



Extended probabilities and their application to statistical inference

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Caprio, M., & Mukherjee, S. (2023). Extended probabilities and their application to statistical inference. In *Understanding Information and Its Role as a Tool: In Memory of Mark Burgin — PART II: Information Realm: Information as a Tool for Exploring Reality*

Published in:

Understanding Information and Its Role as a Tool

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Chapter 1

Extended probabilities and their application to statistical inference

Michele Caprio and Sayan Mukherjee*

*University of Pennsylvania and
Max Planck Institute for Mathematics in the Sciences,
caprio@seas.upenn.edu, sayan.mukherjee@mis.mpg.de*

We propose a new, more general definition of extended probability measures. We study their properties and provide a behavioral interpretation. We put them to use in an inference procedure, whose environment is canonically represented by the probability space (Ω, \mathcal{F}, P) , when both P and the composition of Ω are unknown. We develop an ex ante analysis – taking place before the statistical analysis requiring knowledge of Ω – in which the true composition of Ω is progressively learned. We describe how to update extended probabilities in this setting, and introduce the concept of lower extended probabilities. We apply our findings to a species sampling problem and to the study of the boomerang effect (the empirical observation that sometimes persuasion yields the opposite effect: the persuaded agent moves their opinion away from the opinion of the persuading agent).

1. Introduction

Researchers use the terminology “extended probabilities” to refer to set functions whose codomain is either a superset of $[0, 1]$, or defined using entirely different number types, such as p -adic numbers [1]. They first came up in physics (noticed by [2] and [3]), where they are still studied today [4–6]. They then became popular in other fields too, including economics and finance [7, 8], machine learning [9], stochastic processes [10], and queuing theory [11]. Of course, they have been extensively studied in mathematics (see e.g. [12] and [13] for early works, and [14] and [1] for more recent ones). A complete account is given in [15].

*This work is dedicated to the memory of Professor Mark Burgin. Without his profound contributions to the field of extended probabilities, this chapter would not have existed.

“Extended probabilities” have been given differing definitions and interpretations across – and even within – fields of study. In quantum theory, for example, [16] point out that “negative probabilities” do not have intrinsic meaning beyond the fact that they constitute a probabilistic model compatible with quantum events. In [5, 17, 18], instead, the authors interpret them as being associated with unsharable bets (see section 2.3).

In this chapter, we aim to give a foundational definition of extended probabilities. We give two properties that a set function must have in order to be called an extended probability, irrespective of the field the scholar works in. This represents a great improvement with respect to previous works on the matter, in which definitions depend on the area of study.

We explain how extended probabilities are different from regular Kolmogorovian probabilities, and we characterize some of their more interesting properties. We also give a behavioral interpretation of extended probabilities taking on negative values that complements the frequentist one given in [15] to “negative probabilities”. Finally, we relate our definition and interpretation to the ones existing in the literature, and we illustrate how these latter can be reconciled with the ones we provide.

After that, we present what is to the best of our knowledge the first application of extended probabilities to an inference procedure. Consider a generic statistical experiment; let Ω be a finite or countable space, and endow it with the sigma-algebra $\mathcal{F} = 2^\Omega$. The space usually adopted to express uncertainty around the elements of \mathcal{F} is the probability space (Ω, \mathcal{F}, P) , for some probability measure $P : \mathcal{F} \rightarrow [0, 1]$. Now, suppose we want to express further uncertainty regarding either the composition of Ω , or which P to consider on (Ω, \mathcal{F}) ; we can do so by using lower probabilities. In particular, in the first case, we consider the probability space (Ω, \mathcal{F}, P) and we follow the example in [19]. There, as a consequence of survey non-response (that is, subjects not answering to the questions in a survey), the author is forced to consider $\tilde{\Omega} = 2^\Omega$. In this case, probabilities cannot be computed exactly: only lower and upper bounds to precise probabilities – lower and upper probabilities, respectively – are available. In the second case, we consider the triple $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{P} is a set of probability measures, and we proceed e.g. as in [20, 21].

To the best of our knowledge, there is no cogent way of expressing uncertainty on both the composition of Ω and which P to consider on (Ω, \mathcal{F}) . In this work, we aim to fill this gap by using extended probabilities. We describe an *ex ante* analysis, meaning one that takes place before the actual statistical analysis that requires the knowledge of the state space.

In the most general case, we start from the number type we believe we are working with (naturals, wholes, integers, or rationals), and call it Ω . Then, at time 0 we divide it into an actual space Ω_0^+ that we deem a plausible state space for our experiment, and a latent one Ω_0^- , which we do not know about; notationally, $\Omega = \Omega_0^- \sqcup \Omega_0^+$, where \sqcup denotes a disjoint union of sets. We assign negative extended probabilities to the subsets of Ω_0^- . To capture the uncertainty around which P to consider, we specify a set \mathcal{P}^{ex} of extended probabilities supported on the whole Ω , instead of a single one. We then describe how we progressively discover the true composition of the state space associated with our experiment (which may be the whole Ω we initially specified, or a proper subset $\Omega' \subsetneq \Omega$), and how to update extended probabilities accordingly. We conclude our analysis by discovering the state space associated with our experiment.

In addition, we show that the limiting set of the sequence (\mathcal{P}_t^{ex}) of updates of the initially-specified set of extended probability measures must be one of the following. It is either a set of regular probability measures, if the state space associated with our experiment is the whole Ω , or a set of extended probability measures if the state space associated with our experiment is a proper subset Ω' of Ω . In this latter case, the limiting set induces a set of regular probability measures on Ω' .

We also develop the concept of lower and upper extended probabilities. They represent the “boundaries” of a generic set \mathcal{P}^{ex} of extended probabilities, and more in general allow for an imprecise elicitation of extended probabilities. They are extremely important because under a mild assumption, knowing lower probability \underline{P}^{ex} is enough to be able to retrieve the whole set \mathcal{P}^{ex} . We provide bounds for the lower extended probability $\underline{P}_t^{ex}(A)$ of any element $A \in \mathcal{F}$, at any time t in our ex ante analysis. For the sake of completeness, we give the definition of an extended Choquet capacity, and we show how lower and upper extended probabilities are extended Choquet capacities.

In Example 1, we provide a simple application to the field of ecology. We illustrate how the analysis we describe in the chapter can be put to use in a species sampling problem, specifically to retrieve the number of birds that inhabit a certain region throughout the year.

We also provide an example – adapted from the one in [22] – in the field of opinion dynamics; in particular, we describe the boomerang effect. We have a persuading agent acting on a persuaded agent, but the latter perceives the former as having low credibility. This can be modeled so that the persuaded agent does not know the composition of the entire state

space, while she suspects the persuading agent does: she thinks the persuading agent may be hiding something from her. As she discovers the true composition of the state space, the credibility of the persuading agent is restored.

This chapter is organized as follows: in section 2 we give the foundational definition of an extended probability, its properties, and its behavioral interpretation. In section 3 we use extended probabilities in an inference procedure. Section 4 deals with lower extended probabilities, and section 5 concludes our work. In appendix A.1, we give the proofs of our results. In appendix A.2 we give the opinion dynamics example, and in appendix A.3 we report two interesting quotes from [23].

Remark 1. To deal with uncertainty in the composition of Ω , one could proceed as in [24, section 4.3.3]. The agent could begin the elicitation by specifying the state space Ω_0 and then, as they analyze the problem in greater detail, could realize that a refinement to a finer-grained Ω_1 is needed. This corresponds to specifying what [25] calls a multivalued mapping from Ω_0 to Ω_1 ; that is,

$$A : \Omega_0 \rightrightarrows \Omega_1, \quad \omega_0 \mapsto A(\omega_0) \subset \Omega_1.$$

So Ω_0 corresponds to a partition of Ω_1 , and each state $\omega_0 \in \Omega_0$ can be identified with the set $A(\omega_0)$ of “refined possibilities” in Ω_1 . To illustrate the complication deriving from this approach, let Ω_0 and Ω_1 be finite or countable, and call $\mathcal{F}_0 = 2^{\Omega_0}$ and $\mathcal{F}_1 = 2^{\Omega_1}$. If the agent specifies a probability measure P_0 on $(\Omega_0, \mathcal{F}_0)$, they then need to come up with a probability measure P_1 on $(\Omega_1, \mathcal{F}_1)$ such that

$$P_0(\{\omega_0\}) = \sum_{\omega_1 \in A(\omega_0)} P_1(\{\omega_1\})$$

holds for all $\omega_0 \in \Omega_0$. This means that a new (subjective) probability elicitation must take place once the agent refines the state space to Ω_1 . This can be avoided using extended probabilities, as we shall argue in section 3.

It is worth noting that while this multivalued mapping approach is pointed out in [24, 25], the authors sidestep its associated complications by means of probability bounding. Furthermore, in [26] too the author tries to solve the incompletely specified possibility space problem by probability bounding; this enables progressive learning of the sample space in a way that representation invariance as well as symmetry are satisfied. We will verify whether the model that we introduce in section 3 meets these conditions in future work.

2. Extended probability measures

In this section, we first illustrate the philosophical reason to introduce extended probabilities, and then we dive into more technical details. We conclude with a thorough analysis on how our interpretation of extended probabilities relates to the existing literature.

2.1. *Philosophical motivation for extended probabilities*

We give a behavioral interpretation of extended probabilities. Consider a generic event A , and suppose we want to express our belief about the likelihood of it taking place. Suppose we can enter a bet about A that gives us \$1 if event A happens and \$0 if it does not happen. Then, we say that the probability we attach to A is given by $p \geq 0$, the supremum buying price as well as the infimum selling price for the bet about A (we call it *fair price*). That is, for every $\varepsilon > 0$, we are willing to enter a bet which gives $1 - p + \varepsilon$ if A happens, and $-p + \varepsilon$ if A does not happen, as well as the bet which gives $p - 1 + \varepsilon$ if A happens and $p + \varepsilon$ if A does not happen. This interpretation is inspired by the classical subjective probability interpretation given by de Finetti in [23, 27]. As pointed out in [28], the other two fathers of subjective probability theory, Ramsey [29] and Savage [30] simultaneously introduced measurement schemes for utility. They tied their definitions of probability to bets in which the payoffs were effectively measured in utiles rather than dollars. In this way, they obtained probabilities that were interpretable as measures of pure belief, uncontaminated by marginal utilities for money. De Finetti later admitted that it might have been better to adopt the seemingly more general approach of Ramsey and Savage, since it leads to a theory of decision-making that does not rely on monetary values. Nevertheless, he found other reasons for preferring the money bet approach. In particular, he maintained that it would be extremely difficult to settle bets based on utiles, because the monetary sums needed to settle them would need to be adjusted to the complex variations in a unit of measure (utiles) that is unobservable.^a This is why we retain the DeFinettian interpretation of (subjective) probability, and more in general why we adopt a betting scheme approach.

Suppose now that we are not given the possibility to enter a bet like the aforementioned one, so we cannot assess a probability for A as before. Instead, such a possibility is given to our doppelgänger, who tells us that

^aThe complete quotes from [23] can be found in Appendix A.3.

their subjective assessment for the probability of A is some $q \in [0, 1]$. Here, it is assumed that the doppelgänger assigns the same probabilities as ourselves to all the events we both can enter a bet about. This procedure of asking the doppelgänger is equivalent to setting the probability of A ourselves; we introduce the doppelgänger – a logical artifice – because, given our interpretation of probability, it is impossible to elicit the probability of event A if we cannot enter a bet similar to the one described before. We conclude that if we were given the opportunity to enter the bet about A , the amount of money we would deem fair to pay would be q dollars. Therefore we are prepared to lose $\$q$ in the case A^c happens.^b We express this by setting $p = -q$.

If after a while we are given the opportunity to enter a bet about an event A whose extended probability we deemed to be $-q < 0$, the price we consider fair to pay need not be $p = |-q|$. If $p \neq |-q|$, it means that we changed idea about how likely event A is once given the possibility of entering the bet. If instead $p = |-q|$, we are obeying to Allen's principle of conservation of knowledge [12]. It states that, like the law of conservation of mass, knowledge is neither created nor destroyed, but only transformed; its total possible amount is constant. By having us choose the probability equal in absolute value, we comply with conservation of knowledge. As we can see, this principle allows us to mechanically retrieve positive probabilities starting from negative ones. It also allows us to work backwards to negative probabilities starting from positive ones. Suppose we have an event $A \in \mathcal{F}$ to which we attach probability $p \geq 0$. Then, we know that if we were not given the opportunity to enter the bet that allowed us to indicate p as the probability that event A happens, then we would have expressed our uncertainty by assigning A probability $-p \leq 0$.

A Dutch book is, informally, the possibility of constructing a bet such that the bookmaker always profits, while the punter always loses money (see Remark 2 for a formal definition). In [31], a coherent subjective probability is defined as one that does not allow for a Dutch book to be made against the punter, however a bet is made. Necessary and sufficient conditions for coherence require that subjective probabilities satisfy the Kolmogorovian axioms of probability (with only finite additivity). As we shall see in section 2.2, this does not hold for extended probabilities. This is not a problem though, in light of the interpretation we give to negative probabilities. The events whose attached probabilities are negative are events for which the

^bProvided that we gain $\$(1 - q)$ if A happens.

punter cannot enter a bet; hence, they cannot be used to build a Dutch book. We give the definition of coherence in the context of extended probabilities and the formal statement that extended probabilities are always coherent in Remark 2. In addition, as we show in section 3, the inferential procedure we consider is such that, starting with extended probabilities, we recover – at the end of our analysis – regular probabilities. This also ensures that no Dutch books can be created.

2.2. Technical definition and properties

Consider a measurable space (Ω, \mathcal{F}) , where $\mathcal{F} \subset 2^\Omega$ is a sigma-algebra. An extended probability $P^{ex} : \mathcal{F} \rightarrow \mathbb{R}$ is a set function on \mathcal{F} such that

- (i*) $P^{ex}(A) \in [-1, 1]$, for all $A \in \mathcal{F}$;
- (ii*) if $\{A_j\}_{j \in I}$ is a countable collection of disjoint events such that $\cup_{j \in I} A_j \in \mathcal{F}$, then

$$P^{ex} \left(\bigcup_{j \in I} A_j \right) = \sum_{j \in I} P^{ex}(A_j);$$

- (iii*) given any partition $\mathcal{E} = \{E\}$ of Ω such that $E \in \mathcal{F}$, for all $E \in \mathcal{E}$, $\sum_{E \in \mathcal{E}} |P^{ex}(E)| = 1$.

Condition (iii*) is needed in light of the interpretation we gave in section 2.1 to extended probabilities: if (iii*) were not to hold, the doppelgänger's opinion would be represented by a signed measure, not a probability measure.

Let $\mathcal{F} \subset 2^\Omega$ be the same sigma-algebra as before. The Kolmogorovian axioms for any regular probability measure $P : \mathcal{F} \rightarrow \mathbb{R}$ are the following

- (K1) $P(A) \in [0, 1]$, for all $A \in \mathcal{F}$;
- (K2) if $\{A_j\}_{j \in I}$ is a countable collection of disjoint events such that $\cup_{j \in I} A_j \in \mathcal{F}$, then $P(\cup_{j \in I} A_j) = \sum_{j \in I} P(A_j)$;^c
- (K3) $P(\Omega) = 1$.

It is easy to see, then, how conditions (i*) and (iii*) are more general than (K1) and (K3), respectively. To this extent, extended probabilities are a generalization of the concept of regular probabilities. From its definition, we see that an extended probability is a finite signed measure. We call the triple $(\Omega, \mathcal{F}, P^{ex})$ an *extended probability space*.

^cWe consider the countably additive version of Kolmogorov axioms.

