), as explained in Section 5.2 and 5.3. It is proven in the following proposition that this is indeed the case.

Proposition 15. *The coherent lower prevision $\underline{R}^{\text{beh}}$ on $\mathcal{V}(\Sigma)$ that is defined by Eq.* (25) *satisfies Eqs.* (13) *and* (21)*.*

At this point, we know that the time-consistent sequence of coherent lower previsions $\{\underline{P}^{\text{beh}}_{\downarrow n}\}_{n\in\mathbb{N}}$, that is given by Eq. (24) satisfies requirements (i), (ii) and (iii). All that is left to prove is (iv) that this sequence is the unique point-wise smallest one that does so. This is established in Theorem 17, which we consider to be the main technical result of this paper. It is a rather direct consequence of Proposition 15 and the following corollary to Theorem 14.

Corollary 16. *Consider any coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ that satisfies Eq.* (23)*. Then $\underline{R}^{\text{beh}}$, given by Eq.* (25)*, is point-wise dominated by $\underline{R}$:*

$$\underline{R}^{\text{beh}}(h) \leq \underline{R}(h) \text{ for all } h \in \mathcal{V}(\Sigma).$$

Theorem 17. *Consider any time-consistent sequence of coherent lower previsions $\underline{P}_{\downarrow n}$, $n \in \mathbb{N}$, that is (i) exchangeable, (ii) identically distributed with a given marginal model $\underline{P}$ and (iii) forward irrelevant, then for every $n \in \mathbb{N}$*

$$\underline{P}^{\mathrm{beh}}_{\downarrow n}(f) \leq \underline{P}_{\downarrow n}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}).$$

Therefore, the time-consistent sequence of coherent lower previsions $\underline{P}^{\mathrm{beh}}_{\downarrow n}$, $n \in \mathbb{N}$, as defined by Eq. (24)*, is the unique least committal, most conservative one to satisfy (i), (ii) and (iii).*

6. The sensitivity analysis approach and how to justify it

After this extensive detour through the behavioural theory of imprecise probabilities, it is now time to focus on the actual topic of this paper: how can we justify the sensitivity analysis approach to imprecise multinomial processes? It will turn out that the rather technical results that were developed in the previous sections enable us to provide this problem with an elegant solution. But first, let us explain in more detail what we actually mean by this sensitivity analysis approach.

6.1. The sensitivity analysis approach to multinomial processes

The starting point of the sensitivity analysis approach to imprecise multinomial process, is that the given local model $\underline{P}$ can be interpreted in a non-behavioural way as well.

As we showed in Section 5.2, such a coherent lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ can be uniquely characterised by its corresponding credal set $\mathcal{M}$, which is a closed and convex subset of the $\mathcal{X}$-simplex Σ. At that point, we regarded this simplex as a purely mathematical concept, but it should be obvious from its definition that it is the set of all probability mass functions on $\mathcal{X}$. Therefore, if X is a variable that assumes values in the non-empty finite set $\mathcal{X}$, then for any $\theta \in \Sigma$ and $x \in \mathcal{X}$, we can interpret θ_x as the probability that X assumes the value x. This means that the credal set $\mathcal{M}$ is a (closed and convex) set of probability mass functions on $\mathcal{X}$. Due to the uniqueness that was mentioned in Section 5.2, it serves as a mathematically equivalent representation for $\underline{P}$.

Every probability mass function θ in $\mathcal{M}$ has a corresponding expectation operator E_θ. For any gamble $f \in \mathcal{G}(\mathcal{X})$, it provides us with an expected value, given by

$$E_\theta(f) = \sum_{x \in \mathcal{X}} f(x)\theta_x.$$

The lower bound for this expected value, taken over all $\theta \in \mathcal{M}$, is given by

$$\underline{E}_{\mathcal{M}}(f) = \min\{E_\theta(f) : \theta \in \mathcal{M}\} = \underline{P}(f),$$

where we have used Eq. (17) to obtain the last equality. We conclude that the relation between $\mathcal{M}$ and $\underline{P}$ is not only of a mathematical nature. For all $f \in \mathcal{G}(\mathcal{X})$, $\underline{P}(f)$ can be interpreted as the minimal expected value of f, taken over all probability mass functions θ in the credal set $\mathcal{M}$ that corresponds to $\underline{P}$.

In light of this alternative representation $\mathcal{M}$, one does not necessarily need to regard $\underline{P}$ as an inherently imprecise behavioural model for some subject's beliefs about a variable

X, as we have done earlier on in this paper. Instead, one can assume that X can be modelled by means of a single probability mass function p, interpreting the credal set $\mathcal{M}$ that corresponds to $\underline{P}$ as a set of possible candidates θ for p. Walley calls this the *sensitivity analysis interpretation* [6, Section 1.1.5].

If we extend this sensitivity analysis interpretation to imprecise multinomial processes, then p should be regarded as the marginal model of a precise multinomial process, which is identically distributed and satisfies stochastic independence. Since under the sensitivity analysis interpretation, p is only known to be an element of $\mathcal{M}$, this leaves us with a set of possible precise multinomial processes, one for every candidate marginal model θ in $\mathcal{M}$. The corresponding imprecise multinomial process is then taken to be the convex hull of this set of precise ones. We will refer to this method of constructing an imprecise multinomial process as the *sensitivity analysis approach*.

Let us make this clearer by focussing on the first n variables $X_1, \dots, X_n$ of this imprecise multinomial process. For each $\theta \in \mathcal{M}$, we can use the assumption of stochastic independence to derive a joint probability mass function p_θ^n on $\mathcal{X}_{\downarrow n}$, defined for all $x_{\downarrow n} \in \mathcal{X}_{\downarrow n}$ by

$$p_\theta^n(x_{\downarrow n}) := \prod_{i=1}^{n} \theta(x_i). \tag{26}$$

The imprecise model for the variables $X_1, \dots, X_n$ is then taken to be the credal set

$$\mathcal{M}_{\downarrow n}^{\mathrm{t2}} := \mathrm{CH}(\{p_\theta^n : \theta \in \mathcal{M}\}), \tag{27}$$

which is the convex hull of the set of probability mass functions p_θ^n on $\mathcal{X}_{\downarrow n}$, $\theta \in \mathcal{M}$.

For any gamble $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$, such a (closed and convex) credal set of joint probability mass functions provides us with a range of expected values. Due the linearity of the expectation operator, this range will be closed and convex as well, and we can therefore fully represent it by means of its lower bound $\underline{E}_{\mathcal{M}}^n(f)$ and upper bound $\overline{E}_{\mathcal{M}}^n(f)$. Since, again due to the linearity of the expectation operator, these bounds are not affected by including convex combinations, they are given for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ by

$$\underline{E}_{\mathcal{M}}^n(f) := \min\{E_\theta^n(f) : \theta \in \mathcal{M}\} \text{ and } \overline{E}_{\mathcal{M}}^n(f) := \max\{E_\theta^n(f) : \theta \in \mathcal{M}\}, \tag{28}$$

where E_θ^n is the expectation operator that corresponds to p_θ^n. It is defined for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ by

$$E_\theta^n(f) := \sum_{x_{\downarrow n} \in \mathcal{X}_{\downarrow n}} f(x_{\downarrow n}) p_\theta^n(x_{\downarrow n}). \tag{29}$$

The lower and upper expected value are related through conjugacy: we have for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ that

$$\overline{E}_{\mathcal{M}}^n(f) = \max\{E_\theta^n(f) : \theta \in \mathcal{M}\} = -\min\{E_\theta^n(-f) : \theta \in \mathcal{M}\} = -\underline{E}_{\mathcal{M}}^n(-f).$$

Therefore, one can focus on the lower expected values, combining them with conjugacy to derive the upper ones.

Although it might seem as if the sensitivity analysis approach has little to do with Walley's theory of coherent lower previsions, the corresponding lower expected values do in fact provide us with a coherent lower prevision $\underline{P}_{\downarrow n}^{\mathrm{t2}}$. We define it by

$$\underline{P}_{\downarrow n}^{\mathrm{t2}}(f) := \underline{E}_{\mathcal{M}}^n(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}). \tag{30}$$

Readers that are familiar with imprecise-probabilistic notions of independence might recognise $\underline{P}^{\mathrm{t2}}_{\downarrow n}$ as the so-called *type-2 product* [6, Section 9.3.5] of n identically distributed variables $X_1, \ldots, X_n$ that have $\underline{P}$ as their marginal lower prevision, or equivalently, $\mathcal{M}$ as their marginal credal set. Due to the one-to-one correspondence between coherent lower previsions and (closed and convex) credal sets, $\underline{P}^{\mathrm{t2}}_{\downarrow n}$ serves as a mathematically equivalent representation for the (closed and convex) joint credal set $\mathcal{M}^{\mathrm{t2}}_{\downarrow n}$, as given by Eq. (27). By constructing this type-2 product for every $n \in \mathbb{N}$, we obtain a time-consistent[3] sequence of coherent lower previsions $\underline{P}^{\mathrm{t2}}_{\downarrow n}$ on $\mathcal{G}(\mathcal{X}_{\downarrow n})$, $n \in \mathbb{N}$, which serves as a mathematical representation for the sensitivity analysis approach to imprecise multinomial processes.

6.2. Problems with its interpretation and a partial solution

The main problem with the sensitivity analysis approach to imprecise multinomial processes, or for that matter, with the type-2 product, is that it combines stochastic independence with convexity. In this way, convex combinations of stochastically independent distributions are included, which need no longer be stochastically independent themselves.