Because we do not require $P^{ex}(\Omega) = 1$, the extended probability of the complement of an event A is given by

$$P^{ex}(A^c) = P^{ex}(\Omega \setminus A) = P^{ex}(\Omega) - P^{ex}(A). \quad (1)$$

Equation (1) comes from (ii*); indeed, consider the disjoint sets $\Omega \setminus A$ and $\Omega \cap A = A$. Then,

$$\begin{aligned} P^{ex}(\Omega) &= P^{ex}([\Omega \setminus A] \sqcup A) = P^{ex}(\Omega \setminus A) + P^{ex}(A) \\ \iff P^{ex}(\Omega \setminus A) &= P^{ex}(\Omega) - P^{ex}(A). \end{aligned}$$

Equation (1) ensures us that $P^{ex}(\emptyset) = P^{ex}(\Omega^c) = P^{ex}(\Omega) - P^{ex}(\Omega) = 0$. We also have that $P^{ex}(\Omega) = \sum_{E_j \in \mathcal{E}} P^{ex}(E_j)$, where $\mathcal{E} = \{E_j\}$ is any finite or countable partition of Ω .

Proposition 1. *Extended probabilities have the adequacy property: for all $A, B \in \mathcal{F}$ such that $A = B$, then $P^{ex}(A) = P^{ex}(B)$.*

We also have that extended probabilities retain a version of the monotonic property of regular probabilities.

Proposition 2. *For all $A, B \in \mathcal{F}$ such that $A \subset B$, if $P^{ex}(A) \geq 0$, $P^{ex}(B) \geq 0$, and $P^{ex}(B \cap A^c) \geq 0$, then $P^{ex}(A) \leq P^{ex}(B)$. If instead $P^{ex}(A) \leq 0$, $P^{ex}(B) \leq 0$, and $P^{ex}(B \cap A^c) \leq 0$, then $P^{ex}(A) \geq P^{ex}(B)$.*

This version of the monotonic property implies a version of the continuity property enjoyed by regular probabilities.

Corollary 1. *If we have a collection $\{A_j\}$ of elements of \mathcal{F} such that*

- *it is nested (i.e. $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$),*
- $\bigcap_j A_j = \emptyset$,
- $P^{ex}(A_j) \geq 0$, for all A_j ,

then $\lim_{j \rightarrow \infty} P^{ex}(A_j) = P^{ex}(\bigcap_j A_j) = 0$.

Extended probabilities satisfy the inclusion-exclusion principle.

Proposition 3. *The following is true:*

$$P^{ex}(A \cup B) = P^{ex}(A) + P^{ex}(B) - P^{ex}(A \cap B),$$

for all $A, B \in \mathcal{F}$.

$$\begin{aligned} \text{This implies immediately that, for any } A, B, C \in \mathcal{F}, \\ P^{ex}((A \cup B) \cap C) &= P^{ex}((A \cap C) \cup (B \cap C)) \\ &= P^{ex}(A \cap C) + P^{ex}(B \cap C) - P^{ex}(A \cap B \cap C), \end{aligned}$$

by Proposition 3 and the De Morgan laws.

We define the extended conditional probability of A given B to be the extended counterpart of regular conditional probabilities,

$$P^{ex}(A | B) := \frac{P^{ex}(A \cap B)}{P^{ex}(B)} \in [-1, 1], \quad (2)$$

for all $A, B \in \mathcal{F}$ such that $P^{ex}(B) \neq 0$.^d

To avoid confusion arising from the sign, we say that $A, B \in \mathcal{F}$ are independent if and only if

$$|P^{ex}(A \cap B)| = |P^{ex}(A) \times P^{ex}(B)|.$$

This way of describing independence corresponds to the one given in [12, section 2].

Remark 2. Consider any extended probability space $(\Omega, \mathcal{F}, P^{ex})$. Call \mathcal{B} the sigma-algebra generated by the events B in Ω that we can enter a bet about, that is, the sigma-algebra generated by the collection $\{B \subset \Omega : P^{ex}(B) \geq 0 \text{ and } B \cap C = \emptyset, \forall C \in \mathcal{F}, P^{ex}(C) < 0\}$. Then, a *Dutch book* is a finite collection $\{B_j\}_{j=1}^n \subset \mathcal{B}$ along with numbers $\{s_j\}_{j=1}^n \subset \mathbb{R}$ such that,

$$\sup_{\omega \in \Omega} f(\omega) := \sup_{\omega \in \Omega} \left\{ \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - P^{ex}(B_j)] \right\} < 0. \quad (3)$$

Notice that the s_j 's are the payout for a winning bet on B_j , and $s_j P^{ex}(B_j)$ is the bet's fair buy-in. The inequality in (3) suggests that a bet can be built such that the bookmaker always gets a profit, while the punter always loses money.

Definition 1. Extended probability P^{ex} is coherent if no Dutch books can be made against the punter.

Then, we have the following important result.

Theorem 1. P^{ex} is always coherent.

In our definition of Dutch book, we do not take into account the events $C \in \mathcal{F}$ for which $P^{ex}(C) < 0$ since those are events for which the punter cannot enter a bet. Hence, they cannot be used to build a Dutch book.

^dNotice that in de Finetti's approach [23, 27] a similar formula follows from the "called off bet" interpretation of (regular) conditional probability. In the future, we plan to provide a direct behavioral motivation for (2), in the spirit of de Finetti's work.

2.3. Related literature

Let us now inspect how the definition and the interpretation of extended probabilities we have given so far relates to the existing literature.

Our framework is related to the one studied in [15]. Burgin studies a static environment (as opposed to the dynamic one we inspect in this work in the next section) where the state space Ω can be divided in two irreducible parts Ω^+ and Ω^- such that $\#\Omega^+ = \#\Omega^-$, where $\#$ denotes the cardinality operator. The elements of Ω^- are called anti-events, and they are usually connected to negative objects: encountering a negative object is a negative event. The author then states that an example of a negative object is given by antiparticles, the antimatter counterpart of quantum particles. Anti-events are given negative probabilities. If we require that the following conditions hold, we obtain a framework very similar to Burgin's one:

- extended probabilities are finitely additive (instead of countably additive);
- our state space can be divided in two irreducible parts Ω^+ and Ω^- ;
- $P^{ex}(A) \geq 0$ if and only if $A \subset \Omega^+$;
- there exists a function $\alpha : \Omega \rightarrow \Omega$ such that $\alpha(\omega) = -\omega$ and $\alpha^2(\omega) = \omega$;
- $P^{ex}(\Omega^+) = 1$;
- $\{v_i, \omega, -\omega : v_i, \omega \in \Omega, i \in I\} = \{v_i : v_i \in \Omega, i \in I\}$, for all $\omega \in \Omega$ and all set of indices I .

The main difference is that in [15] $P^{ex}(\Omega^+) + |P^{ex}(\Omega^-)| = 2$. Our interpretation of negative probabilities reconciles with Burgin's one if we consider anti-events as events for which we cannot enter a bet, which seems a reasonable assumption. He also gives a frequentist interpretation of negative probabilities, that complements our interpretation of negative extended probabilities.

In a subsequent paper [14], Burgin generalizes his own setup, mainly by not requiring $\#\Omega^+ = \#\Omega^-$ and by allowing any event to have either positive or negative probability, depending on external conditions. If we require that the following conditions hold, we obtain Burgin's generalized framework:

- extended probabilities are finitely additive;
- $A \in \mathcal{F}$ implies $-A := \{-\omega : \omega \in A\} \in \mathcal{F}$;
- for all $A \in \mathcal{F}$, $P^{ex}(-A) = -P^{ex}(A)$.

Notice that this last condition is very similar to Allen's principle of conservation of knowledge. The interpretation he gives for negative probabilities is similar to the one given in his previous work, with the peculiarity that now negative probabilities are not assigned exclusively to anti-events. Our interpretation can be seen as being a step behind Burgin's one: we give a specific reason for why an event \tilde{A} is assigned a negative probability, namely that we cannot enter a bet on it. Then, Burgin states that the anti-event of \tilde{A} , $-\tilde{A}$, will have a positive (regular) probability. Of course, the vice versa holds: if an event is assigned a positive probability (because we can enter the bet), then according to Burgin its anti-event will have a negative probability.

Another interesting interpretation is given by Székely in [32]. The author proves that we can encounter a negative probability if we work with a random variable having a signed distribution. In addition, if X has a signed distribution, then there exist two random variables Y, Z having an ordinary (not signed) distribution such that $X + Y = Z$ in distribution. Therefore X can be seen as a "difference" of two ordinary random variables Z and Y . We reconcile our interpretation and Székely's one as follows. A signed distribution P^s is simply an extended probability on the space of outcomes of a random variable. A pullback argument can then be used to define an extended probability P^{ex} on Ω : $P^{ex}(X^{-1}(I)) = P^s(I)$, for all subsets I of the outcome space of random variable X . So there are going to be events in Ω having negative probabilities: those are events we cannot enter a bet about.

The interpretation Kronz [6] gives of negative probabilities is the following: he calls negative probabilities inferred probabilities, that can only be obtained indirectly by inference from operational (regular) probabilities. The associated events (that is, events having negative probabilities) are called virtual events in that they are non-operational, and so do not give rise to directly accessible relative frequencies. Actual events have non-negligible effect on them, and the reverse is also true. As it appears clear, virtual events are equivalent to latent events e.g. in the psychology [33], economics [34] and medicine [35] literatures: Kronz assigns negative probabilities to latent events. These latter are events we do not observe, so it is fair to think we cannot enter a bet involving them.

In [36], Wigner allows probabilities to go negative to sidestep the uncertainty principle of quantum mechanics. This latter states that given a particle that moves in one dimension, there is a limit to the precision with which position x and momentum p can be determined simultaneously in

a given state. The observer can know the distribution of x and that of p , but there is no joint probability distribution of (x, p) with the correct marginals of x and p . To overcome this shortcoming, Wigner came up with a quasiprobability distribution $W_\psi(x, p)$, that is, a countably additive function on the measurable subsets of state space $\Omega = \mathbb{R}^2$ such that $W_\psi(\Omega) = 1$. Wigner's function $W_\psi(x, p)$ is a signed probability distribution; all values of $W_\psi(x, p)$ are real, but some values may be negative. In [37], the authors show that Wigner's quasiprobability distribution is the unique signed probability distribution yielding the correct marginal distributions for position and momentum and all their linear combinations, and in [38], the authors show that Wigner's function can be tested using the canonical approach of game-theoretic probability. We can give Wigner's function two interpretations. The first one is purely "utilitarian": because it is the unique quasiprobability distribution that gives the correct marginal distributions for x and p , scientists should use it even if it allows for probabilities being negative-valued. The second one can be reconciled with the interpretation we give in the present work to negative extended probabilities. Because a scientist cannot observe x and p simultaneously, it is fair to think that they cannot enter a bet about events involving both position and momentum of a particle; hence, they resort to negative probabilities.

In [39], the authors study bounds on the algorithmic capabilities of a mathematical theory and analyze their implications. In particular, they require that the theory is logically consistent – that is, it has to be based on a few axioms and rules from which mathematical truths can be unambiguously derived – and that inferences in the theory should be computable in polynomial time – that is, there should be an efficient way to execute the theory. The postulates of consistency and computation are apparently in conflict with each other: intuitively, if only polynomial time computations are allowed, the theory is consistent only up to what polynomial calculus allows, an instance that the authors call external-internal clash. They formalise such a clash via a "weirdness theorem", which shows that any theory obeying their two postulates necessarily departs in a very peculiar way from the probabilistic point of view. In particular, the theorem proves that all models compatible with the theory will present some negative probabilities (quasiprobabilities à la Wigner). They also show that quantum paradoxes are a consequence of the weirdness theorem. To develop their results, the authors rely on a characterization of probability in terms of lotteries (or gambles, a particular type of bets). They then provide a subjective foundation, à la de Finetti, of so-called generalised probability

theories. Negative probabilities arise when an agent would like to enter a given bet involving an event, but computational limitations related to the event prevent them from doing so. This interpretation coincides with the one we give in the present chapter: the authors give a reason for which an agent is denied the possibility of entering a bet.

Finally, the interpretation of negative probabilities in [17] and [5, 18] is very similar – although not identical – to the one in the present work. We give the summary of their way of interpreting negative probabilities as reported in [40]. First, they point out that the probabilities dictated by a physical theory instruct (rational) agents on how to bet on the outcomes of phenomena. Standard arguments from subjective Bayesian probability theory – the Dutch book arguments – demand that the Kolmogorovian axioms for classical probability theory must hold for the degrees of belief of any rational agent, which determine which bets the agent regard as fair. However, they point out that these arguments only apply to bets which are settleable, that is, bets about events the agent will certainly know at some point. They then argue that only bets on sufficiently coarse-grained alternative histories will be settleable, where this coarse-graining guarantees that the alternatives’ probabilities will lie in the unit interval. The main difference between their interpretation and the one presented in this chapter is that for Gell-Mann and Hartle the bets on events that have negative probabilities cannot be settled, whereas we deem those bets to be settleable. The agent simply cannot enter them at the moment, but they may be given the opportunity in the future.

3. Extended probabilities in statistical inference

In this section, we are going to consider an ex ante analysis in which we progressively learn the composition of the state space. It is ex ante in that it takes place before the actual statistical analysis. As time goes by, we collect new observations via a learning procedure specified in advance, e.g. an urn with or without replacement.

Consider a measurable space (Ω, \mathcal{F}) with Ω at most countable and $\mathcal{F} = 2^\Omega$. At any time t , we have $\Omega = \Omega_t^- \sqcup \Omega_t^+$. For all $t \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, Ω_t^- represents the “latent” part of Ω at time t , that is, the part that we have not yet observed. Ω_t^+ represents the “actual” part of Ω , that is, the portion of Ω that we have observed at time t (at time $t = 0$, Ω_0^+ is the portion of Ω that we know ex ante, which is assumed nonempty). This approach is similar to the “ex ante humility” one introduced in [12, section

1], where latent and actual portions of the state space are first introduced, and a dynamic is described. A graphical representation of $\Omega = \Omega_t^- \sqcup \Omega_t^+$ is given in Figure 1. At time $t = 0$, we specify the finest possible partition

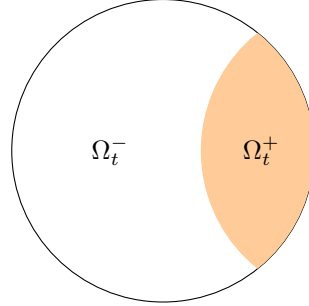


Fig. 1. Graphical representation of $\Omega = \Omega_t^- \sqcup \Omega_t^+$.

of Ω , $\mathcal{E} = \{\{\omega\}\}_{\omega \in \Omega}$. It is such that $\mathcal{E} = \mathcal{E}_0^- \sqcup \mathcal{E}_0^+$; in this notation, \mathcal{E}_0^- partitions Ω_0^- , and \mathcal{E}_0^+ partitions Ω_0^+ . The most general way of proceeding is by specifying the number type (natural numbers \mathbb{N} , whole numbers \mathbb{N}_0 , integers \mathbb{Z} , or rationals \mathbb{Q}) we subjectively believe we are working with, and setting it to be Ω . To this extent, we give to Ω the apparently possible interpretation of [24, section 2.1.2]: Ω is the space of apparently possible states if it contains all the states ω that we believe are logically consistent with our available information. To give an example, a space of apparently possible states associated with a coin toss is

$$\Omega = \{\text{heads, tails, coin landing on its edge, coin breaking into pieces on landing, coin disappearing down a crack in the floor}\}.$$

Considering the space of apparently possible states, then, amounts to considering the broadest state space associated with the statistical experiment of interest.

After that, we subjectively specify what we initially think the state space is, and we set it to be Ω_0^+ ; for example, we may have previous information coming from similar (but not equal) experiments.