This is usually justified in the following way. Each of the factorising distributions is taken to correspond to a different expert. These experts all agree on an assessment of stochastic independence, but disagree on specific probability values. Including convex combinations of their individual distributions is then justified by assuming that these will not affect the collective preferences of the experts.

This justification holds if one assumes that collective preferences can be fully represented by means of binary comparisons, because these are not affected by including convex combinations. However, the argument breaks down once one allows for non-binary preferences. Indeed, as pointed out by Kyburg and Pittarelli in Ref. [17], choosing amongst several acts using E-admissibility [18] (a set-valued decision criterion that chooses those acts that maximise expected utility for at least one of the candidate probabilistic models) does depend on whether or not one includes convex combinations. In that same reference, Kyburg and Pittarelli also provide a direct argument against the combination of stochastic independence and convexity. They show that if an agent bets according to a convex combination of product measures, whilst knowing that stochastic independence should hold, he can be made to incur a sure loss in the long run.

In light of these problems, one can take two different routes. The most obvious one is perhaps to conclude that the sensitivity analysis approach to imprecise multinomial processes should drop convexity and that it should use the original non-convex set of precise multinomial processes instead. This approach comes at a cost though. First off, by giving up on convexity, the resulting model can no longer be represented by means of lower previsions. This enforces an interpretation in terms of partially known probabilities and excludes the possibility of adopting a direct interpretation in terms of betting behaviour. Other than that, dropping convexity also has computational consequences because most imprecise probability algorithms are designed for convex sets of probabilities only. Many of them will for example make use of linear programming techniques, which are applicable to convex optimisation problems only.

[3] This is a consequence of stochastic independence being preserved under marginalisation.

The current paper offers a different route, which does not require dropping convexity. In order to avoid the apparent conflict between stochastic independence and convexity, we will justify the sensitivity analysis approach to multinomial processes directly, without even mentioning stochastic independence, let alone assuming it. To our knowledge, the only justification of this kind has been proposed by Cozman[4] [5, Section 3.5]. Actually, Cozman does not speak of imprecise multinomial processes, but instead provides a justification for the type-2 product of a finite number of variables. It should however become clear from the following brief recapitulation of his approach, that he in fact starts by justifying a model for countable sequences (identical to the one that corresponds to the sensitivity analysis approach to imprecise multinomial processes, as defined in the current paper), and that he subsequently uses it to derive his justification for the type-2 product.

The first step of Cozman's justification is to impose exchangeability on a countable sequence $X_1, \ldots, X_n, \ldots$ of discrete variables. He uses the imprecise-probabilistic definition that was given in Section 3.2 to do so, thereby providing this assumption with a behavioural meaning. As noted by Walley [6, Section 9.5.4], such an assumption is equivalent with requiring every element of the corresponding convex set of probability distributions to be exchangeable in the usual de Finetti's sense [19, 20, 9]. Using de Finetti's representation theorem for countable sequences [19, 20, 9, 21], this in turn implies that all these probability distributions should be convex mixtures of precise multinomial processes.

At this point, Cozman has already managed to justify the use of a convex set of mixtures of precise multinomial processes, without invoking stochastic independence, merely by a judgement of exchangeability. He then goes on to impose a second assumption on this convex set. Its extreme points (called vertices by Cozman) should be single multinomial processes, instead of mixtures of them. The other distributions in this set will then be convex mixtures of these particular multinomial processes. The largest[5] set of distributions to satisfy both of Cozman's assumptions and additionally marginalise to a given local credal set $\mathcal{M}$ is identical to the one that is obtained by the sensitivity analysis approach to imprecise multinomial processes. Consequently, if one judges a finite set $X_1, \ldots, X_n$ of variables to be the initial fragment of a countable sequence $X_1, \ldots, X_n, \ldots$ of exchangeable variables, additionally assuming that Cozman's second requirement holds, one obtains the type-2 product.[6] According to Cozman, this provides the type-2 product with a behavioural justification.

While we agree with the first part of this justification, being a judgement of exchangeability, Cozman's second requirement has a conceptual weakness. It seems to be based on the assumption that one should only consider those exchangeable distributions that are mixtures of multinomial processes whose marginal model is contained in the local credal set $\mathcal{M}$. Or in other words, those exchangeable distributions of which the corre-

[4]Similar justifications do have been proposed for other imprecise-probabilistic models that combine stochastic independence with convexity; see Section 7 for more information. However, for the specific case of the sensitivity analysis approach to imprecise multinomial processes and the related type-2 product, Ref. [5, Section 3.5] seems to be the only justification of this kind.

[5]This corresponds to the least committal, most conservative approach of Section 3.5.

[6]For binary variables, this was already mentioned by Walley in Ref. [6, Section 9.5.4]. In contrast with Cozman, Walley seems to regard this as a purely mathematical result, without interpreting it as a justification for the type-2 product.

sponding prior assigns positive probability to elements of $\mathcal{M}$ only. However, this seems to be an arbitrary mathematical assumption that does not have any behavioural ground. Indeed, the credal set $\mathcal{M}$ does not impose any restrictions on the prior, but only on the corresponding predictive marginal models for single variables.

Let us make this clearer by considering the basic example of a sequence of tosses with a fair coin. The marginal model for a single fair coin toss is a credal set $\mathcal{M}$ that consists of only a single probability mass function p, which is given by $p(\text{heads}) = p(\text{tails}) = 1/2$. Therefore, the sensitivity analysis approach to imprecise multinomial process yields a single precise multinomial process, having p as its marginal probability mass function. However, an assumption of exchangeability is not capable of justifying this model. Indeed, under exchangeability, one could just as well use a distribution that assigns probability $1/2$ to an infinite sequence of heads, $1/2$ to an infinite sequence of tails, and 0 to every other possible sequence. This distribution marginalises to the given local model p and is exchangeable because it is a uniform convex mixture of two (trivial) precise multinomial processes that assign probability 1 to heads and tails respectively. Similarly, any other distribution that marginalises to p and is a convex mixture of multinomial processes, is compatible with both p and the assumption of exchangeability. We conclude that, if no additional assumption is made, the only approach that can be justified through an assumption of exchangeability, is to consider the set of all these compatible models. This is the unique largest convex set containing all exchangeable distributions that have p as their marginal model. The corresponding time-consistent sequence of coherent lower previsions is the unique least committal, most conservative one to be exchangeable and have a marginal model $\underline{P}$ that has $\mathcal{M} = \{p\}$ as its corresponding credal set.

It should be clear from this example that Cozman's second requirement has no behavioural justification. We therefore feel that it should be regarded as a purely mathematical assumption. This implies that Cozman's justification for the type-2 product is only partially valid and that the second part of it should be reconsidered, replacing it with an assessment that does have behavioural meaning.

6.3. *Providing a full behavioural justification*

Our proposal is to replace Cozman's second requirement with a judgement of either forward irrelevance or epistemic independence. We will show that by combining either one of these judgements with a judgement of exchangeability, one obtains a full behavioural justification for the sensitivity analysis approach to imprecise multinomial processes and thereby also for the type-2 product. The starting point is the following simple proposition.

Proposition 18. *The time-consistent sequence of coherent lower previsions $\{\underline{P}^{\mathrm{t2}}_{\downarrow n}\}_{n\in\mathbb{N}}$, as given by Eq.* (30)*, and $\{\underline{P}^{\mathrm{beh}}_{\downarrow n}\}_{n\in\mathbb{N}}$, as given by Eq.* (24)*, are identical:*

$$\underline{P}^{\mathrm{beh}}_{\downarrow n}(f) = \underline{P}^{\mathrm{t2}}_{\downarrow n}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}) \text{ and } n \in \mathbb{N}.$$

By combining this proposition with Theorem 17, we immediately obtain one of the most important results of this paper.

Theorem 19. *The time-consistent sequence of coherent lower previsions $\{\underline{P}^{\mathrm{t2}}_{\downarrow n}\}_{n\in\mathbb{N}}$, as defined by Eq.* (30)*, is the unique least committal, most conservative one that is (i) exchangeable, (ii) identically distributed with a given marginal model $\underline{P}$ and (iii) forward irrelevant.*

This result provides us with a first justification for the sensitivity analysis approach to imprecise multinomial processes and the type-2 product. For a given marginal model, the sensitivity analysis approach corresponds to the most conservative imprecise multinomial process that is both exchangeable and forward irrelevant. This justification is fully behavioural because all of the properties that were used to obtain it can be given a behavioural meaning. From it, one can derive a justification for the type-2 product by judging $X_1, \ldots, X_n$ to be an initial fragment of the imprecise multinomial process that is justified by Theorem 19.

Theorem 19 also serves as a starting point to derive the following slightly weaker, but perhaps more intuitive result.

Theorem 20. *Theorem 19 remains valid if 'forward irrelevance' is replaced by 'many-to-one epistemic independence' or 'many-to-many epistemic independence'.*

The main idea of the proof is that many-to-many epistemic independence, many-to-one epistemic independence and forward irrelevance are increasingly weaker imprecise-probabilistic notions of independence, each of which is satisfied by the sensitivity analysis approach.

Theorem 20 provides us with a second behavioural justification for the sensitivity analysis approach. For a given marginal model, it corresponds to the most conservative imprecise multinomial process that is both exchangeable and epistemically independent (either many-to-many or many-to-one). The type-2 product is again justified by judging $X_1, \ldots, X_n$ to be the initial fragment of such a process.

We leave it to the reader to choose amongst these different behavioural justifications. Personally, we prefer the one that is provided by Theorem 19 because it corresponds to the weakest set of assessments. Also, the asymmetric nature of forward irrelevance seems fitting for a process that takes place over time. However, since symmetry is imposed anyway by exchangeability, one could argue that the symmetric notion of epistemic independence is more natural to impose. This seems especially compelling for the type-2 product of a finite amount of variables. In that case, we would prefer to impose many-to-one epistemic independence, again because it is a mathematically weaker assumption then its many-to-many counterpart.

7. What about strong independence?

The sensitivity analysis approach to imprecise multinomial processes and the related type-2 product are not the only imprecise-probabilistic models that combine stochastic independence with convexity, thereby suffering from the lack of interpretation that was outlined in Section 6.2.

The associated imprecise-probabilistic notion of independence that is at stake here is called strong independence.[7] By definition, a collection of variables is said to satisfy *strong independence* [5] if their representing set of probability distributions is the convex

[7] This concept is known under a number of alternative names as well; see Ref. [5, Section 3.1] for an overview.

hull of a set of stochastically independent ones.[8] This should be contrasted with the notion of *complete independence* [23, 5], which requires every probability distribution in this representing set to be stochastically independent. The former is often preferred over the latter for the same reasons that were mentioned in Section 6.2: the convexity that comes with strong independence allows for a representation in terms of lower previsions and enables the use of convex optimisation techniques such as linear programming.

There is also a special case of strong independence, called *repetition independence* [4]. It again consists in considering the convex hull of a set of stochastically independent probability distributions, but the elements of this set are now also required to be identically distributed—making them iid.[9] This special case of repetition independence is what we have been concerned with in this paper. In particular, we have provided behavioural justifications for specific models that satisfy it: for countable sequences $X_1, \ldots, X_n, \ldots$, we justified the use of the sensitivity analysis approach to imprecise multinomial processes, and for finite sequences $X_1, \ldots, X_n$, we justified the use of the type-2 product. Both of these models correspond to using the largest (least committal, most conservative) set of joint distributions that marginalises to a given local credal set $\mathcal{M}$ and satisfies repetition independence. For every other kind of imprecise-probabilistic model that satisfies strong (or repetition) independence, or in other words, that is constructed as the convex hull of stochastically independent distributions, the question of whether or not it can be provided with a behavioural justification remains largely open. The existing literature provides only a few partial answers, all of which seem to be due to Cozman. We give a short overview.

In the context of credal networks, Cozman has proposed to use what he calls the strong Markov condition [24, Section 6]. The main idea is to impose (conditional) epistemic independence and to require this assessment to keep on holding even after the joint model has been altered in order to incorporate a series of belief changes. He shows that by imposing this condition, one obtains a credal network under strong independence, without making an explicit assessment of stochastic independence. By combining this with a least committal strategy, he obtains the so-called strong extension of a credal network. The main weakness of this approach, as Cozman points out himself, is that it expresses belief changes in terms of individual probability distributions, rather than credal sets. Therefore, the strong Markov condition seems to lack a behavioural meaning.

Cozman comes back to this approach in a recent overview paper [5, Section 3.4], where he focuses on the case of two independent variables, rather than on general credal networks. He compares it with a justification that was proposed by Moral and Cano [25, Theorem 2], arguing that both approaches are similar but that the one by Moral and

[8] Strong independence has been used to refer to other imprecise-probabilistic notions of independence as well; we provide an example further on in this section. As another example: Ref. [4] uses independence in the selection to refer to what we call strong independence, and it uses strong independence to refer to the largest set of distributions that satisfies independence in the selection. We prefer to follow Walley in calling this largest set the type-1 product [6, Section 9.3.5], or alternatively, the strong extension [22]; see also Section 8.

[9] For countable sequences $X_1, \ldots, X_n, \ldots$, this extra assumption can also be expressed in terms of exchangeability: a countable sequence satisfies repetition independence if and only if it is both strongly independent and exchangeable. The direct implication in this statement is immediate because every exchangeable stochastic process is trivially iid. The converse one follows because the only way for an exchangeable stochastic process—which, by de Finetti's representation theorem, is a convex mixture of iid processes—to be stochastically independent is for it to be iid.

Cano is to be preferred because it expresses belief changes in terms of credal sets rather than individual probability distributions. Cozman then goes on to present a 'generalised' conditional version of Moral and Cano's result [5, Theorem 1]. Although this approach seems promising, it does suffer from a number of problems.

The first one is that, in contrast with what Cozman seems to suggest, his result is in fact fundamentally different from that by Moral and Cano. The reason is that Moral and Cano adopt a different definition of strong independence than Cozman does. The definition of Moral and Cano corresponds to the use of a specific *non-convex* credal set, consisting of factorising distributions only,[10] whereas Cozman uses the convexified concept that is also adopted in the present paper. Also, Moral and Cano's definition of strong independence corresponds to a least committal strategy,[11] whereas Cozman's adaptation of their result does not restrict attention to this particular case. Therefore, the result by Moral and Cano should not be regarded as a justification for using convex hulls of arbitrary sets of stochastically independent distributions. This has important consequences for Cozman's result as well, since he did not provide [5, Theorem 1] with an explicit proof. Instead, he claimed that such a proof can be obtained by simply following the steps in the proof of [25, Theorem 2]. In light of the different definitions that were adopted in both results, it is not clear whether this is indeed the case.

The second problem is a conceptual weakness that Cozman points out himself [5, Section 3.4]: the argument that is embedded in [5, Theorem 1] only holds if one assumes that new assessments should always be combined with the currently held joint credal set, thereby disregarding the available assessment of epistemic independence. However, one can argue that an assessment of epistemic independence should not be neglected upon receiving new information and should instead be taken into account to construct the new joint credal set. This alternative approach would break the argument.

Finally, Cozman has also presented a fundamentally different way of providing strong independence with a justification [5, Section 3.5.3], which is based on a judgement of partial exchangeability. The approach is very similar to his justification for the type-2 product, as we discussed in detail in Section 6.2. The main idea is to combine an assessment of partial exchangeability with a condition on the set of priors that results from such an assessment. However, similarly to the approach for the type-2 product, this added condition seems to be a purely mathematical assumption that has no behavioural meaning.

8. Conclusions and suggestions for future research

Starting from a given imprecise-probabilistic model for a single variable X (represented by means of a coherent lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$, or equivalently, a credal set $\mathcal{M}$ consisting of probability mass functions on $\mathcal{X}$), we have considered two fundamentally different approaches to constructing an imprecise multinomial process that describes a countable sequence $X_1, \ldots, X_n, \ldots$ of identically distributed variables that may assume a finite number of values.[12]

[10] It is a special case of the notion of complete independence that was mentioned in Section 7.

[11] They use the *largest* joint model that satisfies complete independence.

[12] For an extension of this discussion to allow for variables X_i assuming values in an *infinite* set $\mathcal{X}$, the existing treatments of exchangeability and (forward) irrelevance for infinite domains must receive much more attention than they have so far in the literature on imprecise probabilities.

Our first approach was to define it using behavioural assessments only. We constructed such an imprecise multinomial process in Section 5, defining it as the least-committal, most conservative model that is exchangeable, has $\underline{P}$ as its marginal model, and additionally satisfies forward irrelevance. Later in the paper, in Section 6.3, we showed that the assessment of forward irrelevance is not essential to this definition. It can be replaced by an assessment of either many-to-one or many-to-many epistemic independence as well, without changing the resulting imprecise multinomial process.

In Section 6.1, we presented a second, perhaps more familiar approach. The main idea was to interpret $\mathcal{M}$ as a set of candidates for the unknown marginal probability mass function of a precise multinomial process. This resulted in a set of precise multinomial processes, one for every candidate mass function in $\mathcal{M}$. The corresponding imprecise multinomial process was then taken to be the convex hull of all these precise ones—if a credal set representation is needed—or, equivalently, their lower envelope—if a lower prevision representation is preferred. This so-called sensitivity analysis approach provided us with simple, intuitive expressions. For any $n \in \mathbb{N}$, the model for the variables $X_1, \ldots, X_n$ was given by the so-called type-2 product of our marginal model.

Our main result, as presented in Section 6.3, is that these two approaches are identical. We find this to be very surprising because the sensitivity analysis approach to imprecise multinomial processes starts from totally different premisses than the behavioural one. Sensitivity analysis is in its core a precise approach, stochastic independence being essential to its definition. Imprecision is introduced in a rather arbitrary manner by taking lower and upper envelopes of precise stochastically independent models. By doing so, it allows for convex combinations that are no longer stochastically independent, thereby contradicting its own premisses; see Section 6.2. The behavioural approach is fundamentally different. It is inherently imprecise and constructed from behavioural assessments only, without making any reference to stochastic independence. The fact that both imprecise multinomial processes turn out to be identical joins both worlds, thereby providing the behavioural approach with simple intuitive expressions and the sensitivity analysis approach with a behavioural justification.

Although this behavioural justification is powerful and consistent, the circumstances under which it is applicable are of course open to discussion. For countable sequences, three assessments are needed: forward irrelevance (or epistemic independence), exchangeability and the use of a least committal strategy. To us, this last requirement is the most compelling one. Whatever the assessments made, the resulting model should reflect but those assessments, and nothing more. An assessment of forward irrelevance is less compelling, but we feel as if it should be accepted whenever one judges the variables at hand to be independent in some way. Indeed, of all imprecise-probabilistic notions of independence that one can impose, epistemic irrelevance is one of the weakest, still meaningful notions. It is furthermore implied by almost every other notion of independence, including the popular notions of strong and epistemic independence. Therefore, the main assessment that remains open for discussion is that of exchangeability. As should have become clear from our results, exchangeability is a remarkably strong property. It should not be taken for granted. We refer to Walley [6, Section 9.5.2] for an extensive discussion on sufficient grounds for exchangeability judgements, stating that in order to justify an assessment of exchangeability, it suffices for the individual experiments to be (physically) similar. However, further discussion on this topic would be more than welcome.