Given the sequences (Ω_t^+) and (Ω_t^-) , we require that, for all t ,

$$\Omega_t^+ \subset \Omega_{t+1}^+ \quad \text{and} \quad \Omega_t^- \supset \Omega_{t+1}^-.$$

This means that (Ω_t^+) is monotone nondecreasing, and (Ω_t^-) is monotone nonincreasing, which implies that the limits of both exist and are well

defined. In particular, we have that

$$\lim_{t \rightarrow \infty} \Omega_t^+ = \bigcup_{t \in \mathbb{N}_0} \Omega_t^+ \quad \text{and} \quad \lim_{t \rightarrow \infty} \Omega_t^- = \bigcap_{t \in \mathbb{N}_0} \Omega_t^-.$$

We then have two possible scenarios. In the first one, $\bigcup_{t \in \mathbb{N}_0} \Omega_t^+ = \Omega$, which implies that $\bigcap_{t \in \mathbb{N}_0} \Omega_t^- = \emptyset$. In the second one, $\bigcup_{t \in \mathbb{N}_0} \Omega_t^+ \equiv \Omega' \subsetneq \Omega$, which implies that $\bigcap_{t \in \mathbb{N}_0} \Omega_t^- \equiv \Omega'' := \Omega \setminus \Omega' \neq \emptyset$.

As we collect more and more observations, we progressively discover the composition of our sample space. In the first scenario, in the limit we discover that the sample space associated with our experiment corresponds to the whole set Ω we initially specified. In this scenario, at time $t+1$ when an element $\tilde{\omega}$ is observed that belongs to Ω_t^- , the element $E_{t,\tilde{\omega}}^- = \{\tilde{\omega}\}$ of the partition \mathcal{E}_t^- consistent with $\tilde{\omega}$ becomes an element of \mathcal{E}_{t+1}^+ . We use the following notation: $E_{t+1,\tilde{\omega}}^+ = E_{t,\tilde{\omega}}^-$. So the partition of the latent space loses an element, while the partition of the actual space gains one. This means that, for all t , $\mathcal{E}_t^+ \subset \mathcal{E}_{t+1}^+$ and $\mathcal{E}_t^- \supset \mathcal{E}_{t+1}^-$. We abuse notation: we write $\mathcal{E}_t^+ \subset \mathcal{E}_{t+1}^+$ to indicate that $\#\mathcal{E}_t^+ \leq \#\mathcal{E}_{t+1}^+$, where $\#\mathcal{E}_t^+$ denotes the number of elements of the partition \mathcal{E}_t^+ , for all t , and we write $\mathcal{E}_t^- \supset \mathcal{E}_{t+1}^-$ to indicate that $\#\mathcal{E}_t^- \geq \#\mathcal{E}_{t+1}^-$. Hence, in the limit, the partition \mathcal{E}_t^+ of the ‘‘actual’’ space coincides with the partition \mathcal{E} of the whole set Ω . This mirrors the behavior of Ω_t^+ , that converges to Ω .

In the second scenario Ω' can be finite or countable. In the former case, there exists a $T \in \mathbb{N}$ after which the observations we collect at time $T+i$ already belong to Ω_T^+ , for all $i \in \mathbb{N}$. If that is the case, we write

$$\Omega' \equiv \Omega_T^+. \quad (4)$$

This corresponds to discovering that the actual sample space is finite and smaller than the whole set Ω that we specified at the beginning of our analysis. If Ω' is countable, we can only say that it is a proper subset of Ω , and that we discover its composition in the limit. This may happen, for example, if we begin our analysis by setting $\Omega = \mathbb{N}_0$, but then realize that the state space associated with our experiment is actually $\Omega' = \mathbb{N}$.

In general, when the second scenario takes place, we have $\Omega_t^+ \uparrow \Omega'$, and $\Omega_t^- \downarrow \Omega'' := \Omega \setminus \Omega'$. We also have that, in the limit, the partition \mathcal{E}_t^+ of the ‘‘actual’’ space coincides with the (finest possible) partition \mathcal{E}' of set Ω' . We call \mathcal{E}'' the partition whose elements belong to \mathcal{E} but not to \mathcal{E}' ; again abusing notation, we write $\mathcal{E}'' := \mathcal{E} \setminus \mathcal{E}'$. We call $\mathcal{F}' = 2^{\Omega'}$ and $\mathcal{F}'' = 2^{\Omega''}$.

Remark 3. In this chapter, our ex ante analysis is concluded by finding the true composition of the state space. We can interpret this using the concept

of a benevolent bookmaker. While our doppelgänger is always able to enter a bet on all the events in the space Ω of apparently possible states, we end up only able to bet on the events of Ω that are crucial to the statistical analysis taking place after our ex ante analysis. This is akin to a benevolent bookmaker preventing us from entering bets on irrelevant events, that, if we were to bet on, would certainly make us lose money.^e

However, there may be cases in which the state space does not get fully discovered. For example, this may happen if we have an urn with a timer attached; once the time runs out, the urn is sealed, so that we do not discover its entire composition. The following statistical analysis, which would require the use of extended probabilities, is explored in a future work. This case corresponds to the existence of a malevolent bookmaker that does not allow us to enter bets that are crucial to the statistical analysis.^f This setting is especially important when studying events about which we will never be able to enter a bet, for example latent events.

Remark 4. If at any time t we collect an observation $\tilde{\omega}$ that does not belong to Ω , this means that the space Ω we initially specified is not rich enough. We have then to specify a richer, larger set $\check{\Omega} \supset \Omega$, and start our analysis over. We can either consider $\check{\Omega} = \Omega \cup \{\tilde{\omega}\}$, or define $\check{\Omega}$ as a larger number type. Let us give an example. Suppose we begin our analysis by setting $\Omega = \mathbb{N}$, and after a while we observe $\tilde{\omega} = 1/2$. Then, we have to restart our analysis and we can either let $\check{\Omega}$ be $\mathbb{N} \cup \{1/2\}$ or $\check{\Omega} = \mathbb{Q}$.

It may also happen that our true sample space is $\check{\Omega} \subsetneq \Omega_0^+$. In this case, equation (4) holds with $T = 0$, and our analysis would still be valid. The drawback is that, if that happens, we are not respecting one of the conditions listed in [41] for a sample space to be valid. In particular, that the sample space Ω must have the right granularity depending on what we are interested in. This means that we must remove irrelevant information from the sample space. In other words, we must choose the right abstraction and forget irrelevant information. This issue can be avoided by initially specifying Ω_0^+ so that it has the fewest possible elements.

On top of dealing with uncertainty on the composition of the sample space, we also address the problem of not being able to specify a unique (extended) probability measure on Ω . That is why we are going to work

^eA more correct expression would be “almost benevolent” since we are considering only one-way bets.

^fHere “malevolence” has to be understood in terms of restricting the menu of available options to the agent.

with sets of extended probability measures. We call this approach *extended sensitivity analysis*, since it corresponds to the extended probabilities counterpart of Bayesian sensitivity analysis [42]. We begin by considering a set $\mathcal{P}^{ex} \equiv \mathcal{P}_0^{ex}$ of extended probabilities that represent the agent's initial beliefs, and we update it as described in section 3.2. We denote the sequence of successive updates of \mathcal{P}_0^{ex} as $(\mathcal{P}_t^{ex})_{t \in \mathbb{N}}$. By working with sets of extended probability measures, we represent the condition of a researcher facing a decision under ambiguity [43]. As we shall see in section 3.2, sequence $(\mathcal{P}_t^{ex})_{t \in \mathbb{N}_0}$ converges in the Hausdorff metric.

Now, fix any $t \in \mathbb{N}_0$, and consider any $P_t^{ex} \in \mathcal{P}_t^{ex}$. We require the following

- (i) $P_t^{ex}(A) \in [0, 1]$ if $A \subset \Omega_t^+$;
- (ii) $P_t^{ex}(A) \in [-1, 0]$ if $A \subset \Omega_t^-$;
- (iii) $P_t^{ex}(A) \in [-1, 1]$ if $A \cap \Omega_t^+ \neq \emptyset \neq A \cap \Omega_t^-$.

In particular, we compute the latter as follows.

Proposition 4. *The following is true.*

$$\begin{aligned} P_t^{ex}(A) = & \sum_{E_{t,j}^+ \in \mathcal{E}_t^+ : P_t^{ex}(E_{t,j}^+) \neq 0} P_t^{ex}(A | E_{t,j}^+) P_t^{ex}(E_{t,j}^+) \\ & + \sum_{E_{t,j}^- \in \mathcal{E}_t^- : P_t^{ex}(E_{t,j}^-) \neq 0} P_t^{ex}(A | E_{t,j}^-) P_t^{ex}(E_{t,j}^-). \end{aligned} \quad (5)$$

Notice that we need to consider the elements of \mathcal{E}_t^+ and \mathcal{E}_t^- whose extended probabilities are not 0 otherwise the conditional extended probabilities $P_t^{ex}(A | E_{t,j}^+)$ and $P_t^{ex}(A | E_{t,j}^-)$ may result in an indeterminate form of the $\frac{0}{0}$ kind. This requirement yields no loss of generality since if an element of a partition is assigned extended probability 0, then it does not convey any information around event A . Notice also that, for the elements of \mathcal{E}_t^+ and \mathcal{E}_t^- whose extended probabilities are not 0, $P_t^{ex}(A | E_{t,j}^+) = \frac{P_t^{ex}(A \cap E_{t,j}^+)}{P_t^{ex}(E_{t,j}^+)} \geq 0$ by (i), and $P_t^{ex}(A | E_{t,j}^-) = \frac{P_t^{ex}(A \cap E_{t,j}^-)}{P_t^{ex}(E_{t,j}^-)} \geq 0$ because it is the ratio of two negative quantities; once multiplied by $P_t^{ex}(E_{t,j}^-)$, which is negative by (ii), it gives us a negative value. So the sign of $P_t^{ex}(A)$ is not predetermined when $A \cap \Omega_t^+ \neq \emptyset \neq A \cap \Omega_t^-$. The interpretation of these conditions is straightforward: we assign negative extended probabilities to events that belong to the latent space at time t (meaning that at time t we cannot enter a bet about them), while we assign positive extended probabilities to events that are in

the actual, observed space (meaning that at time t we can enter a bet about them). If a given event is only partially known, its extended probability has not a predetermined sign: it will depend on whether we know enough about it (then the probability will be positive), or not (vice versa). This means that we can enter a bet about “sub-event” $A \cap \Omega_t^+$, but not about $A \cap \Omega_t^-$. For example, let $A = \{\text{tomorrow there will be a thunderstorm}\}$. Then, for some t , suppose that

$$\begin{aligned} A \cap \Omega_t^+ &= \{\text{tomorrow will rain}\}, \\ A \cap \Omega_t^- &= \{\text{tomorrow there will be a dry thunderstorm}\}. \end{aligned}$$

Then we can place a bet on $A \cap \Omega_t^+$ but not on $A \cap \Omega_t^-$, so the extended probability we assign to A does not have a predetermined sign.

Notice also that we can define a set of events $\mathcal{C}_{P_t^{ex}} := \{A \in \mathcal{F} : P_t^{ex}(A \cap \Omega_t^-) = -P_t^{ex}(A \cap \Omega_t^+)\}$, which we call critical events according to P_t^{ex} , with the property that $P_t^{ex}(A) = 0$, for all $A \in \mathcal{C}_{P_t^{ex}}$ (immediate from the definition). We assign the sub-event we can enter a bet about the same probability that our doppelgänger assigns to the sub-event that we are not allowed to bet on. This means that we deem their “actual” portion (the one we know/we have observed so far) to be just as likely than their “latent” portion (the one we do not know/we have not yet observed). The set of critical events according to the whole set of extended probability measures \mathcal{P}_t^{ex} is given by

$$\mathcal{C}_t := \bigcap_{P_t^{ex} \in \mathcal{P}_t^{ex}} \mathcal{C}_{P_t^{ex}}.$$

3.1. Properties of this environment

We now give some results concerning the environment we depicted so far.

Proposition 5. *Let $A_t^+ := A \cap \Omega_t^+$, $A_t^- := A \cap \Omega_t^-$, $B_t^+ := B \cap \Omega_t^+$, and $B_t^- := B \cap \Omega_t^-$. Then, $A \cup B = (A_t^+ \cup B_t^+) \cup (A_t^- \cup B_t^-)$ and $A \cap B = (A_t^+ \cap B_t^+) \cup (A_t^- \cap B_t^-)$, for all t . Also, $A \setminus B = (A_t^+ \setminus B_t^+) \cup (A_t^- \setminus B_t^-)$, for all t .*

Recall now that a set ring is a system of sets \mathbb{B} such that $A, B \in \mathbb{B}$ implies $A \cap B \in \mathbb{B}$ and $(A \setminus B) \cup (B \setminus A) =: A \Delta B \in \mathbb{B}$. A set ring \mathbb{B} with a unit element, i.e. $E \in \mathbb{B}$ such that for all $A \in \mathbb{B}$, $A \cap E = A$, is called a set algebra. Let $\mathcal{F}_t^+ = 2^{\Omega_t^+}$ and $\mathcal{F}_t^- = 2^{\Omega_t^-}$.

Proposition 6. *\mathcal{F}_t^+ and \mathcal{F}_t^- are set algebras, for all t .*

We also point out that if $A = \{\omega_1, \dots, \omega_k\}$ and $\omega_1, \dots, \omega_k \in \Omega$, then

$$P^{ex}(A) = \sum_{j=1}^k P^{ex}(\{\omega_j\}).$$

This is immediate from the countable additivity of extended probabilities.

Another property is the following. Fix any t and let $A, B \in \mathcal{F}_t^+$ such that $A \subset B$. Then, $P_t^{ex}(A) \leq P_t^{ex}(B)$. Let then $C, D \in \mathcal{F}_t^-$ such that $C \subset D$. Then, $P_t^{ex}(C) \geq P_t^{ex}(D)$. Both these results come from Proposition 2.

The following is also interesting.

Proposition 7. *Consider $A \subset \cup_{j \in \mathbb{N}_0} A_j$, where $A_j \in \mathcal{F}$ for all j , and also $A \in \mathcal{F}$. Then $P_t^{ex}(A) \leq \sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j)$ if $\cup_{j \in \mathbb{N}_0} A_j \in \mathcal{F}_t^+$, and $P_t^{ex}(A) \geq \sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j)$ if $\cup_{j \in \mathbb{N}_0} A_j \in \mathcal{F}_t^-$. If instead some of the A_j 's are in \mathcal{F}_t^+ and some are in \mathcal{F}_t^- , then $P_t^{ex}(A) \gtrless \sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j)$, that is, $P_t^{ex}(A)$ can be larger, smaller, or equal to $\sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j)$.*

In the setting we have outlined so far, there is a way of operationalizing equation (1). Pick any $A \in \mathcal{F}$; we have

$$P_t^{ex}(A^c) = P_t^{ex}(\Omega_t^+ \setminus [\Omega_t^+ \cap A]) + P_t^{ex}(\Omega_t^- \setminus [\Omega_t^- \cap A]) \quad (6)$$

We retain the fact that $P_t^{ex}(\emptyset) = 0$; indeed

$$\begin{aligned} P_t^{ex}(\Omega^c) &= P_t^{ex}(\emptyset) = P_t^{ex}(\Omega_t^+ \setminus [\Omega_t^+ \cap \Omega]) + P_t^{ex}(\Omega_t^- \setminus [\Omega_t^- \cap \Omega]) \\ &= P_t^{ex}(\emptyset) + P_t^{ex}(\emptyset) \\ &\iff P_t^{ex}(\emptyset) = 0. \end{aligned}$$

We can also write $P_t^{ex}(A^c) = P_t^{ex}(A^c \cap \Omega_t^+) + P_t^{ex}(A^c \cap \Omega_t^-)$, because, as we know, $A^c = (A^c \cap \Omega_t^+) \sqcup (A^c \cap \Omega_t^-)$.

3.2. Interpretation and updating procedure

Let us now discuss the probability assigned to the whole sample space Ω . From (ii*), we know that since, for all t , $\Omega_t^+ \sqcup \Omega_t^- = \Omega$, then $P_t^{ex}(\Omega) = P_t^{ex}(\Omega_t^+) + P_t^{ex}(\Omega_t^-)$. Also, from (i) and (ii) we know that $P_t^{ex}(\Omega_t^+) \geq 0$ and $P_t^{ex}(\Omega_t^-) \leq 0$. So

$$P_t^{ex}(\Omega) = \begin{cases} p > 0 & \text{if } P_t^{ex}(\Omega_t^+) > |P_t^{ex}(\Omega_t^-)| \\ p = 0 & \text{if } P_t^{ex}(\Omega_t^+) = |P_t^{ex}(\Omega_t^-)| \\ p < 0 & \text{if } P_t^{ex}(\Omega_t^+) < |P_t^{ex}(\Omega_t^-)| \end{cases}$$

What does it mean, then, for $P_t^{ex}(\Omega)$ to be equal to 0? And to be negative? And to be positive, but not 1? Given the benevolent bookmaker

interpretation, we have the following. If $P_t^{ex}(\Omega) = 0$, it means that we have no sufficient information to say whether the sample space associated with our experiment is in fact Ω . If $P_t^{ex}(\Omega) < 0$, it means that, for the time being, the sample space associated with our experiment appears to be some $\tilde{\Omega} \subsetneq \Omega$. If $P_t^{ex}(\Omega) \in (0, 1)$, it means that there is evidence that the sample space associated with our experiment could be in fact Ω , but we cannot state it with certainty.

The natural question that one might ask now is, for some $t \in \mathbb{N}_0$, how do we come up with negative numbers to assign to events that belong to the latent space Ω_t^- . Or, for that matter, how do we come up with positive numbers to assign to events that belong to the actual space Ω_t^+ . Call $\Delta(\Omega, \mathcal{F})$ the set of probability measures on (Ω, \mathcal{F}) , and $\Delta^{ex}(\Omega, \mathcal{F})$ the set of extended probability measures on (Ω, \mathcal{F}) . The latter is a linear space, as shown in [44]. In a future work, we will argue that it is a Dedekind complete Banach lattice with respect to the norm induced by the total variation of an element P^{ex} of $\Delta^{ex}(\Omega, \mathcal{F})$. We will also show that if Ω is a compact separable space, then the subset $\Delta_{Baire}^{ex}(\Omega, \mathcal{F}) \subset \Delta^{ex}(\Omega, \mathcal{F})$ of Baire extended probability measures is the dual of the real Banach space of all continuous real-valued functions on Ω .

Consider any $P \in \Delta(\Omega, \mathcal{F})$ such that, for all $\tilde{A} \in \mathcal{F}_0^+$, $P(\tilde{A}) = p \in [0, 1]$ is the amount we deem fair to pay to enter a bet about \tilde{A} ; we call it the *oracle probability measure*. Then, for a generic $A \in \mathcal{F}$ we have that

$$P_0^{ex}(A \cap \Omega_0^+) = P(A \cap \Omega_0^+) \quad (7)$$

and

$$P_0^{ex}(A \cap \Omega_0^-) = -P(A \cap \Omega_0^-). \quad (8)$$

It is easy to see that P_0^{ex} satisfies (i*) and (ii*). Notice also that

$$\sum_{E \in \mathcal{E}} |P_0^{ex}(E)| = \sum_{E_0^+ \in \mathcal{E}_0^+} P_0^{ex}(E_0^+) + \sum_{E_0^- \in \mathcal{E}_0^-} |P_0^{ex}(E_0^-)| = 1, \quad (9)$$

so P_0^{ex} satisfies (iii*) as well. It is hence a properly defined extended probability measure. Clearly, (9) holds for all $t \in \mathbb{N}_0$, not just for $t = 0$.

This way of assessing initial extended probabilities well reconciles with the interpretation we gave in general for extended probabilities: we cannot enter bets about events $A \notin \mathcal{F}_0^+$ (in this case, because we cannot observe them for the time being). So we assess the probabilities of the events $A \in \mathcal{F}_0^+$ as specified in section 2, and then we ask our doppelgänger the probabilities $P_{0,D}(B) \geq 0$ they assign to the elements $B \in \mathcal{F}_0^-$.⁸ Then, we flip the sign

⁸Subscript D stands for “doppelgänger”.

to those, that is, $P_0^{ex}(B) = -P_{0,D}(B)$, for all $B \in \mathcal{F}_0^-$. In this notation, $P_{0,D}(B)$ is the probability the doppelgänger assigns to event B at time $t = 0$.

Because the agent faces ambiguity, they need to specify a set $\mathcal{P} \subset \Delta(\Omega, \mathcal{F})$ of probability measures. Every element $P \in \mathcal{P}$ induces an extended probability P_0^{ex} as we just described. In this way, we build the set \mathcal{P}_0^{ex} .

Let us now discuss how to update extended probabilities, that is, how to update $P_t^{ex}(A)$ to $P_{t+1}^{ex}(A)$, for all $A \in \mathcal{F}$, for all $P_t^{ex} \in \mathcal{P}_t^{ex}$, for all $t \in \mathbb{N}_0$.^h We first consider a procedure to discover the components of the sample space that is equivalent to an urn without replacement. That is, after specifying Ω and Ω_0^+ , we start our analysis with an urn whose content is unknown and possibly countable; it represents the true sample space associated with our experiment. At any time point t , we extract a ball (an element ω of the sample space). Once we learn about that element, our knowledge about the composition of the urn increases. We do not put the ball back into the urn; we discuss the case in which the discover procedure is equivalent to an urn with replacement later in the chapter.

Let us begin with the updating procedure from $P_0^{ex} \in \mathcal{P}_0^{ex}$ to $P_1^{ex} \in \mathcal{P}_1^{ex}$.

We collect a new observation $\omega \in \Omega$. If $\omega \in \Omega_0^-$, this means that, at time $t = 1$, we learn a new element of the true sample space. Notice that ω is consistent with an element $E_{0,\omega}^-$ of \mathcal{E}_0^- (that is, $E_{0,\omega}^- = \{\omega\}$), which – given that we observe such ω – at time $t = 1$ becomes an element of \mathcal{E}_1^+ ; in formulas, $E_{0,\omega}^- = E_{1,\omega}^+$. Then, $\mathcal{E}_1^+ \supset \mathcal{E}_0^+$, $\mathcal{E}_1^- \subset \mathcal{E}_0^-$, and we update the extended probability assigned to $E_{0,\omega}^- = E_{1,\omega}^+$ as follows

$$P_1^{ex}(E_{1,\omega}^+) = |P_0^{ex}(E_{0,\omega}^-)|. \quad (10)$$

The extended probabilities assigned to the other elements of \mathcal{E} (the finest possible partition of the whole Ω) are held constant. From (5), the updated extended probability associated with event A is given by

$$\begin{aligned} P_1^{ex}(A) = & \sum_{E_{1,j}^+ \in \mathcal{E}_1^+ : P_1^{ex}(E_{1,j}^+) \neq 0} P_1^{ex}(A | E_{1,j}^+) P_1^{ex}(E_{1,j}^+) \\ & + \sum_{E_{1,j}^- \in \mathcal{E}_1^- : P_1^{ex}(E_{1,j}^-) \neq 0} P_1^{ex}(A | E_{1,j}^-) P_1^{ex}(E_{1,j}^-), \end{aligned} \quad (11)$$

where

$$P_1^{ex}(A | E_{1,j}^+) = \frac{P_1^{ex}(A \cap E_{1,j}^+)}{P_1^{ex}(E_{1,j}^+)}$$

^hIt is worth mentioning that in [26] too the author discusses a process of learning the state space combined with learning about the probability of the elements of \mathcal{F} .

22

M. Caprio and S. Mukherjee

and

$$P_1^{ex}(A | E_{1,j}^-) = \frac{P_1^{ex}(A \cap E_{1,j}^-)}{P_1^{ex}(E_{1,j}^-)}.$$

Notice that we are working with the finest possible partition of Ω , that is, $\mathcal{E} = \{\{\omega\}\}_{\omega \in \Omega}$. Then, for any $A \in \mathcal{F}$ and any $E \in \mathcal{E}$, $A \cap E = A \cap \{\omega\}$, which is equal to the empty set if $\omega \notin A$, and it is equal to $\{\omega\} = E$ if $\omega \in A$. Hence, for all $E \in \mathcal{E}$ such that $P_1^{ex}(E) \neq 0$,

$$P_1^{ex}(A | E) = \frac{P_1^{ex}(A \cap E)}{P_1^{ex}(E)} = \begin{cases} \frac{P_1^{ex}(\emptyset)}{P_1^{ex}(E)} = 0 & \text{if } A \cap E = \emptyset \\ \frac{P_1^{ex}(E)}{P_1^{ex}(E)} = 1 & \text{if } A \cap E = \{\omega\} \end{cases},$$

for all $E \in \mathcal{E}$.

So, equation (11) can be rewritten as

$$\begin{aligned} P_1^{ex}(A) &= \sum_{E_{1,j}^+ \in \mathcal{E}_1^+ : A \cap E_{1,j}^+ \neq \emptyset, P_1^{ex}(E_{1,j}^+) \neq 0} P_1^{ex}(E_{1,j}^+) \\ &+ \sum_{E_{1,j}^- \in \mathcal{E}_1^- : A \cap E_{1,j}^- \neq \emptyset, P_1^{ex}(E_{1,j}^-) \neq 0} P_1^{ex}(E_{1,j}^-). \end{aligned} \quad (12)$$

Of course this holds for all $P_1^{ex} \in \mathcal{P}_1^{ex}$. Notice that we are implicitly assuming Allen's principle of conservation of knowledge. Indeed, suppose at time \mathbf{t} the event A is entirely in the latent portion of Ω , that is, $A \cap \Omega_{\mathbf{t}}^- = A$, and at time $\mathbf{t} + k$ it is entirely in the actual portion of Ω , that is, $A \cap \Omega_{\mathbf{t}+k}^+ = A$. Then, by (10) and (12), we have that $P_{\mathbf{t}+k}^{ex}(A) = |P_{\mathbf{t}}^{ex}(A)|$. This has the practical advantage of making the updating procedure mechanical (we do not need to reassess any subjective extended probability when new observations become available), and of preserving the initial opinion of the researcher. The importance of this last point is discussed in Remark 5.

If instead we observe $\omega \in \Omega_0^+$, this means that we made a good job in specifying Ω_0^+ , and so we keep the extended probability constant

$$P_1^{ex}(E_{1,\omega}^+) = P_0^{ex}(E_{0,\omega}^+) \geq 0. \quad (13)$$

Of course, the extended probabilities assigned to the other elements of \mathcal{E} remain constant as well.

We follow the procedures in (10) and (13) to update P_t^{ex} to P_{t+1}^{ex} for all $t \in \mathbb{N}_0$, not just from $t = 0$ to $t = 1$. We call (P_t^{ex}) the sequence of updates of initial extended probability P_0^{ex} . Now consider the extended probability $P_\infty^{ex} \in \Delta^{ex}(\Omega, \mathcal{F})$ such that

$$P_\infty^{ex} \left(A \cap \bigcup_{t \in \mathbb{N}_0} \Omega_t^+ \right) = P \left(A \cap \bigcup_{t \in \mathbb{N}_0} \Omega_t^+ \right)$$

and

$$P_\infty^{ex} \left(A \cap \bigcap_{t \in \mathbb{N}_0} \Omega_t^- \right) = -P \left(A \cap \bigcap_{t \in \mathbb{N}_0} \Omega_t^- \right),$$

for all $A \in \mathcal{F}$, where P is the oracle probability measure we used in (7) and (8) to specify P_0^{ex} . Let us denote by d_{ETV} the extended total variation distance,

$$d_{ETV}(P^{ex}, Q^{ex}) := \sup_{A \in \mathcal{F}} |P^{ex}(A) - Q^{ex}(A)|.$$

It is routine to check that d_{ETV} is a metric: the proof goes along the lines of showing that the total variation distance is a metric. Then, the following holds.

Proposition 8. $P_t^{ex} \rightarrow P_\infty^{ex}$ as $t \rightarrow \infty$ in the extended total variation metric.

The following claim is especially important.

Proposition 9. If $\Omega_t^+ \uparrow \Omega$, then $P_\infty^{ex}(\Omega) = 1$.

This result implies that if $\Omega_t^+ \uparrow \Omega$, then P_∞^{ex} is a regular probability measure. Indeed, it is easy to see that it satisfies the (countably additive) Kolmogorovian axioms for regular probability measures.

Now call $\mathcal{P}_\infty^{ex} \subset \Delta^{ex}(\Omega, \mathcal{F})$ the following set

$$\mathcal{P}_\infty^{ex} := \left\{ P_\infty^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : d_{ETV}(P_t^{ex}, P_\infty^{ex}) \xrightarrow{t \rightarrow \infty} 0, P_t^{ex} \in \mathcal{P}_t^{ex} \right\}.$$

That is, \mathcal{P}_∞^{ex} is the set of limits (in the extended total variation metric) of the sequences (P_t^{ex}) whose elements P_t^{ex} belong to \mathcal{P}_t^{ex} .

The Hausdorff distance between an element of $(\mathcal{P}_t^{ex})_{t \in \mathbb{N}_0}$ and \mathcal{P}_∞^{ex} is given by

$$d_H(\mathcal{P}_t^{ex}, \mathcal{P}_\infty^{ex}) = \max \left\{ \sup_{P_t^{ex} \in \mathcal{P}_t^{ex}} \inf_{P_\infty^{ex} \in \mathcal{P}_\infty^{ex}} d_{ETV}(P_t^{ex}, P_\infty^{ex}), \sup_{P_\infty^{ex} \in \mathcal{P}_\infty^{ex}} \inf_{P_t^{ex} \in \mathcal{P}_t^{ex}} d_{ETV}(P_t^{ex}, P_\infty^{ex}) \right\}. \quad (14)$$

Then, the sequence $(\mathcal{P}_t^{ex})_{t \in \mathbb{N}_0}$ of successive updates of set \mathcal{P}_0^{ex} representing the initial beliefs of the agent facing ambiguity converges in the Hausdorff distance to \mathcal{P}_∞^{ex} . This result is an immediate consequence of Proposition 8: every element of \mathcal{P}_t^{ex} converges to an element of \mathcal{P}_∞^{ex} , so the distance between the ‘‘borders’’ of these two sets – measured by the Hausdorff metric – converges to 0.

Now, there are two possible scenarios: one in which we continue discovering the elements of the state space until we retrieve the full Ω we specified ex ante (this corresponds to $\Omega_t^+ \uparrow \Omega$), and another one in which we discover that the actual sample space associated with our experiment is $\Omega' \subsetneq \Omega$. In the first scenario, any $P_\infty^{ex} \in \mathcal{P}_\infty^{ex}$ is a regular probability measure, a consequence of Proposition 9. So, after discovering the composition of the state space, we have that $(\Omega, \mathcal{F}, P_\infty^{ex})$ is a regular probability space, for all $P_\infty^{ex} \in \mathcal{P}_\infty^{ex}$.

In the second scenario, Ω' can be finite or countable. In the former case, we have that, for some $T \in \mathbb{N}$, $\Omega_T^+ \equiv \Omega' \subsetneq \Omega$, so $\sum_{E_T^+ \in \mathcal{E}_T^+} P_T^{ex}(E_T^+) = q < 1$ because, by (9) and (10),

$$\sum_{E_T^+ \in \mathcal{E}_T^+} P_T^{ex}(E_T^+) + \sum_{E_T^- \in \mathcal{E}_T^-} |P_T^{ex}(E_T^-)| = 1.$$

This may seem problematic: \mathcal{P}_t^{ex} converges to \mathcal{P}_T^{ex} (that is, \mathcal{P}_∞^{ex} coincides with \mathcal{P}_T^{ex}), which is not a set of regular probability measures. To solve this issue, we need to describe the regular probability measure induced by every $P_T^{ex} \in \mathcal{P}_T^{ex}$. It would be desirable to find \tilde{P} such that $\tilde{P}(A) = cP_T^{ex}(A)$, for all $A \in \mathcal{F}_T^+$. This because such a regular probability measure preserves the ratios between extended probabilities of the elements of \mathcal{F}_T^+ , that is,

$$\frac{\tilde{P}(A)}{\tilde{P}(B)} = \frac{cP_T^{ex}(A)}{cP_T^{ex}(B)} = \frac{P_T^{ex}(A)}{P_T^{ex}(B)},$$

for all $A, B \in \mathcal{F}_T^+$ such that $P_T^{ex}(B) \neq 0$.ⁱ To find such a c , the following needs to hold

$$\sum_{E_T^+ \in \mathcal{E}_T^+} cP_T^{ex}(E_T^+) = 1,$$

which happens if and only if

$$c = \frac{1}{\sum_{E_T^+ \in \mathcal{E}_T^+} P_T^{ex}(E_T^+)} = \frac{1}{P_T^{ex}(\Omega_T^+)}.$$

Hence, $\tilde{P} = cP_T^{ex}$ is the regular probability measure induced by P_T^{ex} that preserves the ratios between extended probabilities of elements of \mathcal{F}_T^+ . Clearly, this holds for all $P_T^{ex} \in \mathcal{P}_T^{ex}$, so $(\Omega_T^+, \mathcal{F}_T^+, \tilde{P})$ is a regular probability space, for all $\tilde{P} \in \tilde{\mathcal{P}}$, where $\tilde{\mathcal{P}}$ is the set of regular probabilities induced by the elements of \mathcal{P}_T^{ex} .