If instead of a countable sequence, one only considers a finite set of variables, our behavioural justification for modelling this set by means of the type-2 product needs an additional assumption. This finite set of variables should then be judged to be a subset of a countable sequence and the assessments of exchangeability and forward irrelevance should be imposed on this countable sequence instead of on the original finite set of variables. The circumstances under which such an assumption can be justified are open to discussion as well.

That said, as explained in Section 7, the sensitivity analysis approach to imprecise multinomial processes and the related type-2 product are not the only imprecise-probabilistic models that combine stochastic independence with convexity. Every other so-called strongly independent model does so as well, thereby suffering from the lack of interpretation that was outlined in Section 6.2. An obvious line of future research would therefore be to try and find a behavioural justification for strongly independent models other than the ones treated in the current paper as well. The main idea would be similar: to try and construct them directly, without the use of stochastic independence. Section 7 gave an overview of some previous attempts at doing so, explaining their conceptual weaknesses. We conclude from this overview that the problem of justifying strong independence remains largely open.

A popular imprecise-probabilistic model for which such a justification would be especially welcome, is the so-called *type-1 product* [6, Section 9.3.5], often referred to as the *strong extension* as well. Similarly to the type-2 product, it is a convex hull of stochastically independent distributions. However, the type-1 product does not require these distributions to be identically distributed. Given a marginal model $\mathcal{M}_i$ for every variable X_i in a finite collection $X_1, \dots, X_n$, their type-1 product is defined as the convex hull of the set of all stochastically independent joint probability mass functions that are of the form $\prod_{i=1}^{n} p_i$, where for all $i \in \{1, \dots, n\}$, p_i is selected from $\mathcal{M}_i$. This type-1 product is the largest set of joint probability mass functions that satisfies strong independence and has $\mathcal{M}_i$ as its marginal model for the variable X_i, $i \in \{1, \dots, n\}$.

As a possible avenue for further research, here are some preliminary ideas that could perhaps lead to a justification of this type-1 product. We suggest to use a strategy similar to the one that we used to justify the type-2 product. The starting point would be Cozman's idea of imposing partial exchangeability. By combining such an assessment with an additional rather technical condition, Cozman obtains strong independence. However, as explained in Section 7, this second condition seems to have no behavioural meaning and, therefore, does not provide strong independence with a behavioural justification. We suggest to try and remedy this situation by replacing Cozman's condition with some well-chosen assessments of forward irrelevance or epistemic independence. By combining these assessments with partial exchangeability and the use of a least committal strategy, one might obtain the type-1 product.

Acknowledgements

Most of Jasper De Bock's work on this manuscript was conducted while he was a PhD Fellow of the Research Foundation – Flanders (FWO); he therefore wishes to acknowledge their financial support. Gert de Cooman's research was partially funded by FWO project G.0125.12.

References

[1] D. M. Gabbay, P. Thagard, J. Woods, P. S. Bandyopadhyay, M. R. Forster, Philosophy of Statistics, Handbook of the Philosophy of Science, Elsevier Science, 2011.

[2] T. Dietz, L. Kalof, Introduction to Social Statistics: The Logic of Statistical Reasoning, Wiley, 2009.

[3] D. Vose, Risk Analysis: A Quantitative Guide, John Wiley & Sons, 2008.

[4] I. Couso, S. Moral, P. Walley, A survey of concepts of independence for imprecise probabilities, Risk Decision and Policy 5 (2000) 165–181.

[5] F. G. Cozman, Sets of probability distributions, independence, and convexity, Synthese 186 (2) (2012) 577–600.

[6] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[7] G. de Cooman, E. Miranda, Forward irrelevance, Journal of Statistical Planning and Inference 139 (2) (2009) 256–276. `doi:10.1016/j.jspi.2008.01.012`.

[8] G. de Cooman, E. Miranda, M. Zaffalon, Independent natural extension, Artificial Intelligence 175 (2011) 1911–1950. `doi:10.1016/j.artint.2011.06.001`.

[9] B. de Finetti, Theory of Probability: A Critical Introductory Treatment, John Wiley & Sons, Chichester, 1974–1975, English translation of [26], two volumes.

[10] P. M. Williams, Notes on conditional previsions, Tech. rep., University of Sussex, reprinted in a revised form in the International Journal of Approximate Reasoning 44(3):366-383, 2007 (1975).

[11] G. de Cooman, E. Miranda, Weak and strong laws of large numbers for coherent lower previsions, Journal of Statistical Planning and Inference 138 (8) (2008) 2409–2432. `doi:10.1016/j.jspi.2007.10.020`.

[12] J. De Bock, A. Van Camp, M. A. Diniz, G. de Cooman, Representation theorems for partially exchangeable random variables, Fuzzy Sets and Systems 284 (2016) 1–30.

[13] G. De Cooman, E. Quaeghebeur, E. Miranda, Exchangeable lower previsions, Bernoulli 15 (3) (2009) 721–735, an expanded version can be found at `http://arxiv.org/abs/0801.1265`. `doi:10.3150/09-BEJ182`.
URL `http://hdl.handle.net/1854/LU-498518`

[14] G. de Cooman, E. Quaeghebeur, Exchangeability and sets of desirable gambles, International Journal of Approximate Reasoning 53 (3) (2012) 363–395, special issue in honour of Henry E. Kyburg, Jr.

[15] H. Prautzsch, W. Boehm, M. Paluszny, Bézier and B-Spline Techniques, Springer, Berlin, 2002.

[16] M. Marshall, Positive Polynomials and Sums of Squares, Mathematical Surveys and Monographs, American Mathematical Society, 2008.
URL `http://books.google.be/books?id=Aj5WP_Knu0sC`

[17] H. E. Kyburg, Jr., M. Pittarelli, Set-based bayesianism, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 26 (3) (1996) 324–339. `doi:10.1109/3468.487958`.

[18] I. Levi, The Enterprise of Knowledge, MIT Press, London, 1980.

[19] B. de Finetti, La prévision: ses lois logiques, ses sources subjectives, Annales de l'Institut Henri Poincaré 7 (1) (1937) 1–68.
URL `http://www.numdam.org/item?id=AIHP_1937__7_1_1_0`

[20] B. de Finetti, Probability, Induction and Statistics (The art of guessing), John Wiley & Sons, 1972.

[21] M. J. Schervish, Theory of Statistics, Springer, 1995.

[22] F. G. Cozman, Credal networks, Artificial Intelligence 120 (2000) 199–233.

[23] T. Seidenfeld, Conditional independence, imprecise probabilities, null-events and graph-theoretic models., Seimour Geisser Distinguished Lecture, School of Statistics, University of Minnesota, Minneapolis, MN (2007).

[24] F. G. Cozman, Separation properties of sets of probabilities, in: C. Boutilier, M. Goldszmidt (Eds.), Uncertainty in Artificial Intelligence (Proceedings of the Sixteenth Conference), Morgan Kaufmann, San Francisco, 2000, pp. 107–115.

[25] S. Moral, A. Cano, Strong conditional independence for credal sets, Annals of Mathematics and Artificial Intelligence 35 (1–4) (2002) 295–321.

[26] B. de Finetti, Teoria delle Probabilità, Einaudi, Turin, 1970.

[27] S. Willard, General Topology, Dover Books on Mathematics Series, Dover, 1970.
URL `http://books.google.be/books?id=UrsHbOjiR8QC`

[28] T. R. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

Appendix A. Proofs of main results

Proof of Proposition 1. Consider any finite subset R of $\mathbb{N}$ that has $r = |R|$ elements and any $f \in \mathcal{G}(\mathcal{X}^r) = \mathcal{G}(\mathcal{X}_{\downarrow r})$. If we call n the highest element of R, then both $\underline{P}_R$ and $\underline{P}_{\downarrow r}$ can be derived from $\underline{P}_{\downarrow n}$ through marginalisation. We find that due to Eq. (2), $\underline{P}_{\downarrow r}(f) = \underline{P}_{\downarrow n}(f)$, where we have used cylindrical extension (Section 3.1) to identify the gamble f on $\mathcal{X}_{\downarrow r}$ with its extension to $\mathcal{X}_{\downarrow n}$, thereby allowing us to apply $\underline{P}_{\downarrow n}$ to f. Now choose a permutation π in $\mathcal{P}_n$ that maps the elements of R onto the first r natural numbers and apply it to the cylindrical extension of f to $\mathcal{X}_{\downarrow n}$. This results in a gamble $\pi f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ that theoretically depends on the value of $X_{\downarrow n}$, but in practice only depends on the value that X_R takes in $\mathcal{X}_R$ and on that domain, coincides with f. Therefore, we can apply Eq. (1), with $N = \{1, \dots, n\}$, to find that $\underline{P}_R(f) = \underline{P}_{\downarrow n}(\pi f)$. Finally, we obtain $\underline{P}_{\downarrow n}(f) = \underline{P}_{\downarrow n}(\pi f)$ by applying the permutability of the sequence. Together with the previous two equalities, this yields $\underline{P}_R(f) = \underline{P}_{\downarrow r}(f)$. $\square$

Proof of Proposition 3. By definition, forward irrelevance of $\underline{P}_{\downarrow n}$ means joint coherence of $\underline{P}_{\downarrow n}$ with the collection of conditional lower previsions $\underline{P}_{\downarrow k+1}(\cdot|X_{\downarrow k})$, $1 \le k \le n-1$, defined for all $x_{\downarrow k} \in \mathcal{X}_{\downarrow k}$ and $f \in \mathcal{G}(\mathcal{X}_{\downarrow k+1})$ by

$$\underline{P}_{\downarrow k+1}(f|x_{\downarrow k}) := \underline{P}_{k+1}(f(\cdot, x_{\downarrow k})|x_{\downarrow k}) := \underline{P}_{\{k+1\}}(f(\cdot, x_{\downarrow k})).$$

We know that this collection is (jointly) coherent on its own because Ref. [8, Proposition 13] proves (joint) coherence for a superset[13] of this collection, implying (joint) coherence of the subset. Therefore, we can invoke Ref. [8, Theorem 2] to find that $\underline{P}_{\downarrow n}$ is forward irrelevant if and only if $\underline{P}_{\downarrow n}$ is weakly coherent with the collection of conditional lower previsions $\underline{P}_{\downarrow k+1}(\cdot|X_{\downarrow k})$, $1 \le k \le n-1$. Taking into account Ref. [8, Theorem 1], this holds if and only if for every $1 \le k \le n-1$

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}_{\downarrow k+1}(f|x_{\downarrow k})]) = 0 \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow k+1}) \text{ and } x_{\downarrow k} \in \mathcal{X}_{\downarrow k},$$

which is in turn equivalent to Eq. (7) because we have $\mathbb{I}_{\{x_{\downarrow k}\}} f = \mathbb{I}_{\{x_{\downarrow k}\}} f(\cdot, x_{\downarrow k})$, $\underline{P}_{\downarrow k+1}(f|x_{\downarrow k}) = \underline{P}_{\{k+1\}}(f(\cdot, x_{\downarrow k}))$ and $f(\cdot, x_{\downarrow k}) \in \mathcal{G}(\mathcal{X}_{k+1})$. $\square$

Proof of Corollary 4. Since a countable sequence $X_1, \dots, X_n, \dots$ is by definition forward irrelevant if and only if the finite sequence $X_1, \dots, X_n$ is forward irrelevant for every $n \in \mathbb{N}$, we can apply Proposition 3 to find that $X_1, \dots, X_n, \dots$ is forward irrelevant if and only if for every $n \in \mathbb{N}$ and $1 \le k \le n-1$

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}_{\{k+1\}}(f)]) = 0 \text{ for all } f \in \mathcal{G}(\mathcal{X}_{k+1}) \text{ and } x_{\downarrow k} \in \mathcal{X}_{\downarrow k}.$$

Due to the additional assumption of identical marginal distributions, this turns into demanding that for every $n \in \mathbb{N}$ and $1 \le k \le n-1$

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]) = 0 \text{ for all } f \in \mathcal{G}(\mathcal{X}_{k+1}) \text{ and } x_{\downarrow k} \in \mathcal{X}_{\downarrow k}.$$

The final equivalence with Eq. (8) is now due to time-consistency (Eq. (2)), because it implies that for all $n \in \mathbb{N}$, $1 \le k \le n-1$, $f \in \mathcal{G}(\mathcal{X}_{k+1})$ and $x_{\downarrow k} \in \mathcal{X}_{\downarrow k}$

$$\underline{P}_{\downarrow n}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]) = \underline{P}_{\downarrow k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)]). \qquad \square$$

[13] The superset that corresponds to epistemic many-to-one independence.

Proof of Proposition 5. For every $\theta \in \Sigma$, we have that $\sum_{x\in\mathcal{X}} \theta_x = 1$ by definition. As a consequence, we can derive that

$$1 = \left(\sum_{x\in\mathcal{X}} \theta_x\right)^n = \sum_{m\in\mathcal{N}^n} B^*_m(\theta) \text{ for all } \theta \in \Sigma,$$

proving that $\mathcal{B}_n$ is a partition of unity. To prove that it is a basis for $\mathcal{V}_n(\Sigma)$, we need to show that every $h \in \mathcal{V}_n(\Sigma)$ can be written as a linear combination of Bernstein basis gambles in $\mathcal{B}_n$ and that all of these Bernstein basis gambles are linearly independent.

We start with the first part. Consider any polynomial gamble $h \in \mathcal{V}_n(\Sigma)$. By definition, there is a polynomial p on $\mathbb{R}^{\mathcal{X}}$, with $\deg(p) \leq n$, for which $h = p^*$. Since $\deg(p) \leq n$, we know that p is a sum $\sum_{i=1}^s p_i$ of a finite amount of monomials p_i, each of which has a degree $d_i \leq n$. One such monomial is of the form

$$p_i(\theta) = c_i \prod_{x\in\mathcal{X}} \theta_x^{d_i(x)} \text{ for all } \theta \in \mathbb{R}^{\mathcal{X}},$$

in which c_i is a coefficient in $\mathbb{R}$ and $d_i(x)$ are exponents in $\mathbb{N}_0$ that sum up to d_i. The polynomial p'_i that is given by

$$p'_i(\theta) = c_i \prod_{x\in\mathcal{X}} \theta_x^{d_i(x)} \left(\sum_{x\in\mathcal{X}} \theta_x\right)^{n-d_i} \text{ for all } \theta \in \mathbb{R}^{\mathcal{X}},$$

is now a linear combination of Bernstein basis polynomials of degree n. Its restriction p'^*_i to Σ is thus a linear combination of Bernstein basis gambles of degree n and coincides with p^*_i because $\sum_{x\in\mathcal{X}} \theta_x = 1$ for all $\theta \in \Sigma$. As a direct consequence, we have that $h = p^* = \sum_{i=1}^s p^*_i = \sum_{i=1}^s p'^*_i$ is a linear combination of Bernstein basis gambles of degree n.

To prove that all Bernstein basis gambles of degree n are linearly independent, assume ex absurdo that there is some gamble $b \in \mathcal{G}(\mathcal{N}^n)$ that differs from zero and for which $h_b(\theta) := \sum_{m\in\mathcal{N}^n} b(m)B^*_m(\theta) = 0$ for all $\theta \in \Sigma$. Now consider the polynomial p_b on $\mathbb{R}^{\mathcal{X}}$, defined by $p_b(\theta) := \sum_{m\in\mathcal{N}^n} b(m)B_m(\theta)$ for all $\theta \in \mathbb{R}^{\mathcal{X}}$. Since h_b is the restriction of p_b to the simplex, we have for all $\theta \in \Sigma$ that $p_b(\theta) = 0$. We also know that it is a homogeneous polynomial of degree n (it consists of a sum of monomials of degree n), which implies for all $\lambda \in \mathbb{R}$ that $p_b(\lambda\theta) = \lambda^n p_b(\theta)$. Because every $\theta \in \mathbb{R}^{\mathcal{X}}$ in the positive orthant ($\theta_x \geq 0$ for all $x \in \mathcal{X}$), has a corresponding $\lambda \in \mathbb{R}$ and $\theta' \in \Sigma$ for which $\theta = \lambda\theta'$, it follows that $p_b(\theta) = p_b(\lambda\theta') = \lambda^n p_b(\theta') = 0$ for all θ in the positive orthant. Since p_b is a polynomial, this is only possible if $p_b(\theta) = 0$ for all $\theta \in \mathbb{R}^{\mathcal{X}}$, which would mean that $p_b = \sum_{m\in\mathcal{N}^n} b(m)B_m$ is identically zero. However, since p_b is a linear combination of monomials on $\mathbb{R}^{\mathcal{X}}$, which are known to be linearly independent, this would imply that all coefficients $b(m)$ have to be zero, leading to a contradiction. Therefore, all Bernstein basis gambles of degree n are linearly independent, which completes the proof. □

Proof of Corollary 6. This is trivial because every polynomial gamble h in $\mathcal{V}(\Sigma)$ is an element of $\mathcal{V}_{\deg(h)}(\Sigma)$ and can therefore be written as a linear combination of Bernstein basis gambles of degree $\deg(h)$ due to proposition 5. □

Proof of Corollary 7. Consider any polynomial gamble $h \in \mathcal{V}(\Sigma)$ and use n to denote its degree $\deg(h)$. Due to proposition 5, h has a unique corresponding $b \in \mathcal{G}(\mathcal{N}^n)$ for which $h = \sum_{m\in\mathcal{N}^n} b(m)B^*_m$. Therefore, $p = \sum_{m\in\mathcal{N}^n} b(m)B_m$ is a homogeneous polynomial that represents h and for which $\deg(p) = n$. To show that it is unique, assume that there is another homogeneous polynomial p' of degree n for which $p'^* = h$. Because it is homogeneous and has degree n, there has to be some $b' \in \mathcal{N}^n$ such that $p' = \sum_{m\in\mathcal{N}^n} b'(m)B_m$ and we would thus have that $h = p'^* = \sum_{m\in\mathcal{N}^n} b'(m)B^*_m$. Due to the uniqueness of the b that was mentioned in the beginning of this proof, we find that $b = b'$, which implies that $p = p'$ and thus completes the proof. □

Proof of Proposition 8. Consider any k and n in $\mathbb{N}$, with $k \leq n$, and any $f_k \in \mathcal{G}(\mathcal{X}_{\downarrow k})$, and let f_n be its cylindrical extension to $\mathcal{G}(\mathcal{X}_{\downarrow n})$. Let $R := \{k+1, \dots, n\}$. Then for all $\theta \in \Sigma$, we know from Proposition 5 that

$$\sum_{x_R\in\mathcal{X}_R} \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_R)} = \sum_{m\in\mathcal{N}^n} B^*_m = 1,$$

which implies that

$$\begin{aligned}
\mathrm{Mn}_n(f_n)(\theta) &= \sum_{x_{\downarrow n}\in\mathcal{X}_{\downarrow n}} f_n(x_{\downarrow n}) \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow n})} \\
&= \sum_{x_{\downarrow n}\in\mathcal{X}_{\downarrow n}} f_k(x_{\downarrow k}) \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow n})} \\
&= \sum_{x_{\downarrow k}\in\mathcal{X}_{\downarrow k}} \sum_{x_R\in\mathcal{X}_R} f_k(x_{\downarrow k}) \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow k})+T_x(x_R)} \\
&= \sum_{x_{\downarrow k}\in\mathcal{X}_{\downarrow k}} f_k(x_{\downarrow k}) \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow k})} \sum_{x_R\in\mathcal{X}_R} \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_R)} = \mathrm{Mn}_k(f_k)(\theta).
\end{aligned}$$

Since this is true for every $\theta \in \Sigma$, we find that

$$\underline{P}_{\downarrow k}(f_k) := \underline{R}(\mathrm{Mn}_k(f_k)) = \underline{R}(\mathrm{Mn}_n(f_n)) = \underline{P}_{\downarrow n}(f_n).$$

Hence, $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$ is a time-consistent sequence of models.