ⁱNotice that if the ratio is greater than 1, we deem A more likely than B , and vice versa if the ratio is smaller than 1. We deem A and B equally likely if the ratio is exactly 1. Because we want to preserve this behavioral interpretation when we move to the regular probabilities induced by the elements of \mathcal{P}_T^{ex} , we want the ratio to be maintained.

If Ω' is countable, we proceed in a similar way. We still want to preserve the ratios between extended probabilities of the elements of \mathcal{F}' , so we have to find c such that $\sum_{E \in \mathcal{E}'} cP_{\infty}^{ex}(E) = 1$. This happens when

$$c = \frac{1}{\sum_{E \in \mathcal{E}'} P_{\infty}^{ex}(E)} = \frac{1}{P_{\infty}^{ex}(\Omega')},$$

so $\tilde{P} = cP_{\infty}^{ex}$ is the regular probability measure we were looking for. This holds for all $P_{\infty}^{ex} \in \mathcal{P}_{\infty}^{ex}$, so $(\Omega', \mathcal{F}', \tilde{P})$ is a regular probability space, for all $\tilde{P} \in \tilde{\mathcal{P}}$.

Remark 5. Notice that, in the first scenario ($\Omega_t^+ \uparrow \Omega$), every element $P_{\infty}^{ex} \in \mathcal{P}_{\infty}^{ex}$ coincides with its corresponding oracle probability measure $P \in \mathcal{P}$ we expressed ex ante on the whole Ω . Indeed, we have that $P_{\infty}^{ex}(E) = |P_{\infty}^{ex}(E)|$, for all $E \in \mathcal{E}$, which implies $P_{\infty}^{ex}(A) = |P_{\infty}^{ex}(A)|$, for all $A \in \mathcal{F}$. But then, by (9), we know that $\sum_{E \in \mathcal{E}} P_{\infty}^{ex}(E) = 1$, and that $P_{\infty}^{ex}(A) = P(A)$, for all $A \in \mathcal{F}$. This should not surprise: the updating procedure we described earlier is not based on collecting new data like the Bayesian one, nor on repeating the experiment many times like the frequentist one. Rather, it is based on discovering the true composition of the state space associated with our inference procedure. Then, it is natural to retrieve the opinion we expressed ex ante on the state space Ω once we get the confirmation that the true state space is indeed Ω . This also reconciles well with the interpretation we gave to negative extended probabilities: once we are given the possibility to enter the bets we were denied before, we tend to agree with our doppelgänger who had the opportunity to bet on those events in the first place (Allen's principle of conservation of knowledge holds).

Notice also that \tilde{P} , albeit not equal, is proportional to P_{∞}^{ex} , $\tilde{P} = cP_{\infty}^{ex}$, for every $\tilde{P} \in \tilde{\mathcal{P}}$ and its corresponding $P_{\infty}^{ex} \in \mathcal{P}_{\infty}^{ex}$. The interpretation is immediate: we maintain our opinion on the elements of the sample space $\Omega' \subsetneq \Omega$ that pertains to our experiment.

If the procedure for discovering the composition of the state space is equivalent to an urn with replacement, everything we discussed so far still holds with just two differences:

- if we extract twice or more times the same element, its extended probability flips sign the first time, and then stays constant;
- the convergence may be slower, because we may need more extractions from the urn to learn the true composition (because we can extract twice or more times the same element).

Let us give a simple example that illustrates one of the possible situations in which the analysis depicted so far can be put to use.

Example 1. We consider a species sampling problem in the field of ecology. Suppose we want to know the number of bird species that inhabit a certain region throughout the year. What we can do is to start our analysis by letting $\Omega = \mathbb{N}$ and $\Omega_0^+ = \{1, \dots, n\}$, where n is the number of species that inhabit a region similar to the one of interest throughout the year. Then, we specify a set \mathcal{P} of probability measures on \mathbb{N} , e.g. a collection $\{\text{Geom}(p)\}_{p \in [0,1]}$ of geometric distributions having parameter $p \in [0, 1]$. We specify a set of probabilities because we are not able to express our initial opinion via a unique probability measure (we face ambiguity). We specify the probability measures on the whole number field \mathbb{N} because we do not know exactly the composition of our state space. Then, after eliciting the set \mathcal{P}_0^{ex} of extended probability measures induced by (the elements of) \mathcal{P} , we begin the ex ante analysis described in the present work. After collecting observations for an entire year, we end up discovering that the state space associated with our experiment is $\Omega' = \{1, \dots, m\} \subsetneq \mathbb{N}$, where $m \geq n$ is the number of species we counted during the year. We recover also the set $\tilde{\mathcal{P}}$ of regular probability measures induced by \mathcal{P}_∞^{ex} . Now, the “real” statistical analysis can take place. Indeed, we know that m is the maximum number of species that live in the region during the year, but it does not take into account migrations to or from the region itself. Hence, the number of bird species that inhabit the region throughout the year may well be smaller than m . So, every $\tilde{P} \in \tilde{\mathcal{P}}$ will be such that $\tilde{P}(\{\omega_k\}) \in [0, 1]$, and $\sum_{k=1}^m \tilde{P}(\{\omega_k\}) = 1$, where $\omega_k = k$, for all $k \in \{1, \dots, m\}$. Gathering data y_1, \dots, y_ℓ and updating these \tilde{P} 's via Bayesian conditioning, we obtain the set $\{\tilde{P}(\cdot | y_1, \dots, y_\ell)\}_{\tilde{P} \in \tilde{\mathcal{P}}}$ of posterior (regular) probability measures. This gives a robust analysis: for all $\omega_k \in \Omega'$, the posterior probability of ω_k being the correct number of species belongs to the interval

$$\left[\underline{\tilde{P}}(\{\omega_k\} | y_1, \dots, y_\ell), \overline{\tilde{P}}(\{\omega_k\} | y_1, \dots, y_\ell) \right],$$

where the lower bound is a lower (regular) probability, and the upper bound is an upper (regular) probability. The narrower the interval, the less imprecise our beliefs resulting from the analysis. \triangle

A criticism that can be made of our ex ante analysis is the following: why should the scholar be concerned with the exact composition of the state space? It should be enough to specify it as richly as possible, and then proceed to a regular statistical analysis. There are two responses to

such a critique. First, as pointed out in Remark 4, in doing so the scholar would not be respecting the condition listed in [41] that the state space must have the right granularity depending on the statistical experiment they are interested in. Second, in the case that the state space associated with our statistical experiment is $\Omega' \subsetneq \Omega$, working with probability measures supported on the whole space Ω of apparently possible states may be computationally costly. Our ex ante analysis allows the scholar to focus on the “minimal” state space – the one containing only the necessary states – that is more meaningful to the analysis.

4. Upper and lower extended probabilities

Fix any $t \in \mathbb{N}_0$, and consider the “boundary elements” of the set \mathcal{P}_t^{ex} ,

$$\underline{P}_t^{ex}(A) = \inf_{P^{ex} \in \mathcal{P}_t^{ex}} P^{ex}(A) \quad (15)$$

and

$$\overline{P}_t^{ex}(A) = \sup_{P^{ex} \in \mathcal{P}_t^{ex}} P^{ex}(\Omega) - \underline{P}_t^{ex}(A^c) = \sup_{P^{ex} \in \mathcal{P}_t^{ex}} P^{ex}(A), \quad (16)$$

for all $A \in \mathcal{F}$. They are not extended probabilities; we call $\underline{P}_t^{ex}(A)$ a lower extended probability measure, and $\overline{P}_t^{ex}(A)$ an upper extended probability measure. Notice that upper extended probability measures differ from upper regular probability measures. These latter are defined as 1 minus the lower regular probability of the complement of the event we are interested in. In (16) we give a similar conjugate type of definition, but we cannot write 1, because we do not require that the lower extended probability of Ω is 1.

Given a generic lower extended probability \underline{P}^{ex} and a generic event $A \in \mathcal{F}$, we interpret $\underline{P}^{ex}(A)$ as follows. If $\underline{P}^{ex}(A) = p > 0$, then p represents the supremum price that we are willing to pay to enter a bet on A that gives us \$1 if A takes place and \$0 otherwise. If $\underline{P}^{ex}(A) = 0$, then in the most conservative of our mental states we deem A impossible to take place. Finally, if $\underline{P}^{ex}(A) = q < 0$, then $|q|$ represents the infimum selling price at which our doppelgänger takes bets on A that pay \$1 if A takes place and \$0 otherwise. As we can see, this interpretation captures the ideas of worst case scenario and of prudent behavior. It can be seen as a betting scheme analogous to the one in [24, section 2.3.1], but with monetary instead of utiles outcomes, and extended to probabilities that can take on negative values as well.

Both lower and upper extended probabilities are extended Choquet capacities. A generic extended Choquet capacity is defined as a set function $\nu^{ex} : \mathcal{F} \rightarrow \mathbb{R}$ such that

- (EC1) $\nu^{ex}(\emptyset) = 0$,
- (EC2) $\nu^{ex}(A) \in [-1, 1]$, for all $A \in \mathcal{F}$,
- (EC3) for any $A, B \in \mathcal{F}$ such that $A \subset B$, if $\nu^{ex}(A) \geq 0$, $\nu^{ex}(B) \geq 0$, and $\nu^{ex}(B \cap A^c) \geq 0$, then $\nu^{ex}(A) \leq \nu^{ex}(B)$. If instead $\nu^{ex}(A) \leq 0$, $\nu^{ex}(B) \leq 0$, and $\nu^{ex}(B \cap A^c) \leq 0$, then $\nu^{ex}(A) \geq \nu^{ex}(B)$.

As we can see, we do not require countable additivity (ii*) to hold for extended capacities. An extended probability measure is an additive extended capacity.

It is easy to see that upper and lower extended probability measures satisfy (EC1)-(EC3); in addition, lower extended probabilities are superadditive, while upper extended probabilities are subadditive. That is, for all $A, B \in \mathcal{F}$,

$$\underline{P}_t^{ex}(A \sqcup B) \geq \underline{P}_t^{ex}(A) + \underline{P}_t^{ex}(B) \quad (17)$$

and

$$\overline{P}_t^{ex}(A \sqcup B) \leq \overline{P}_t^{ex}(A) + \overline{P}_t^{ex}(B). \quad (18)$$

These inequalities come immediately from the properties of the infimum and supremum operators.

Remark 6. Notice that the behavioral interpretation that we gave to lower extended probabilities entails that they can be specified even without eliciting a set of extended probability measures first. To this extent, our behavioral interpretation can be called minimal, similarly to [24, section 2.3.1]. An immediate question the reader may ask is: “If we were to specify a lower extended probability without resorting to a set of extended probabilities, are we sure it is subadditive?”. The answer is yes, under a mild assumption: Example 2 – based on the example in [24, section 1.6.4] – shows that if lower extended probabilities avoid sure loss, then they are superadditive, and Theorem 2 shows that if \underline{P}^{ex} can be obtained as the infimum of a set of extended probabilities, then it avoids sure loss.

We first define *sure loss* for lower extended probabilities. It is the immediate lower counterpart of Definition 1.

Definition 2. Lower extended probability \underline{P}^{ex} avoids sure loss if no Dutch books can be made against the punter, that is, if we cannot find a finite

collection $\{B_j\}_{j=1}^n \subset \mathcal{B}'$ along with numbers $\{s_j\}_{j=1}^n \subset \mathbb{R}_+$ such that

$$\sup_{\omega \in \Omega} \left\{ \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - \underline{P}^{ex}(B_j)] \right\} < 0, \quad (19)$$

where \mathcal{B}' is the sigma-algebra generated by the collection $\{B \subset \Omega : \underline{P}^{ex}(B) \geq 0 \text{ and } B \cap C = \emptyset, \forall C \in \mathcal{F}, \underline{P}^{ex}(C) < 0\}$.

We have the following interesting result, that is a version of the lower envelope theorem in [24, Corollary 2.8.6].^j

Theorem 2. *Consider a generic lower extended probability \underline{P}^{ex} defined on a measurable space (Ω, \mathcal{F}) . If there exists a nonempty set \mathcal{P}^{ex} of extended probabilities on (Ω, \mathcal{F}) such that $\underline{P}^{ex}(A) = \inf_{P^{ex} \in \mathcal{P}^{ex}} P^{ex}(A)$, for all $A \in \mathcal{F}$, then \underline{P}^{ex} avoids sure loss.*

The following example shows that if \underline{P}^{ex} avoids sure loss, then it is super-additive.

Example 2. Pick two mutually exclusive events A and B . By our behavioral interpretation, the highest amount we are willing to pay to get \$1 if A occurs (or the highest amount we would deem reasonable to lose if we were given the possibility to enter the bet) is $\underline{P}^{ex}(A)$. The same holds for $\underline{P}^{ex}(B)$. Because we want to avoid sure loss, the net outcome is equivalent to paying $\underline{P}^{ex}(A) + \underline{P}^{ex}(B)$ to get \$1 if $A \sqcup B$ occurs. Now, $\underline{P}^{ex}(A \sqcup B)$ is the highest price we are willing to pay to obtain \$1 if $A \sqcup B$ occurs. So we recover the superadditivity constraint

$$\underline{P}^{ex}(A \sqcup B) \geq \underline{P}^{ex}(A) + \underline{P}^{ex}(B).$$

Similarly, an upper lower probability \overline{P}^{ex} should satisfy the subadditivity constraint $\overline{P}^{ex}(A \sqcup B) \leq \overline{P}^{ex}(A) + \overline{P}^{ex}(B)$. \triangle

An important concept worth introducing is the core of a lower extended probability measure.

Definition 3. Given a generic lower extended probability \underline{P}^{ex} , we call core of \underline{P}^{ex} the set

$$\text{core}(\underline{P}^{ex}) := \{P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : P^{ex}(A) \geq \underline{P}^{ex}(A), \forall A \in \mathcal{F} \\ \text{and } P^{ex}(\Omega) = \underline{P}^{ex}(\Omega)\}.$$

^jNotice that the lower envelope theorem in [24, Corollary 2.8.6] comprises an envelope of probabilities that are merely finitely additive, and it states a necessary and sufficient condition. We plan to prove the opposite direction of Theorem 2 in future work.

This definition tells us that the core of \underline{P}^{ex} is the set of all suitably normalized extended probabilities that setwise dominate \underline{P}^{ex} . Notice that in general it may be empty, and that

$$\begin{aligned} \text{core}(\underline{P}^{ex}) &= \{P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : \underline{P}^{ex} \leq P^{ex} \leq \overline{P}^{ex}\} \\ &= \{P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : P^{ex}(A) \leq \overline{P}^{ex}(A), \forall A \in \mathcal{F} \\ &\quad \text{and } P^{ex}(\Omega) = \underline{P}^{ex}(\Omega)\}, \end{aligned}$$

so the core can be seen as the set of extended probabilities “sandwiched” between lower extended probability \underline{P}^{ex} and upper extended probability \overline{P}^{ex} , as well as the set of extended probabilities setwise dominated by \overline{P}^{ex} . The following is a corollary to Theorem 2.

Corollary 2. *If $\text{core}(\underline{P}^{ex}) \neq \emptyset$, then \underline{P}^{ex} avoids sure loss.*

A crucial property of the core is the following.

Proposition 10. *Given a generic lower extended probability \underline{P}^{ex} , its core is convex and weak* -compact.*

Being compact and convex, the core of \underline{P}^{ex} is completely characterized by \underline{P}^{ex} . This means that it is enough to know \underline{P}^{ex} to be able to retrieve every element in its core. So in our analysis we can focus on updating \underline{P}_t^{ex} to \underline{P}_{t+1}^{ex} , and then require that $\mathcal{P}_{t+1}^{ex} = \text{core}(\underline{P}_{t+1}^{ex})$, instead of updating \mathcal{P}_t^{ex} to \mathcal{P}_{t+1}^{ex} elementwise. This justifies our focus in the remainder of this section on studying how to update lower extended probabilities. The procedure to update upper extended probability measures is going to be similar (their relation is described by equation (16)).

Recall that the conditions to perform the update in the additive case are given by (10) and (13).

Proposition 11. *The sublinear counterpart of (10) is the following,*

$$\underline{P}_{t+1}^{ex}(E_{t+1,\omega}^+) = \left| \overline{P}_t^{ex}(E_{t,\omega}^-) \right| \quad \text{and} \quad \overline{P}_{t+1}^{ex}(E_{t+1,\omega}^+) = \left| \underline{P}_t^{ex}(E_{t,\omega}^-) \right|. \quad (20)$$

It holds when we learn a new element ω of our sample space, that is, when the new observation ω belongs to the latent space at time t , but “moves” to the actual space at time $t + 1$. In formulas, $\omega \in \Omega_t^-$, but $\omega \in \Omega_{t+1}^+$. The lower extended probabilities assigned to the other elements of \mathcal{E} are held constant.

Notice that we are working with the finest possible partition of Ω , so, as before, the intersection $A \cap E$ between any $A \in \mathcal{F}$ and any $E = \{\omega\} \in \mathcal{E}$

is either the empty set, $A \cap E = \emptyset$, if $\omega \notin A$, or the element ω itself, $A \cap E = \{\omega\} = E$, if $\omega \in A$. Hence, for any $t \in \mathbb{N}_0$, for any $E = \{\omega\} \in \mathcal{E}$, and for any $A \in \mathcal{F}$, we have that

$$\underline{P}_t^{ex}(A \cap E) = \begin{cases} \underline{P}_t^{ex}(E) & \text{if } \omega \in A \\ \underline{P}_t^{ex}(\emptyset) = 0 & \text{if } \omega \notin A \end{cases}. \quad (21)$$

Given any $A, B \in \mathcal{F}$, if $\underline{P}_t^{ex}(B) \neq 0$, we define

$$\underline{P}_t^{ex}(A | B) := \frac{\inf_{P^{ex} \in \mathcal{P}_t^{ex}} P^{ex}(A \cap B)}{\inf_{P^{ex} \in \mathcal{P}_t^{ex}} P^{ex}(B)} = \frac{\underline{P}_t^{ex}(A \cap B)}{\underline{P}_t^{ex}(B)}, \quad (22)$$

for all $t \in \mathbb{N}_0$. We call it the *extended Geometric rule*. We notice immediately that, combining (21) and (22), we get

$$\underline{P}_t^{ex}(A | E) = \frac{\underline{P}_t^{ex}(A \cap E)}{\underline{P}_t^{ex}(E)} = \begin{cases} \frac{\underline{P}_t^{ex}(E)}{\underline{P}_t^{ex}(E)} = 1 & \text{if } \omega \in A \\ \frac{\underline{P}_t^{ex}(\emptyset)}{\underline{P}_t^{ex}(E)} = 0 & \text{if } \omega \notin A \end{cases}, \quad (23)$$

for all $E = \{\omega\} \in \mathcal{E}$ such that $\underline{P}_t^{ex}(E) \neq 0$ and all $A \in \mathcal{F}$. So, we have that $\underline{P}_t^{ex}(A | E) = P_t^{ex}(A | E)$, for all A , all E , and all $P_t^{ex} \in \mathcal{P}_t^{ex}$. This is true only because we are working with the finest possible partition of Ω , and because we endorse the extended Geometric rule.

The sublinear counterpart of (13) is the following

$$\underline{P}_{t+1}^{ex}(E_{t+1,\omega}^+) = \underline{P}_t^{ex}(E_{t,\omega}^+) \geq 0, \quad (24)$$

for all $t \in \mathbb{N}_0$. This comes from the fact that we draw an element already belonging to the actual space, so we do not need to update its lower extended probability (similarly to what is described in equation (13)). The lower extended probabilities assigned to the other elements of \mathcal{E} are held constant.

At this point, a natural question one may ask is how to compute $\underline{P}_t^{ex}(A)$, for any t , for any $A \in \mathcal{F}$. It would be tempting to write that

$$\begin{aligned} \underline{P}_t^{ex}(A) &= \sum_{E_{t,j}^+ \in \mathcal{E}_t^+ : \underline{P}_t^{ex}(E_{t,j}^+) \neq 0} \underline{P}_t^{ex}(A | E_{t,j}^+) \underline{P}_t^{ex}(E_{t,j}^+) \\ &+ \sum_{E_{t,j}^- \in \mathcal{E}_t^- : \underline{P}_t^{ex}(E_{t,j}^-) \neq 0} \underline{P}_t^{ex}(A | E_{t,j}^-) \underline{P}_t^{ex}(E_{t,j}^-). \end{aligned}$$

This would mimic exactly (11), with lower extended probabilities in place of additive extended probabilities. Alas, that would not be true, since lower extended probabilities are not additive. Instead, we have the following.

Fix any $t \in \mathbb{N}_0$. Call $\mathfrak{E}_t \equiv \{E_{t,A}\}$ the collection of elements of \mathcal{E} such that $A \cap E_{t,A} \neq \emptyset$. We index \mathfrak{E}_t to time t to highlight the fact that although

its elements stay the same, some of them may “move” from the latent to the actual space as time goes by and we collect more observations. Since we are working with the finest possible partition of Ω , we can write

$$A = \bigsqcup_{E_{t,A} \in \mathfrak{E}_t} E_{t,A}. \quad (25)$$

In the most general case, some of these $E_{t,A}$'s belong to the actual space, and some to the latent space. Let us denote the former by $E_{t,A}^+$'s and the latter by $E_{t,A}^-$'s. Formally, we have that $E_{t,A}^+ \cap \Omega_t^+ \neq \emptyset$ and $E_{t,A}^+ \cap \Omega_t^- = \emptyset$, and vice versa for the $E_{t,A}^-$'s. Hence, (25) can be rewritten as

$$A = \bigsqcup_{E_{t,A}^+ \in \mathfrak{E}_t} E_{t,A}^+ \sqcup \bigsqcup_{E_{t,A}^- \in \mathfrak{E}_t} E_{t,A}^-. \quad (26)$$

Now, from (20) and (24), we know how to update the lower extended probabilities of all the elements of the partition \mathfrak{E} of Ω , which implies that we know the value of $\underline{P}_{t+1}^{ex}(E_{t+1,A})$, for all $E_{t+1,A} \in \mathfrak{E}_{t+1}$. Then, consider now the summation

$$\sum_{E_{t+1,A} \in \mathfrak{E}_{t+1}} \underline{P}_{t+1}^{ex}(E_{t+1,A}).$$

From equation (17), we have that

$$\begin{aligned} \underline{P}_{t+1}^{ex}(A) &\geq \sum_{E_{t+1,A} \in \mathfrak{E}_{t+1}} \underline{P}_{t+1}^{ex}(E_{t+1,A}) \\ &= \sum_{E_{t+1,A}^+ \in \mathfrak{E}_{t+1}} \underline{P}_{t+1}^{ex}(E_{t+1,A}^+) + \sum_{E_{t+1,A}^- \in \mathfrak{E}_{t+1}} \underline{P}_{t+1}^{ex}(E_{t+1,A}^-). \end{aligned}$$

From equation (18), we can give an upper bound for the upper extended probability of A ,

$$\begin{aligned} \overline{P}_{t+1}^{ex}(A) &\leq \sum_{E_{t+1,A} \in \mathfrak{E}_{t+1}} \overline{P}_{t+1}^{ex}(E_{t+1,A}) \\ &= \sum_{E_{t+1,A}^+ \in \mathfrak{E}_{t+1}} \overline{P}_{t+1}^{ex}(E_{t+1,A}^+) + \sum_{E_{t+1,A}^- \in \mathfrak{E}_{t+1}} \overline{P}_{t+1}^{ex}(E_{t+1,A}^-). \end{aligned}$$

5. Conclusion

In this work we give a definition of extended probability measures that does not depend on the field a scholar works in. We give some of their more interesting properties, and a behavioral interpretation to positive and negative values of extended probabilities. We then apply extended probabilities to

statistical inference. Given the probability space (Ω, \mathcal{F}, P) associated with the experiment we want to conduct, we use extended probabilities to express uncertainty about both the composition of the state space Ω , and which probability measure P to select. We develop an ex-ante analysis; our method describes how the researcher progressively discovers the true composition of the state space, so that, at the end of the process, a regular statistical analysis (that requires the knowledge of Ω) can take place. We introduce the concept of extended Choquet capacities, and in particular of upper and lower extended probabilities that represent the “borders” of sets of extended probabilities, and we give bounds for the lower extended probability of any element $A \in \mathcal{F}$. We also apply our model to the fields of opinion dynamics and species sampling models. This chapter is important because it gives a foundational definition of extended probability measures and makes these latter relevant to the field of statistics.

In the future, we will deal with the possibility that in our ex-ante analysis we are not able to discover the whole composition of the state space associated with our experiment. In that case, the statistical analysis itself will have to be carried out using extended probabilities. We also plan to relax the assumption we made about working with a finite or countable state space. We will also provide direct behavioral motivation for the properties of extended probabilities so as to be able to relate our approach to those of de Finetti [23, 27] and Walley [24]. Furthermore, since it is one of the main benefits of a betting approach, we will study partially specified extended probabilities (that is, extended probabilities defined not on a sigma-algebra, but rather on a generic collection of events) and their coherent extension. Finally, we would like to deepen the study of lower extended probabilities.

Acknowledgments

This work is supported in part by the following grants: NSF CCF-193496, NSF DEB-1840223, NIH R01 DK116187-01, HFSP RGP0051/2017, NIH R21 AG055777-01A, NSF DMS 17-13012, NSF ABI 16-61386, ARO MURI W911NF2010080. The authors would like to thank Aaditya Ramdas, Mark Burgin, Shounak Chattopadhyay, Roberto Corrao, Ruobin Gong, Vittorio Orlandi, Glenn Shafer, Peter Walley, Marco Zaffalon and Alessandro Zito for helpful comments.

A.1. Proofs

Proof of Proposition 1. Suppose for the sake of contradiction that for some $A = B$,

$$P^{ex}(A) \neq P^{ex}(B). \quad (\text{A.1})$$

Consider then $A \setminus B = B \setminus B = \emptyset$; we have that

$$P^{ex}(A \setminus B) \neq P^{ex}(B \setminus B) = P^{ex}(\emptyset) = 0. \quad (\text{A.2})$$

This is true because $A = \{A \setminus B\} \sqcup \{A \cap B\}$ and $B = \{B \setminus B\} \sqcup \{B \cap B\}$; by (ii*), this implies that $P^{ex}(A \setminus B) = P^{ex}(A) - P^{ex}(A \cap B)$ and $P^{ex}(B \setminus B) = P^{ex}(B) - P^{ex}(B \cap B)$. Then, we have that $P^{ex}(A \setminus B) = P^{ex}(B \setminus B)$ if and only if

$$P^{ex}(A) = P^{ex}(B) - P^{ex}(B \cap B) + P^{ex}(A \cap B);$$

but $B \cap B = B$, and $A \cap B = B$ because we assumed $A = B$, so $P^{ex}(A \setminus B) = P^{ex}(B \setminus B)$ if and only if $P^{ex}(A) = P^{ex}(B)$. But by (A.1) $P^{ex}(A) \neq P^{ex}(B)$, so we have that the inequality in (A.2) holds.

By assumption we know that $A = B$, so $A \setminus B = \emptyset$, and so $P^{ex}(A \setminus B) = P^{ex}(\emptyset)$. Then, by (A.2), we have that $P^{ex}(\emptyset) = P^{ex}(A \setminus B) \neq P^{ex}(B \setminus B) = P^{ex}(\emptyset)$, a contradiction. \square

Proof of Proposition 2. Let $A \subset B$; then, we can write B as $B = A \sqcup (B \cap A^c)$. By (ii*) this implies that $P^{ex}(B) = P^{ex}(A) + P^{ex}(B \cap A^c)$. Then, if $P^{ex}(B)$, $P^{ex}(A)$, and $P^{ex}(B \cap A^c)$ are all nonnegative, then $P^{ex}(B) = P^{ex}(A) + P^{ex}(B \cap A^c) \geq P^{ex}(A)$. If instead they are all nonpositive, then $P^{ex}(B) = P^{ex}(A) + P^{ex}(B \cap A^c) \leq P^{ex}(A)$. \square

Proof of Proposition 3. Pick $A, B \in \mathcal{F}$; if they are disjoint, the equation follows immediately from (ii*). If they are not, consider $A \cup B = \{A \setminus B\} \sqcup \{A \cap B\} \sqcup \{B \setminus A\}$. Then, again by (ii*), we have that

$$P^{ex}(A \cup B) = P^{ex}(A \setminus B) + P^{ex}(A \cap B) + P^{ex}(B \setminus A) \quad (\text{A.3})$$

Notice then that $P^{ex}(A \setminus B) = P^{ex}(A) - P^{ex}(A \cap B)$; indeed

$$P^{ex}(A) = P^{ex}(\{A \cap B\} \sqcup \{A \setminus B\}) = P^{ex}(A \cap B) + P^{ex}(A \setminus B).$$

Similarly, $P^{ex}(B \setminus A) = P^{ex}(B) - P^{ex}(B \cap A)$. So, (A.3) becomes

$$\begin{aligned} P^{ex}(A \cup B) &= P^{ex}(A) - P^{ex}(A \cap B) + P^{ex}(A \cap B) + P^{ex}(B) - P^{ex}(B \cap A) \\ &= P^{ex}(A) + P^{ex}(B) - P^{ex}(A \cap B). \end{aligned}$$

\square

Proof of Theorem 1. Notice that P^{ex} restricted to \mathcal{B} is a finite signed measure. By the Hahn-Jordan decomposition theorem [45], there exists a unique decomposition of P^{ex} into a difference $P^{ex} = P_+^{ex} - P_-^{ex}$ of two finite positive measures. Pick any $\{B_j\}_{j=1}^n \subset \mathcal{B}$, $\{s_j\}_{j=1}^n \subset \mathbb{R}$. We have that, for any payoff

$$\begin{aligned} f(\omega) &= \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - P^{ex}(B_j)] \\ &= \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - P_+^{ex}(B_j) + P_-^{ex}(B_j)], \end{aligned}$$

the following holds

$$\begin{aligned} \int_{\Omega} f dP^{ex} &= \sum_{j=1}^n s_j \left[\int_{\Omega} \mathbb{1}_{B_j} dP^{ex} - P_+^{ex}(B_j) + P_-^{ex}(B_j) \right] \\ &= \sum_{j=1}^n s_j \left[\int_{\Omega} \mathbb{1}_{B_j} dP_+^{ex} - \int_{\Omega} \mathbb{1}_{B_j} dP_-^{ex} - P_+^{ex}(B_j) + P_-^{ex}(B_j) \right] \\ &= \sum_{j=1}^n s_j \left[\int_{\Omega} \mathbb{1}_{B_j} dP_+^{ex} - P_+^{ex}(B_j) \right] - \sum_{j=1}^n s_j \left[\int_{\Omega} \mathbb{1}_{B_j} dP_-^{ex} - P_-^{ex}(B_j) \right] \\ &= 0. \end{aligned}$$