Now fix any $n \in \mathbb{N}$. It remains to prove that $\underline{P}_{\downarrow n}$ is coherent and exchangeable.

We start with coherence. Since $\underline{R}$ is coherent, it satisfies P1–P3. Now notice from Equation (11) and Proposition 5 that, for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\theta \in \Sigma$, $\mathrm{Mn}_n(f)(\theta)$ is a convex combination of the values of f and therefore dominates its infimum $\inf f$. Since this is true for every $\theta \in \Sigma$, we have that $\inf \mathrm{Mn}_n(f) \geq \inf f$. Since $\underline{R}$ satisfies P1, this implies that $\underline{P}_{\downarrow n}$ satisfies P1 as well. That $\underline{P}_{\downarrow n}$ satisfies P2 and P3 follows directly from the linearity of the operator Mn_n and the fact that $\underline{R}$ satisfies P2 and P3. Hence, $\underline{P}_{\downarrow n}$ is a coherent lower prevision on $\mathcal{G}(\mathcal{X}_{\downarrow n})$.

For exchangeability, it suffices to notice that for any $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$ and $\pi \in \mathcal{P}_n$ the polynomial gambles $\mathrm{Mn}_n(f)$ and $\mathrm{Mn}_n(\pi f)$ are identical, implying that $\mathrm{Mn}_n(\pi f - f) = 0$. Hence, since $\underline{R}$ satisfies (P1), we find that $\underline{P}_{\downarrow n}(\pi f - f) \geq 0$, as desired. □

Proof of Equation (20). For all $k \in \mathbb{N}$, $x_{\downarrow k} \in \mathcal{X}_{\downarrow k}$, $f \in \mathcal{G}(\mathcal{X}_{k+1})$ and $\theta \in \Sigma$

$$\begin{aligned}
&\nu(T(x_{\downarrow k}))\,\mathrm{Mn}_{k+1}(\mathbb{I}_{\{x_{\downarrow k}\}}[f - \underline{P}(f)])(\theta) \\
&= \nu(T(x_{\downarrow k})) \sum_{z_{\downarrow k+1} \in \mathcal{X}_{\downarrow k+1}} \mathbb{I}_{\{x_{\downarrow k}\}}(z_{\downarrow k})[f(z_{k+1}) - \underline{P}(f)] \prod_{x \in \mathcal{X}} \theta_x^{T_x(z_{\downarrow k+1})} \\
&= \nu(T(x_{\downarrow k})) \sum_{\substack{z_{\downarrow k+1} \in \mathcal{X}_{\downarrow k+1} \\ z_{\downarrow k} = x_{\downarrow k}}} [f(z_{k+1}) - \underline{P}(f)] \prod_{x \in \mathcal{X}} \theta_x^{T_x(z_{\downarrow k+1})} \\
&= \nu(T(x_{\downarrow k})) \sum_{z_{k+1} \in \mathcal{X}_{k+1}} [f(z_{k+1}) - \underline{P}(f)] \prod_{x \in \mathcal{X}} \theta_x^{T_x(x_{\downarrow k})} \theta_{z_{k+1}} \\
&= \nu(T(x_{\downarrow k})) \prod_{x \in \mathcal{X}} \theta_x^{T_x(x_{\downarrow k})} \sum_{z_{k+1} \in \mathcal{X}_{k+1}} [f(z_{k+1}) - \underline{P}(f)] \theta_{z_{k+1}} \\
&= B^*_{T(x_{\downarrow k})}(\theta) \sum_{z_{k+1} \in \mathcal{X}_{k+1}} [f(z_{k+1}) - \underline{P}(f)] \theta_{z_{k+1}} \\
&= B^*_{T(x_{\downarrow k})}(\theta) \sum_{x \in \mathcal{X}} [f(x) - \underline{P}(f)] \theta_x \\
&= B^*_{T(x_{\downarrow k})}(\theta) \bigg(\sum_{x \in \mathcal{X}} f(x)\theta_x - \underline{P}(f) \bigg) = B^*_{T(x_{\downarrow k})}(\theta)(\mathrm{Mn}_1(f)(\theta) - \underline{P}(f)),
\end{aligned}$$

where the first and last equality follow from Eqs. (11) and (14), respectively. ☐

Proof of Proposition 10. Consider any $g \in \mathcal{V}_{>0}(\Sigma)$. Due to Corollary 7 it has a corresponding homogeneous polynomial p on $\mathbb{R}^{\mathcal{X}}$ of degree $\deg(p) = \deg(g)$ that represents it, meaning that $g = p^*$. Since g is by definition strictly positive over Σ, this also holds for p. If we then use the homogeneity of p, we can derive that p is strictly positive on the set

$$K := \bigg\{ \theta \in \mathbb{R}^{\mathcal{X}} : \sum_{x \in \mathcal{X}} \theta_x \neq 0 \text{ and } \theta_x \geq 0 \text{ for all } x \in \mathcal{X} \bigg\}.$$

This allows us to apply Pólya's result [16, Theorem 5.5.1], which says that if a homogeneous polynomial p on $\mathbb{R}^{\mathcal{X}}$ is strictly positive over K, then there exists an integer $k \geq 0$ such that the polynomial p' on $\mathbb{R}^{\mathcal{X}}$, defined for all $\theta \in \Sigma$ by

$$p'(\theta) := \bigg(\sum_{x \in \mathcal{X}} \theta_x \bigg)^k p(\theta),$$

has non-negative coefficients. Consequently, there exist non-negative coefficients c_m, $m \in \mathcal{N}^n$, with $n = \deg(p) + k$, such that $p' = \sum_{m \in \mathcal{N}^n} c_m B_m$ and therefore also $g = p^* = p'^* = \sum_{m \in \mathcal{N}^n} c_m B^*_m$.

Now consider a coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ that satisfies Eq. (23) and any $h \in \mathcal{H}_1$. The proof of this proposition then follows from the chain of (in)equalities

$$\underline{R}(hg) = \underline{R}\bigg(\sum_{m \in \mathcal{N}^n} c_m B^*_m h \bigg) \geq \sum_{m \in \mathcal{N}^n} \underline{R}(c_m B^*_m h) = \sum_{m \in \mathcal{N}^n} c_m \underline{R}(B^*_m h) \geq 0,$$

in which the first equality follows from the previous paragraph. The first inequality and the subsequent equality are due to coherence (P3 and P2 respectively) and the final inequality is a consequence of Eq. (23). ☐

Lemma 21. *Consider a non-empty, convex and compact $C \subseteq \mathbb{R}^{\mathcal{X}}$ and any $\delta > 0$. Let*

$$C_\delta := \big\{\theta \in \mathbb{R}^{\mathcal{X}} : \|\theta - y\| \leq \delta \text{ for at least one } y \in C\big\},$$

with $\|\theta - y\|$ the Euclidean distance between θ and y. Then C_δ is a convex and compact set whose interior $\operatorname{int} C_\delta$ is a strict superset of C.

Proof. Throughout this proof, for any $\theta \in \mathbb{R}^{\mathcal{X}}$ and $\epsilon > 0$, we let

$$B_{\theta,\epsilon} := \{\theta' \in \mathbb{R}^{\mathcal{X}} : \|\theta - \theta'\| \leq \epsilon\}$$

be the closed ϵ-ball around θ.

We now start with convexity. Consider any $\theta_1, \theta_2 \in C_\delta$ and any $\lambda \in [0,1]$ and consider the convex combination $\theta := \lambda\theta_1 + (1-\lambda)\theta_2$. Since $\theta_1, \theta_2 \in C_\delta$, there are $y_1, y_2 \in C$ such that $\|\theta_1 - y_1\| \leq \delta$ and $\|\theta_2 - y_2\| \leq \delta$. Let $y := \lambda y_1 + (1-\lambda)y_2$. The convexity of C then implies that $y \in C$. Hence, since

$$\begin{aligned}
\|\theta - y\| &= \|\lambda\theta_1 + (1-\lambda)\theta_2 - (\lambda y_1 + (1-\lambda)y_2)\| \\
&= \|\lambda(\theta_1 - y_1) + (1-\lambda)(\theta_2 - y_2)\| \\
&\leq \lambda\|\theta_1 - y_1\| + (1-\lambda)\|\theta_2 - y_2\| \leq \lambda\delta + (1-\lambda)\delta = \delta,
\end{aligned}$$

we find that $\theta \in C_\delta$.

For compactness, we prove that C_δ is bounded and closed.

We begin with boundedness. Since C is compact, it is bounded, meaning that there is some constant $\alpha \in \mathbb{R}$ such that $\|\theta\| \leq \alpha$ for all $\theta \in C$. Now let $\alpha' := \alpha + \delta$ and consider any $\theta \in C_\delta$, implying that there is some $y \in C$ such that $\|\theta - y\| \leq \delta$. Then $\|\theta\| \leq \|\theta - y\| + \|y\| \leq \delta + \alpha = \alpha'$. Since $\theta \in C_\delta$ is arbitrary, this implies that C_δ is bounded.

To establish that C_δ is closed, we will prove that its complement is open. So fix any $\theta \in \mathbb{R}^{\mathcal{X}} \setminus C_\delta$, meaning that $\|\theta - y\| > \delta$ for all $y \in C$. Since $\|\theta - y\|$ is a continuous function of y, the compactness of C implies that it achieves a minimum $M := \min_{y \in C} \|\theta - y\|$ on C. Since $\|\theta - y\| > \delta$ for all $y \in C$, we know that $M > \delta$ and therefore also that $M - \delta > 0$. Consider now any $\epsilon > 0$ such that $\epsilon < M - \delta$. For all $\theta' \in B_{\theta,\epsilon}$ we then find that $\theta' \in \mathbb{R}^{\mathcal{X}} \setminus C_\delta$ because

$$\|\theta' - y\| = \|\theta - y + \theta' - \theta\| \geq \|\theta - y\| - \|\theta' - \theta\| \geq M - \epsilon > \delta$$

for all $y \in C$. Since $\theta \in \mathbb{R}^{\mathcal{X}} \setminus C_\delta$ was arbitrary, this implies that $\mathbb{R}^{\mathcal{X}} \setminus C_\delta$ is open, and therefore, that C_δ is closed.

We end by proving that $\operatorname{int} C_\delta$ is a strict superset of C. Consider any $\theta \in C$. Then for all $\theta' \in B_{\theta,\delta}$, we know that $\|\theta - \theta'\| \leq \delta$ and therefore, since $\theta \in C$, also that $\theta' \in C_\delta$. Hence, we find that $B_{\theta,\delta} \subseteq C_\delta$, which implies that $\theta \in \operatorname{int} C_\delta$. Since $\theta \in C$ was arbitrary, we conclude that C is a subset of $\operatorname{int} C_\delta$. In order to prove that it is a strict subset, we assume *ex absurdo* that $\operatorname{int} C_\delta = C$. Consider now any $\theta \in C$ and any $y \in \mathbb{R}^{\mathcal{X}}$ such that $\|y\| = \delta/2$ and, for all $n \in \mathbb{N} \cup \{0\}$, let $\theta_n := \theta + ny$. For all $\theta' \in B_{\theta_1,\delta/2}$, we then have that

$$\|\theta_0 - \theta'\| = \|\theta_1 - y - \theta'\| \leq \|y\| + \|\theta_1 - \theta'\| \leq \|y\| + \delta/2 = \delta,$$

which, since $\theta_0 = \theta \in C$, implies that $\theta_1 \in \operatorname{int} C_\delta = C$. Similarly, for all $\theta' \in B_{\theta_2, \delta/2}$,

$$\|\theta_1 - \theta'\| = \|\theta_2 - y - \theta'\| \le \|y\| + \|\theta_2 - \theta'\| \le \|y\| + \delta/2 = \delta.$$

Since $\theta_1 \in C$, this implies that $\theta_2 \in \operatorname{int} C_\delta = C$. By continuing in this way, we find that $\theta_n \in C$ for all $n \in \mathbb{N}$. As mentioned before, the compactness (and hence boundedness) of C implies that there is some $\alpha \in \mathbb{R}$ such that $\|\theta'\| \le \alpha$ for all $\theta' \in C$. Consider now any $n \in \mathbb{N}$ such that $n\delta > 4\alpha$. Then since $\theta_n \in C$, we know that $\|\theta_n\| \le \alpha$. However, we also find that

$$\|\theta_n\| = \|\theta + ny\| \ge \|ny\| - \|\theta\| = n\|y\| - \|\theta\| \ge n\|y\| - \alpha = n\frac{\delta}{2} - \alpha > 2\alpha - \alpha = \alpha,$$

a contradiction. □

Proof of Proposition 12. Consider any non-empty, convex and compact subset C of $\mathbb{R}^{\mathcal{X}}$ and any polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$ that is strictly positive over C. Due to the compactness of C and the continuity of p, the minimum of p over C is well defined and we will denote it by $\epsilon = \min\{p(\theta) : \theta \in C\}$. Note that $\epsilon > 0$, because p was assumed to be strictly positive over C. We now define, for any $\delta > 0$, the set

$$C_\delta := \big\{\theta \in \mathbb{R}^{\mathcal{X}} : \|\theta - y\| \le \delta \text{ for at least one } y \in C\big\},$$

in which $\|\theta - y\|$ is the Euclidean distance between θ and y. We then know from Lemma 21 that C_δ is a convex and compact set whose interior $\operatorname{int} C_\delta$ is a strict superset of C. Next, we fix an arbitrary $\delta_1 > 0$ and its corresponding set C_{δ_1}. Due to the compactness of C_{δ_1}, the continuous polynomial p is uniformly continuous over C_{δ_1} [27, Corollary 36.20], implying the existence of a $\delta_2 > 0$ for which it holds for all θ and y in C_{δ_1} that

$$\|\theta - y\| < \delta_2 \Longrightarrow |p(\theta) - p(y)| < \epsilon.$$

Now choose a $\delta > 0$ that is strictly smaller than both δ_1 and δ_2. For any $\theta \in C_\delta$, we know that there is some $y \in C$ for which $\|\theta - y\| \le \delta < \delta_2$. Since θ and y clearly both belong to C_{δ_1}, this in turn implies that $|p(\theta) - p(y)| < \epsilon$. Hence, we find that

$$p(\theta) \ge p(y) - |p(\theta) - p(y)| > p(y) - \epsilon \ge 0,$$

in which the final inequality follows from the fact that ϵ is the minimum of p over C. Since $\theta \in C_\delta$ was arbitrary, we have thus proved the existence of some $\delta > 0$ for which p is strictly positive over C_δ.

We now invoke a theorem that was proved in [28, Theorem 20.4], stating that if C is a non-empty closed and bounded convex set, and D is a convex set such that $C \subset \operatorname{int} D$, then there exists a polyhedral convex set P such that $P \subset \operatorname{int} D$ and $C \subset \operatorname{int} P$. Because in Euclidean space, compactness is equivalent to being closed and bounded, our set C satisfies all properties needed for this Theorem to hold. If we now choose D to be C_δ we can derive from this theorem the existence of a polyhedral set P such that $P \subset \operatorname{int} C_\delta$ and $C \subset \operatorname{int} P$. By definition, such a polyhedral set P has a corresponding finite subset $S = \{q_1, \dots, q_s\}$ of $\mathcal{V}_1(\mathbb{R}^{\mathcal{X}})$ for which $P = K_S$. The set K_S is always closed by definition and in this case is also bounded and thus compact because it is a subset of the compact (and hence bounded) set C_δ. It is also a superset of C because $C \subset \operatorname{int} P \subseteq P = K_S$, thereby completing the proof that K_S is a compact superset of C. The proof that p is an element of M_S now follows directly from Theorem 11 because K_S is compact and p is strictly positive over C_δ and thus also over its subset K_S. □

Proof of Corollary 13. Consider any non-empty, closed and convex subset $\mathcal{M}$ of Σ and a polynomial gamble $h \in \mathcal{V}(\Sigma)$ that is strictly positive over $\mathcal{M}$. By definition, h has a corresponding polynomial $p \in \mathcal{V}(\mathbb{R}^{\mathcal{X}})$ for which $p^* = h$, implying that p is also strictly positive over $\mathcal{M}$. The set $\mathcal{M}$ is compact since it is closed by assumption and bounded because it is a subset of Σ. Therefore, the minimum of p over $\mathcal{M}$ is well-defined and we will denote it as $\epsilon = \min\{p(\theta) : \theta \in \mathcal{M}\}$. Note that $\epsilon > 0$ because p is strictly positive over $\mathcal{M}$. We now choose some α such that $0 < \alpha < \epsilon$ and define $p_\alpha := p - \alpha$. It should be clear that p_α is also strictly positive over $\mathcal{M}$.