So f cannot have a negative supremum. \square

Proof of Proposition 4. We have that

$$\begin{aligned} P_t^{ex}(A) &= P_t^{ex}(A \cap \Omega_t^+) + P_t^{ex}(A \cap \Omega_t^-) \\ &= \sum_{E_{t,j}^+ \in \mathcal{E}_t^+} P_t^{ex}(A \cap E_{t,j}^+) + \sum_{E_{t,j}^- \in \mathcal{E}_t^-} P_t^{ex}(A \cap E_{t,j}^-) \\ &= \sum_{E_{t,j}^+ \in \mathcal{E}_t^+ : P_t^{ex}(E_{t,j}^+) \neq 0} P_t^{ex}(A | E_{t,j}^+) P_t^{ex}(E_{t,j}^+) \quad (\text{A.4}) \\ &\quad + \sum_{E_{t,j}^- \in \mathcal{E}_t^- : P_t^{ex}(E_{t,j}^-) \neq 0} P_t^{ex}(A | E_{t,j}^-) P_t^{ex}(E_{t,j}^-), \end{aligned}$$

where the first equality comes from $A = (A \cap \Omega_t^+) \sqcup (A \cap \Omega_t^-)$ and (ii*), the third equality comes from (2), and the second equality comes from $\Omega_t^+ = \sqcup_{E_{t,j}^+ \in \mathcal{E}_t^+} E_{t,j}^+$, so

$$A \cap \Omega_t^+ = A \cap (\sqcup_{E_{t,j}^+ \in \mathcal{E}_t^+} E_{t,j}^+) = \sqcup_{E_{t,j}^+ \in \mathcal{E}_t^+} (A \cap E_{t,j}^+),$$

where the last equality comes from De Morgan laws. \square

Proof of Proposition 5. Pick any $A, B \in \mathcal{F}$. Since $\Omega_t^+ \cap \Omega_t^- = \emptyset$, we have that

$$\begin{aligned} A \cup B &= ((A \cup B) \cap \Omega_t^+) \sqcup ((A \cup B) \cap \Omega_t^-) \\ &= ((A \cap \Omega_t^+) \cup (B \cap \Omega_t^+)) \sqcup ((A \cap \Omega_t^-) \cup (B \cap \Omega_t^-)) \\ &= (A_t^+ \cup B_t^+) \sqcup (A_t^- \cup B_t^-). \end{aligned}$$

A similar argument shows that $A \cap B = (A_t^+ \cap B_t^+) \cup (A_t^- \cap B_t^-)$, for all t .

For the second part, notice that $A \setminus B = \{\omega \in \Omega : \omega \in A, \omega \notin B\}$. Then, $A \setminus B \cap \Omega_t^+ \equiv A_t^+ \setminus B_t^+ = \{\omega \in \Omega_t^+ : \omega \in A, \omega \notin B\}$; similarly, $A \setminus B \cap \Omega_t^- \equiv A_t^- \setminus B_t^- = \{\omega \in \Omega_t^- : \omega \in A, \omega \notin B\}$. But $\Omega_t^+ \sqcup \Omega_t^- = \Omega$, so the claim follows. \square

Proof of Proposition 6. Let us focus on \mathcal{F}_t^+ ; it is closed with respect to countable intersections because it is a sigma-algebra. Then, pick $A, B \in \mathcal{F}_t^+$ such that $A \neq B$. If $A \cap B = \emptyset$, then $A \setminus B = A \in \mathcal{F}_t^+$; if $A \cap B \neq \emptyset$, then $A \setminus B = A \cap B^c$. Then, $B \in \mathcal{F}_t^+$ implies $B^c \in \mathcal{F}_t^+$ because \mathcal{F}_t^+ is a sigma-algebra; also, $A \cap B^c \in \mathcal{F}_t^+$ because \mathcal{F}_t^+ is closed with respect to countable intersections. So $A \setminus B \in \mathcal{F}_t^+$. Finally, the unit element is Ω_t^+ : for any $A \in \mathcal{F}_t^+$, $A \cap \Omega_t^+ = A$, because $A \in \mathcal{F}_t^+$. Hence \mathcal{F}_t^+ is a set algebra. We show in a similar fashion that \mathcal{F}_t^- is a set algebra as well. \square

Proof of Proposition 7. To ease notation, let $C \equiv \cup_{j \in \mathbb{N}_0} A_j$. C belongs to \mathcal{F} , because \mathcal{F} is a sigma-algebra. Then, let $C \in \mathcal{F}_t^+$. This implies that $P_t^{ex}(C) \geq 0$, but also that $P^{ex}(A) \geq 0$ and that $P^{ex}(C \cap A^c) \geq 0$, since $C = A \sqcup (C \cap A^c)$. Then,

$$P_t^{ex}(A) \leq P_t^{ex}(C) \leq \sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j),$$

where the first inequality comes from Proposition 2, and the second one from Proposition 3.

If $C \in \mathcal{F}_t^-$, then $P_t^{ex}(C) \leq 0$, and also that $P^{ex}(A) \leq 0$ and that $P^{ex}(C \cap A^c) \leq 0$. Then,

$$P_t^{ex}(A) \geq P_t^{ex}(C) \geq \sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j),$$

where again the first inequality comes from Proposition 2, and the second one from Proposition 3.

If $C \equiv \cup_{j \in \mathbb{N}_0} A_j$ is such that $C \cap \Omega_t^+ \neq \emptyset \neq C \cap \Omega_t^-$, then we cannot say anything general about the relation between $P_t^{ex}(A)$ and $\sum_{j \in \mathbb{N}_0} P_t^{ex}(A_j)$. \square

Proof of Proposition 8. Notice that by (7) and(8), we have that, for all t ,

$$P_t^{ex}(A \cap \Omega_t^+) = P(A \cap \Omega_t^+)$$

and

$$P_t^{ex}(A \cap \Omega_t^-) = -P(A \cap \Omega_t^-),$$

for all $A \in \mathcal{F}$, where $P \in \Delta(\Omega, \mathcal{F})$. Now, to save some notation, let us denote by $\Omega^+ \equiv \cup_{t \in \mathbb{N}_0} \Omega_t^+$ and by $\Omega^- \equiv \cap_{t \in \mathbb{N}_0} \Omega_t^-$. Then, we have the following.

$$d_{ETV}(P_t^{ex}, P_\infty^{ex}) = \sup_{A \in \mathcal{F}} |P_t^{ex}(A) - P_\infty^{ex}(A)| \equiv |P_t^{ex}(\mathbf{A}) - P_\infty^{ex}(\mathbf{A})|.$$

Let us denote by $\mathbf{A}^+ \equiv \mathbf{A} \cap \Omega^+$, and by $\mathbf{A}^- \equiv \mathbf{A} \cap \Omega^-$, so $\mathbf{A} = \mathbf{A}^+ \sqcup \mathbf{A}^-$. Notice that even though in the limit we fully discover \mathbf{A}^+ , there may be some $t \in \mathbb{N}_0$ such that $\mathbf{A}^+ \cap \Omega_t^+ \neq \emptyset \neq \mathbf{A}^+ \cap \Omega_t^-$. So we have

$$|P_t^{ex}(\mathbf{A}) - P_\infty^{ex}(\mathbf{A})| = |P_t^{ex}(\mathbf{A}^+) + P_t^{ex}(\mathbf{A}^-) - P_\infty^{ex}(\mathbf{A}^+) - P_\infty^{ex}(\mathbf{A}^-)| \quad (\text{A.5})$$

$$= |P_t^{ex}(\mathbf{A}^+) - P_\infty^{ex}(\mathbf{A}^+)| \quad (\text{A.6})$$

$$= |P(\mathbf{A}^+ \cap \Omega_t^+) - P(\mathbf{A}^+ \cap \Omega_t^-) - P(\mathbf{A}^+ \cap \Omega_t^+) - P(\mathbf{A}^+ \cap \Omega_t^-)| \quad (\text{A.7})$$

$$= 2P(\mathbf{A}^+ \cap \Omega_t^-).$$

Equation (A.5) comes from $\mathbf{A} = \mathbf{A}^+ \sqcup \mathbf{A}^-$ and (ii*). Equation (A.6) comes from the fact that, for all t , $P_t^{ex}(\mathbf{A}^-) = P_\infty^{ex}(\mathbf{A}^-)$, because \mathbf{A}^- is the portion of \mathbf{A} that never leaves the latent space. Equation (A.7) comes from the updating procedure described in Section 3.2, from the countable additivity of P , and from $\mathbf{A}^+ = (\mathbf{A}^+ \cap \Omega_t^+) \sqcup (\mathbf{A}^+ \cap \Omega_t^-)$.

Now, notice that the limit as $t \rightarrow \infty$ of $\mathbf{A}^+ \cap \Omega_t^-$ is

$$\mathbf{A}^+ \cap \bigcap_{t \in \mathbb{N}_0} \Omega_t^- = \mathbf{A}^+ \cap \Omega^- = \emptyset,$$

where the last equality is by construction. Then, by the continuity of P we have that

$$\lim_{t \rightarrow \infty} d_{ETV}(P_t^{ex}, P_\infty^{ex}) = 2 \lim_{t \rightarrow \infty} P(\mathbf{A}^+ \cap \Omega_t^-) = 2P(\emptyset) = 0,$$

which concludes the proof. \square

Proof of Proposition 9. Suppose that $P_\infty^{ex}(\Omega) \neq 1$. This of course implies that $P_\infty^{ex}(\Omega) < 1$ because, from the definition of extended probabilities, $P_\infty^{ex}(A) \in [-1, 1]$, for all $A \in \mathcal{F}$. Then, this means that there exists a set $A \in \mathcal{F}$ such that $P_\infty^{ex}(A) \leq 0$, which implies $A \subset \Omega_\infty^-$. But we know that, if $\Omega_t^+ \uparrow \Omega$, then $\Omega_t^- \downarrow \Omega_\infty^- = \emptyset$, which contradicts $A \subset \Omega_\infty^-$. \square

Proof of Theorem 2. Pick any lower extended probability \underline{P}^{ex} and suppose there exists $\emptyset \neq \mathcal{P}^{ex} \subset \Delta^{ex}(\Omega, \mathcal{F})$ such that $\underline{P}^{ex}(A) = \inf_{P^{ex} \in \mathcal{P}^{ex}} P^{ex}(A)$, for all $A \in \mathcal{F}$. Now, by Theorem 1, we know that P^{ex} is coherent, for all $P^{ex} \in \mathcal{P}^{ex}$. This means that for all $\{B_j\}_{j=1}^n \subset \mathcal{B}'$, all $\{s_j\}_{j=1}^n \subset \mathbb{R}$, and all $P^{ex} \in \mathcal{P}^{ex}$,

$$\sup_{\omega \in \Omega} \left\{ \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - P^{ex}(B_j)] \right\} \geq 0. \quad (\text{A.8})$$

Notice that we use \mathcal{B}' because that is the sigma-algebra generated by the events we can enter a bet about. Then, (A.8) entails that

$$\sup_{\omega \in \Omega} \left[\sum_{j=1}^n s_j \mathbb{1}_{B_j}(\omega) \right] \geq \sum_{j=1}^n s_j P^{ex}(B_j). \quad (\text{A.9})$$

The inequality in (A.9) implies that

$$\begin{aligned} \sup_{\omega \in \Omega} \left[\sum_{j=1}^n s_j \mathbb{1}_{B_j}(\omega) \right] &\geq \inf_{P^{ex} \in \mathcal{P}^{ex}} \sum_{j=1}^n s_j P^{ex}(B_j) \\ &\geq \sum_{j=1}^n s_j \inf_{P^{ex} \in \mathcal{P}^{ex}} P^{ex}(B_j) \\ &= \sum_{j=1}^n s_j \underline{P}^{ex}(B_j). \end{aligned} \quad (\text{A.10})$$

The results in (A.10) hold if and only if $\sup_{\omega \in \Omega} \left\{ \sum_{j=1}^n s_j [\mathbb{1}_{B_j}(\omega) - \underline{P}^{ex}(B_j)] \right\} \geq 0$. The claim follows. \square

Proof of Corollary 2. Pick a generic lower extended probability \underline{P}^{ex} , and suppose $\text{core}(\underline{P}^{ex})$ is nonempty. Then, \underline{P}^{ex} avoids sure loss by Theorem 2. \square

Proof of Proposition 10. We first show that $\text{core}(\underline{P}^{ex})$ is convex. Pick any P_1^{ex}, P_2^{ex} in the core of \underline{P}^{ex} , any $\alpha \in (0, 1)$, and any $A \in \mathcal{F}$. We have

$$\alpha P_1^{ex}(A) + (1 - \alpha) P_2^{ex}(A) \geq \alpha \underline{P}^{ex}(A) + (1 - \alpha) \underline{P}^{ex}(A) = \underline{P}^{ex}(A),$$

so $\alpha P_1^{ex} + (1 - \alpha) P_2^{ex} \in \text{core}(\underline{P}^{ex})$.

We then show that $\text{core}(\underline{P}^{ex})$ is weak*-compact. Recall that, in the weak* topology, a net $(P_\alpha^{ex})_{\alpha \in I}$ converges to P^{ex} if and only if $P_\alpha^{ex}(A) \rightarrow P^{ex}(A)$, for all $A \in \mathcal{F}$. This proof is similar to the proof of [46, Proposition

3], where the authors prove the same claim for the core of a bounded game. Pick any $P^{ex} \in \text{core}(\underline{P}^{ex})$, and let $k := 2 \sup_{A \in \mathcal{F}} |\underline{P}^{ex}(A)|$. For any $A \in \mathcal{F}$, it holds that $P^{ex}(A) \geq \underline{P}^{ex}(A) \geq -k$. On the other hand, for all $A \in \mathcal{F}$, we have that

$$P^{ex}(A) = P^{ex}(\Omega) - P^{ex}(A^c) \leq \underline{P}^{ex}(\Omega) - \underline{P}^{ex}(A^c) \leq 2 \sup_{A \in \mathcal{F}} |\underline{P}^{ex}(A)|.$$

This implies that $|P^{ex}(A)| \leq k$, for all $A \in \mathcal{F}$. By [47, Page 94], we have that

$$\|P^{ex}\| := \sup \sum_{j=1}^n |P^{ex}(A_j) - P^{ex}(A_{j-1})| \leq 2k, \quad (\text{A.11})$$

where the supremum is taken over all finite chains $\emptyset = A_0 \subset A_1 \subset \dots \subset A_n = \Omega$. Then, (A.11) implies that

$$\text{core}(\underline{P}^{ex}) \subset \{P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : \|P^{ex}\| \leq 2k\}.$$

By the Alaoglu Theorem [47, Theorem 2, Page 424], we know that $\{P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F}) : \|P^{ex}\| \leq 2k\}$ is weak*-compact. Hence, to complete the proof, we are left to show that $\text{core}(\underline{P}^{ex})$ is weak*-closed. Let then $(P_\alpha^{ex})_{\alpha \in I}$ be a net in $\text{core}(\underline{P}^{ex})$ that weak*-converges to $P^{ex} \in \Delta^{ex}(\Omega, \mathcal{F})$. Using the properties of the weak* topology, it is easy to see that $P^{ex} \in \text{core}(\underline{P}^{ex})$. Hence, $\text{core}(\underline{P}^{ex})$ is weak*-closed. The claim follows. \square

Proof of Proposition 11. Equation (20) comes from the fact that, for all t ,

$$P_{t+1}^{ex}(E_{t+1,\omega}^+) = |P_t^{ex}(E_{t,\omega}^-)|,$$

that is, $P_{t+1}^{ex}(E_{t+1,\omega}^+) = -P_t^{ex}(E_{t,\omega}^-)$, and that

$$-\overline{P}_t^{ex}(E_{t,\omega}^-) \leq -P_t^{ex}(E_{t,\omega}^-) \leq -\underline{P}_t^{ex}(E_{t,\omega}^-).$$

\square

A.2. Application to opinion dynamics

This example comes from the model in [22]. There, opinion dynamics between a persuaded and a persuading agents is studied, in particular when the persuaded agent evaluates new information in a way that is consistent with her own preexisting belief. In this example we are going to adoperate extended probabilities to model the boomerang effect. This phenomenon corresponds to the empirical observation that sometimes persuasion yields

the opposite effect: the persuaded agents moves her opinion away from the opinion of the persuading agent. That is, she enforces her old opinion.

In [22], the authors assume that the state of the world does not change, that the agents are aware of this fact, and that the persuaded agent changes her opinion only under the influence of the opinion of the persuading agent. They model this dynamic in a linear fashion. The first iteration is the following

$$\hat{P}_1(\{\omega_k\}) = \epsilon P_0(\{\omega_k\}) + (1 - \epsilon)Q(\{\omega_k\}), \quad \epsilon \in [0, 1], \quad (\text{A.12})$$

where $\omega_k \in \Omega$, a finite sample space, and P_0 and Q are regular probability measures that represent the initial opinions of the persuaded and the persuading agents, respectively. Q is not indexed to time because they make the simplifying assumption that the persuading agent does not change his mind: he tries to persuade the other agent of the same thing at every iteration. ϵ is a weight, and several qualitative factors contribute to its subjective assessment: egocentric attitude of the persuaded agent, the fact that the persuaded agent has access to internal reasons for choosing her opinion, while she is not aware of the internal reasons of the persuading agent, and many more. The authors relate ϵ to the credibility of the persuading agent: the higher ϵ , the less credible he is. The successive iterations are modeled as follows

$$\hat{P}_{t+1}(\{\omega_k\}) = \epsilon \hat{P}_t(\{\omega_k\}) + (1 - \epsilon)Q(\{\omega_k\}), \quad (\text{A.13})$$

for $t \geq 1$. This means that at every iteration, the persuading agent tries to shift the persuaded agent's opinion closer to his own. This continues until either the persuaded agent is content with her opinion (and hence does not further change her beliefs), or the persuading agent completely convinces the other agent.

Now, one way to model the boomerang effect is to consider $\epsilon > 1$. This conveys the idea that the persuading agent has an extremely low credibility. This results in the updated opinion of the persuaded agent to be an extended probability measure. Indeed, for $\epsilon > 1$, $\hat{P}_{t+1}(\{\omega_k\})$ is negative whenever $\epsilon \hat{P}_t(\{\omega_k\}) < |1 - \epsilon|Q(\{\omega_k\})$. In [22] the authors consider the induced regular probability measure to avoid working with extended probabilities.

We modify slightly the linear model in [22]. The main differences are three: we allow the use of extended probabilities, we describe a state space that is divided in latent and known, and whose composition is gradually discovered as the time passes by, and we do not let ϵ be a free parameter.

It depends on both the element ω_k of the state space we examine and on the iteration t we are considering.

Mathematically, we can describe the low credibility of the persuading agent through hidden states: the persuaded agent may think that she does not know the state space well enough, that is, that there are some hidden portions of Ω she is not (yet) aware of.

We assume that the state space $\Omega = \{\omega_1, \dots, \omega_N\}$ is a finite set (a common simplifying assumption in opinion dynamics). Notice that the finest possible partition of Ω is $\mathcal{E} = \{\{\omega_j\}\}_{j=1}^N$. At the beginning of the interaction between persuading and persuaded agents, the former is fully aware of the composition of the state space, while the latter is only aware of the composition of $\Omega_0^+ \subsetneq \Omega$, but suspects that the actual state space is larger. This corresponds to having suspects on the persuading agent hiding some pieces of information. In particular, she correctly guesses that the true state space is Ω . This correct guess is without loss of generality for our analysis: we are in the $\Omega_t^+ \uparrow \Omega$ case; we also assume that the discovering procedure is equivalent to an urn without replacement. She then defines an extended probability on Ω the way we explained in section 3. That is, she specifies the oracle probability distribution P on the whole Ω and then flips the sign to the probabilities of the latent events: $P_0^{ex}(\{\omega_k\}) = P(\{\omega_k\})$ if $\omega_k \in \Omega_0^+$, and $P_0^{ex}(\{\omega_k\}) = -P(\{\omega_k\})$ if $\omega_k \notin \Omega_0^+$.

To obtain the influenced extended probability at any time $t \geq 0$, we modify slightly equation (A.13) to get

$$\hat{P}_t^{ex}(\{\omega_k\}) = \epsilon_{k,t} P_t^{ex}(\{\omega_k\}) + (1 - \epsilon_{k,t}) Q(\{\omega_k\}). \quad (\text{A.14})$$

The nonnegative $\epsilon_{k,t}$'s have to be chosen such that \hat{P}_t^{ex} is an extended probability measure, that is, $\hat{P}_t^{ex}(\{\omega_k\}) \in [-1, 1]$ for all k , for all t , and

$$\hat{P}_t^{ex}(\Omega) = \sum_{k=1}^N \hat{P}_t^{ex}(\{\omega_k\}) \leq 1.$$

Here, \hat{P}_t^{ex} denotes the influenced extended probability at time t . Notice that \hat{P}_t^{ex} is analytically similar to an ϵ -contaminated probability measure. There are of course two major differences: for some ω_k , $\epsilon_{k,t}$ is greater than 1, and also one of the elements of the mixture is an extended probability measure (rather than a regular one). Notice also that Q is not indexed to time because we too make the simplifying assumption that the persuading agent does not change his mind. Another characteristic worth noting is that in our model, at every iteration t , the persuaded agent combines her updated belief (expressed via the extended probability P_t^{ex}) with the other

agent's belief to obtain \hat{P}_t^{ex} . This is different from the model in [22] where at every iteration t the persuaded agent combines her influenced belief at iteration $t - 1$ (expressed through \hat{P}_{t-1}) with the other agent's belief to obtain \hat{P}_t . This because in their model the world does not change, so the only way of describing an opinion dynamics is the one the authors illustrate.

The persuaded agent updates P_t^{ex} as specified in section 3. That is, when at time t she observes ω_k that used to belong to the latent space, $P_t^{ex}(\{\omega_k\}) = |P_{t-1}^{ex}(\{\omega_k\})|$, while the extended probabilities for $\omega_s \neq \omega_k$ are kept constant. Let us be more precise about the differences with equation (A.12); $\epsilon_{k,t}$ depends on both k and t . The persuaded agent has a different perception of the opinion of the persuading agent depending on whether she is not sure the topic they are debating about belongs to the state space Ω , so for ω_k in the latent space, $\epsilon_{k,t}$ is greater than 1. In addition, as time passes by, the hidden elements of the state space become known, so (part of) the credibility of the persuading agent is restored. The $\epsilon_{k,t}$ associated with ω_k observed at time t becomes smaller than 1, for all ω_k .

Notice that, for $t \geq N$, P_t^{ex} is a regular probability measure, because the persuaded agent discovers the composition of the whole state space, so the latent space shrinks to the empty set.

Throughout this section we made the tacit assumption that \mathcal{P} , the set of probability measures on Ω that induces the set of extended probabilities \mathcal{P}_0^{ex} at time $t = 0$, is the singleton $\{P\}$, so that $\mathcal{P}_0^{ex} = \{P_0^{ex}\}$. This simplifying assumption can be dropped, and the analysis stays the same. The only difference is that we have to repeat it for all the elements of \mathcal{P}_t^{ex} , for all $t \in \mathbb{N}_0$. Every set \mathcal{P}_t^{ex} of extended probabilities induces a set $\hat{\mathcal{P}}_t^{ex}$ of influenced extended probabilities. For all $t \geq N$, \mathcal{P}_t^{ex} is a set of regular probability measures, and $\hat{\mathcal{P}}_t^{ex}$ is a set of influenced regular probability measures.

A.3. DeFinettian interpretation of subjective probability

De Finetti admits that it might have been better to adopt the seemingly more general approach of Ramsey and Savage of defining bets whose payoffs are in utils [23, Page 79]:

The formulation [...] could be made watertight [...] by working in terms of the utility instead of with monetary value. This would undoubtedly be the best course from the theoretical point of view, because one could construct, in an integrated fashion, a theory of decision-making [...] whose meaning would be unexceptionable from an economic viewpoint, and which would

establish simultaneously and in parallel the properties of probability and utility on which it depends.

Nevertheless, he found “other reasons for preferring” the money bet approach [23, Page 81]:

The main motivation lies in being able to refer, in a natural way to combinations of bets, or any other economic transactions, understood in terms of monetary value (which is invariant). If we referred ourselves to the scale of utility, a transaction leading to a gain of amount S if the event E occurs would instead appear as a variety of different transactions, depending on the outcome of other random transactions. These, in fact, cause variations in one’s fortune, and therefore in the increment of utility resulting from the possible additional gain S : conversely, suppose that in order to avoid this one tried to consider bets, or economic transactions, expressed, let us say, in “utiles” (units of utility, definable as the increment between two fixed situations). In this case, it would be practically impossible to proceed with the transactions, because the real magnitudes in which they have to be expressed (monetary sums or quantities of goods, etc.) would have to be adjusted to the continuous and complex variations in a unit of measure that nobody would be able to observe.

References

- [1] A. Khrennikov, *Interpretations of probability*. Berlin/New York : Walter de Gruyter (2009).
- [2] P. A. M. Dirac, Note on exchange phenomena in the Thomas atom, *Mathematical Proceedings of the Cambridge Philosophical Society*. **26**, 376–395 (1930).
- [3] W. K. Heisenberg, Über die inkohärente Streuung von Röntgenstrahlen, *Physikalische Zeitschrift*. **32**, 737–740 (1931).
- [4] C. Ferrie, Quasi-probability representations of quantum theory with applications to quantum information science, *Reports on Progress in Physics*. **74** (11), 116001 (2011).
- [5] J. B. Hartle, Quantum mechanics with extended probabilities, *Physical Review A*. **78:012108** (2008).
- [6] F. Kronz, Non-monotonic probability theory and photon polarization, *Journal of Philosophical Logic*. **36**, 449–472 (2007).
- [7] M. Burgin and G. Meissner, Negative probabilities in modeling random financial processes, *Integration: Mathematical Theory and Applications*. **2(3)**, 305–322 (2011).
- [8] R. A. Jarrow and S. M. Turnbull, Pricing derivatives on financial securities subject to credit risk, *Journal of Finance*. **L(1)**, 53–85 (1995).
- [9] D. Lowe, Machine learning, uncertain information, and the inevitability

- of negative ‘probabilities’, Available at *Machine Learning Workshop 2004* (2004/2007).
- [10] E. W. Wiewiora. *Modeling probability distributions with predictive state representations*. PhD thesis, University of California, San Diego (2008).
- [11] H. C. Tijms and K. Staats, Negative probabilities at work in the M/D/1 queue, *Probability in the Engineering and Informational Sciences*. **21**(1), 67–76 (2007).
- [12] E. H. Allen, Negative probabilities and the uses of signed probability theory, *Philosophy of Science*. **43**(1), 53–70 (1976).
- [13] M. S. Bartlett, Negative probability, *Mathematical Proceedings of the Cambridge Philosophical Society*. **41**, 71–73 (1945).
- [14] M. Burgin, Negative probability in the framework of combined probability, Available at *arXiv:1306.1166* (2013).
- [15] M. Burgin, Integrating random properties and the concept of probability, *Integration: Mathematical Theory and Applications*. **3**(2), 137–181 (2012).
- [16] A. Benavoli, A. Facchini, and M. Zaffalon, Computational complexity and the nature of quantum mechanics (extended version), Available at *arXiv:1902.03513* (2019).
- [17] M. Gell-Mann and J. B. Hartle, Decoherent histories quantum mechanics with one “real” fine-grained history, *Physical Review A*. **85:062120** (2012).
- [18] J. B. Hartle, Linear positivity and virtual probability, *Physical Review A*. **70:022104** (2004).
- [19] R. Gong. Modeling uncertainty with sets of probabilities. In *Foundations of Probability seminar series*, Rutgers University (2018).
- [20] M. Caprio and R. Gong, Dynamic precise and imprecise probability kinematics, Available on *arXiv:2110.04382* (2022).
- [21] M. Caprio and S. Mukherjee, Ergodic theorems for dynamic imprecise probability kinematics, Available on *arXiv:2003.06502* (2022).
- [22] A. E. Allahverdyan and A. Galstyan, Opinion dynamics with confirmation bias, *PLoS One*. **9**(7) (2014).
- [23] B. de Finetti, *Theory of Probability*. vol. 1, New York : Wiley (1974).
- [24] P. Walley, *Statistical reasoning with imprecise probabilities*. vol. 42, *Mono-graphs on Statistics and Applied Probability*, London : Chapman and Hall (1991).
- [25] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *The Annals of Mathematical Statistics*. **38**(2), 325–339 (1967).
- [26] P. Walley, Inferences from multinomial data: Learning about a bag of marbles, *Journal of the Royal Statistical Society. Series B (Methodological)*. **58**(1), 3–57 (1996).
- [27] B. de Finetti, *Theory of Probability*. vol. 2, New York : Wiley (1975).
- [28] R. F. Nau, De Finetti was right: Probability does not exist, *Theory and Decision*. **51**, 89–124 (2001).
- [29] F. P. Ramsey. Truth and Probability. In eds. H. E. K. Jr. and H. E. Smokler, *Studies in Subjective Probability*, pp. 61–92. New York : Wiley (1964).
- [30] L. J. Savage, *The Foundations of Statistics*. New York : John Wiley and Sons (1954).

- [31] B. de Finetti, La prévision : ses lois logiques, ses sources subjectives, *Annales de l'institut Henri Poincaré*. **7**(1), 1–68 (1937).
- [32] G. J. Székely, Half of a coin: Negative probabilities, *Wilmott Magazine*. pp. 66–68 (2005).
- [33] K. A. Bollen, Latent variables in psychology and the social sciences, *Annual Review of Psychology*. **53**, 605–634 (2002).
- [34] Y. Hu, The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics, *Journal of Econometrics*. **200**(2), 154–168 (2017).
- [35] S. Rabe-Hesketh and A. Skrondal, Classical latent variable models for medical research, *Statistical Methods in Medical Research*. **17**(1), 5–32 (2008).
- [36] E. P. Wigner, On the quantum correction for thermodynamic equilibrium, *Physical Review*. **40**, 749–759 (1932).
- [37] A. Blass and Y. Gurevich, Negative probabilities: What they are and what they are for, *Available at arXiv:2009.10552* (2022).
- [38] Y. Gurevich and V. Vovk, Betting with negative probabilities, *The Game-Theoretic Probability and Finance Project, Working Paper #58* (2021).
- [39] A. Benavoli, A. Facchini, and M. Zaffalon, The weirdness theorem and the origin of quantum paradoxes, *Foundations of Physics*. **51**(95) (2021).
- [40] B. H. Feintzeig and S. C. Fletcher, On noncontextual, non-Kolmogorovian hidden variable theories, *Foundations of Physics*. **47**, 294–315 (2017).
- [41] J. Tsitsiklis. Sample spaces. In *Introduction to Probability (online course)*, Massachusetts Institute of Technology (2018).
- [42] J. O. Berger. The robust Bayesian viewpoint. In ed. J. B. Kadane, *Robustness of Bayesian Analyses*. Amsterdam : North-Holland (1984).
- [43] D. Ellsberg, Risk, ambiguity, and the Savage axioms, *The Quarterly Journal of Economics*. **75**(4), 643–669 (1961).
- [44] K. P. S. B. Rao and M. B. Rao, *Theory of charges, a study of finitely additive measures*. London : Academic Press (1983).
- [45] T. Fischer, Existence, uniqueness, and minimality of the Jordan measure decomposition, *Available at arXiv:1206.5449* (2012).
- [46] M. Marinacci and L. Montrucchio. Introduction to the mathematics of ambiguity. In ed. I. Gilboa, *Uncertainty in economic theory: a collection of essays in honor of David Schmeidler's 65th birthday*, London : Routledge (2004).
- [47] N. Dunford and J. T. Schwartz, *Linear operators, part I: general theory*. London : Wiley Interscience (1958).