Next, because $\mathcal{M}$ is non-empty, convex and compact, we can invoke Proposition 12 to show the existence of a finite subset $S = \{q_1, \ldots, q_s\}$ of linear polynomials in $\mathcal{V}_1(\mathbb{R}^{\mathcal{X}})$ for which K_S is a superset of $\mathcal{M}$ and p_α is an element of M_S. We thus find that

$$p = \alpha + p_\alpha = \alpha + \sigma_0 + q_1\sigma_1 + \ldots + q_s\sigma_s = \alpha + \sigma_0 + \sum_{i=1}^{s} q_i\sigma_i,$$

where for all $i \in \{0, \ldots, s\}$, $\sigma_i \in \sum \mathcal{V}^2(\mathbb{R}^{\mathcal{X}})$. We now introduce the polynomial $q := \sum_{i=1}^{s} q_i$, of which the maximum $\max\{q(\theta) : \theta \in \Sigma\}$ over Σ is well-defined due to the compactness of Σ. It is then always possible to choose some $\delta > 0$ such that $\delta \max\{q(\theta) : \theta \in \Sigma\} < \alpha$ and we use it to define $p_0 := \alpha - \delta q + \sigma_0$, $q_0 := 1$ and $p_i := \delta + \sigma_i$ for all $i \in \{1, \ldots, s\}$, allowing us to write

$$p = \alpha + \sigma_0 + \sum_{i=1}^{s} q_i\sigma_i = \alpha + \sigma_0 - \delta q + \sum_{i=1}^{s} \delta q_i + \sum_{i=1}^{s} q_i\sigma_i = \sum_{i=0}^{s} q_i p_i.$$

Consequently, if we define $h_i := q_i^*$ and $g_i := p_i^*$ for all $i \in \{0, \ldots, s\}$, we find that

$$h = p^* = \left(\sum_{i=0}^{s} q_i p_i\right)^* = \sum_{i=0}^{s} q_i^* p_i^* = \sum_{i=0}^{s} h_i g_i.$$

The proof is then concluded if we can show for all $i \in \{0, \ldots, s\}$ that $h_i \in \mathcal{H}_1$ and $g_i \in \mathcal{V}_{>0}(\Sigma)$. We start with the polynomials h_i. For $i = 0$, $h_0 = 1$ by definition and is thus trivially an element of $\mathcal{H}_1$. For $i \in \{1, \ldots, s\}$, the linear polynomials q_i are by definition non-negative over K_S and thus also over the subset $\mathcal{M}$, which makes their restrictions h_i to Σ elements of $\mathcal{H}_1$.

Each of the polynomials σ_i, with $i \in \{0, \ldots, s\}$, is an element of $\sum \mathcal{V}^2(\mathbb{R}^{\mathcal{X}})$, implying that it is non-negative everywhere. For $i \in \{1, \ldots, s\}$, this means that $p_i = \delta + \sigma_i$ is strictly positive everywhere, and its restriction g_i tot Σ is thus certainly an element of $\mathcal{V}_{>0}(\Sigma)$. For the special case $i = 0$, we have that $g_0 = p_0^*$, with $p_0 = \alpha - \delta q + \sigma_0$. Since $\delta \max\{q(\theta) : \theta \in \Sigma\} < \alpha$, we know that $\alpha - \delta q$ and consequently also $p_0 = \alpha - \delta q + \sigma_0$ are strictly positive over Σ. Therefore, $g_0 = p_0^*$ is an element of $\mathcal{V}_{>0}(\Sigma)$, concluding the proof. □

Proof of Theorem 14. Consider any polynomial gamble $h \in \mathcal{V}(\Sigma)$ that is non-negative over $\mathcal{M}$. Then for all $\epsilon > 0$, the polynomial gamble $h_\epsilon = \epsilon + h$ will be strictly positive over $\mathcal{M}$, allowing us to use Corollary 13 to find that $h_\epsilon = \sum_{i=0}^{s} h_i g_i$, with $h_i \in \mathcal{H}_1$ and $g_i \in \mathcal{V}_{>0}(\Sigma)$ for all $i \in \{0, \dots, s\}$. As a consequence, due to Proposition 10 and the superadditivity that follows from coherence, we obtain that $\underline{R}(h+\epsilon) = \underline{R}(h_\epsilon) \geq 0$, which is in turn equivalent with $\underline{R}(h) \geq -\epsilon$ due to coherence property P5. Since this is true for all $\epsilon > 0$, we find that $\underline{R}(h) \geq 0$. $\square$

Proof of Proposition 15. Fix any $k \in \mathbb{N}$, $x_{\downarrow k} \in \mathcal{X}^k$ and $f \in \mathcal{G}(\mathcal{X})$. Then we know from Eq. (16) that $\min\{\mathrm{Mn}_1(f)(\theta) : \theta \in \mathcal{M}\} = \underline{P}(f)$, implying that $\underline{R}^{\mathrm{beh}}$ satisfies Eq. (13). Also, it implies that $\min\{\mathrm{Mn}_1(f)(\theta) - \underline{P}(f) : \theta \in \mathcal{M}\} = 0$. Since the Bernstein basis gamble $B^*_{T(x_{\downarrow k})}$ is non-negative by definition, this in turn implies that

$$\min\big\{B^*_{T(x_{\downarrow k})}(\theta)[\mathrm{Mn}_1(f)(\theta) - \underline{P}(f)] : \theta \in \mathcal{M}\big\} = 0,$$

and therefore that $\underline{R}^{\mathrm{beh}}$ satisfies Eq. (21). $\square$

Proof of Corollary 16. Consider any coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ that satisfies Eq. (23). Now choose any polynomial gamble $h \in \mathcal{V}(\Sigma)$ and define the corresponding polynomial gamble h' as

$$h' := h - \min\{h(\theta) : \theta \in \mathcal{M}\}.$$

Since h' is positive on $\mathcal{M}$ and $\underline{R}$ satisfies Eq. (23), we can apply Theorem 14 to find that $\underline{R}(h') \geq 0$, or equivalently, due to coherence property P5, that

$$\underline{R}(h) \geq \min\{h(\theta) : \theta \in \mathcal{M}\} = \underline{R}^{\mathrm{beh}}(h). \qquad \square$$

Proof of Theorem 17. Consider any time-consistent sequence of coherent lower previsions $\{\underline{P}_{\downarrow n}\}_{n\in\mathbb{N}}$ that is (i) exchangeable, (ii) identically distributed with a given marginal model $\underline{P}$ and (iii) forward irrelevant. Then due to (i) and Theorem 9, there is a unique corresponding coherent lower prevision $\underline{R}$ on $\mathcal{V}(\Sigma)$ such that for every $n \in \mathbb{N}$

$$\underline{P}_{\downarrow n}(f) = \underline{R}(\mathrm{Mn}_n(f)) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}). \tag{A.1}$$

Due to (ii) and (iii), $\underline{R}$ satisfies Eqs. (19) and (22) and therefore also Eq. (23). We can thus invoke Corollary 16 to find that

$$\underline{R}^{\mathrm{beh}}(h) \leq \underline{R}(h) \text{ for all } h \in \mathcal{V}(\Sigma)$$

and therefore due to Eq. (A.1) and Eqs. (24) and (25) that for every $n \in \mathbb{N}$

$$\underline{P}^{\mathrm{beh}}_{\downarrow n}(f) \leq \underline{P}_{\downarrow n}(f) \text{ for all } f \in \mathcal{G}(\mathcal{X}_{\downarrow n}).$$

Since we know from Section 5.5 that $\{\underline{P}^{\mathrm{beh}}_{\downarrow n}\}_{n\in\mathbb{N}}$ is itself a time-consistent sequence of coherent lower previsions that satisfies (i), (ii) and (iii), it follows that it is the unique least committal, or most conservative one to do so. $\square$

Proof of Proposition 18. Fix an arbitrary $n \in \mathbb{N}$ and choose $\theta \in \mathcal{M}$. Then due to Eqs. (26) and (10), we have for all $x_{\downarrow n} \in \mathcal{X}_{\downarrow n}$ that

$$p^n_\theta(x_{\downarrow n}) = \prod_{i=1}^{n} \theta(x_i) = \prod_{x\in\mathcal{X}} \theta_x^{T_x(x_{\downarrow n})}.$$

Consequently, we can apply Eqs. (29) and (11) to find that for all $f \in \mathcal{G}(\mathcal{X}_{\downarrow n})$

$$E_\theta^n(f) = \sum_{x_{\downarrow n} \in \mathcal{X}_{\downarrow n}} f(x_{\downarrow n}) p_\theta^n(x_{\downarrow n}) = \sum_{x_{\downarrow n} \in \mathcal{X}_{\downarrow n}} f(x_{\downarrow n}) \prod_{x \in \mathcal{X}} \theta_x^{T_x(x_{\downarrow n})} = \mathrm{Mn}_n(f)(\theta).$$

The proof is now a direct consequence of Eqs. (24), (30) and (28). □

Proof of Theorem 19. This is a direct consequence of Theorem 17 and Proposition 18. □

Proof of Theorem 20. It follows from Theorem 19 that the time-consistent sequence of coherent lower previsions $\{\underline{P}_{\downarrow n}^{\mathrm{t2}}\}_{n \in \mathbb{N}}$, as defined by Eq. (30), is (i) exchangeable and (ii) identically distributed with a given marginal model $\underline{P}$. Since by construction, this sequence is a lower envelope of precise, stochastically independent ones, it is also (iii') epistemically independent (both many-to-many and many-to-one). Indeed, stochastic independence implies both many-to-many and many-to-one independence [8, Proposition 10] and both of these notions are preserved under taking lower envelopes; see Section 3.5. We are thus left to prove that the sequence $\{\underline{P}_{\downarrow n}^{\mathrm{t2}}\}_{n \in \mathbb{N}}$ is the unique least committal, most conservative one to satisfy these three properties.

Consider therefore any other time-consistent sequence of coherent lower previsions $\{\underline{P}_{\downarrow n}\}_{n \in \mathbb{N}}$ that is (i) exchangeable, (ii) identically distributed with a given marginal model $\underline{P}$ and (iii') epistemically independent (either many-to-many or many-to-one). We need to show that it is less committal (not as conservative) as the sequence $\{\underline{P}_{\downarrow n}^{\mathrm{t2}}\}_{n \in \mathbb{N}}$. Since joint coherence with a set of conditional lower previsions implies joint coherence with any subset, we immediately have that condition (iii') implies forward irrelevance and we can therefore apply Theorem 19 to conclude the proof. □