BAYESIAN DECISION THEORY AND STOCHASTIC INDEPENDENCE[1]

by Philippe Mongin (CNRS & HEC Paris)
mongin@greg-hec.com

March 2018

Abstract. Stochastic independence (SI) has a complex status in probability theory. It is not part of the definition of a probability measure, but it is nonetheless an essential property for the mathematical development of this theory, hence a property that any theory on the foundations of probability should be able to account for. Bayesian decision theory, which is one such theory, appears to be wanting in this respect. In Savage's classic treatment, postulates on preferences under uncertainty are shown to entail a subjective expected utility (SEU) representation, and this permits asserting only the existence and uniqueness of a subjective probability, regardless of its properties. What is missing is a preference postulate that would specifically connect with the SI property. The paper develops a version of Bayesian decision theory to fill this gap. In a framework of multiple sources of uncertainty, we introduce preference conditions that jointly entail the SEU representation and the property that the subjective probability in this representation treats the sources of uncertainty as being stochastically independent. We give two representation theorems of graded complexity to demonstrate the power of our preference conditions. Two sections of comments follow, one connecting the theorems with earlier results in Bayesian decision theory, and the other connecting them with the foundational discussion on SI in probability theory and the philosophy of probability.

1 Introduction and preview

The property of stochastic (or statistical, or probabilistic) independence occupies a rather special place in the mathematical theory of probability. It does

[1]A first version of this paper was presented at a seminar at the Munich Center for Mathematical Philosophy and at the TARK 2017 conference. The present version has particularly benefited from comments made by Donald Gillies, Joseph Halpern, Marcus Pivato, Burkhard Schipper, Paul Weirich and two anonymous TARK referees. Special thanks are due to Richard Bradley and Peter Wakker for their detailed suggestions. We thank the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) for supporting our research.

not belong to the properties that this theory singles out to define a probability measure axiomatically. Indeed, its familiar definitions by the multiplication rule, or the equality of conditional with unconditional probability, do not enter the Kolmogorov axiomatization of a probability measure. Rather, they capture properties of given events (and more generally, of given partitions or random variables) for a given probability measure and can thus be adopted only to model particular situations. At the same time, probability theory obviously makes extensive use of independence assumptions, as evidenced by the Laws of Large Numbers, many theorems on stochastic processes and some core results of statistical theory. For Kolmogorov himself, this property occupies a "central position in the theory of probability" (1933-1956, p. 8), making it significantly different from the theory of positive measures. One would thus expect all theories of the foundations of probability to pay careful attention to stochastic independence, but this is not the case. In this paper, we investigate Bayesian decision theory, one of the most influential among these theories, and at the same time an example of neglect of this major property.

As is well known, Bayesian decision theorists have developed a brand of subjective interpretation for the probability calculus. They claim that an agent's uncertain beliefs should be represented by a probability measure, and ground their claim on a pragmatic argument. They show formally that if the agents' preferences over uncertain prospects - typically, but not exclusively over monetary bets - obey certain requirements of practical rationality, these agents' beliefs should conform to the axiomatic definition of a probability measure. Bayesian decision theorists hardly go beyond this demonstration. As long as they have nothing to add on stochastic independence, it remains unclear whether they have established an appropriate connection between the probability calculus and rational belief. They can be criticized for handling only the basics of the probability calculus, and not its actual working.

More technically, Bayesian decision theory offers a representation theorem for preferences over uncertain prospects that involves two sets of quantities, utilities (over the consequences of prospects) and probabilities (over the uncertain events), these two items being combined by the familiar rule of expected utility (EU). After Ramsey's and de Finetti's sketches, this strategy was implemented in full axiomatic detail by Savage (1954-1972). At one point, he extends the representation theorem to obtain a posterior probability measure, i.e., one that represents beliefs after a partial resolution of uncertainty, and proves that this posterior obeys Bayes's rule of revision; properly speaking, the "Bayesian" label of the school becomes fully justified only at this stage. This is also where Savage stops. As we will report below, however, he acknowledges that a treatment of stochastic independence should have come next, but his honest admission of an unfinished business was lost on most Bayesian decision theorists. The rare exceptions will be discussed below.

What is missing is a further condition put on the agent's preferences under uncertainty that would account for that property specifically, and the aim of the present paper is to provide one. We set up a framework in which there are two distinct sources of uncertainty S and T, and states of nature thus have the form of two-component vectors $(s,t) \in S \times T$. Our proposed condition is stated in terms of the agent's *conditional preferences*: those defined conditionally on s should be the same for all s in S, and those defined conditionally on t should be the same for all t in T. What this intuitively says is that if the agent is uncertain on s but not on t, or is uncertain on t but not on s, then such a partial knowledge cannot improve the decisions made by the agent under the residual uncertainty. We thus recover in preference terms one of the standard informal justifications of stochastic independence: if the occurrence of an event A *carries no information on* the occurrence of B, and vice versa, then the two events should be declared to be independent. An alternative informal justification, which is also common, goes as follows: if the occurrence of an event A *does not influence* the occurrence of B, and vice versa, then the two events should be declared to be independent. The two lines are semantically distinct, but easily get mixed up in probability texts and even some works in the philosophy of probability. In the Bayesian decision theory of this paper, there is no danger of confusion since the information carried by events matters only if it enters the agent's decision process, hence is subjective in character; objective connections holding between events, for instance causal connections, play a role only if the agent considers them relevant. Thus, one merit of the theory is to give the informational reading of stochastic independence a foundation that clearly sets it apart from other possible readings. Besides this contribution, the theory casts some light on the reflective discussion of stochastic independence in probability theory and the philosophy of probability. In particular, it has something to say on the *symmetry* of the multiplicative definition - A's independence of B implying B's independence of A -, a property that these literatures have often called into question.

That stochastic independence follows from the above mentioned preference condition can be checked by *assuming* that the EU formula holds, and seeing what this condition entails for the shape of subjective probability entering the formula. However, this can only be an heuristic step towards the theoretical work, for Bayesian decision theory makes a strong point of taking the agent's preferences as its only primitive terms. Consistently with this, one should devise a system of preference conditions that *jointly* delivers the EU rule and the stochastic independence property of the subjective probability for relevant events. The main task of this paper is to set out such a large system. Its EU part is more elementary than Savage's system, but for that reason also much handier, and we will argue that it compares favourably with other variants of this canonical system such as Anscombe and Aumann's (1963). Our technical source is in the recent work by Mongin and Pivato (2015, 2017).

We offer two representation theorems in succession. Both are adapted to the multiple uncertainty framework sketched above, and both deliver an EU formula in which the subjective probability is multiplicative in the values of two sources, hence satisfies stochastic independence. The highly condensed axiomatic system of Theorem 1 makes it possible to reach both the EU formula and the multiplicative property in one go. This brevity is both an advantage and a disadvantage. To derive the EU formula and the multiplicative property in two succesive logical steps helps one better understand how each of the preference axioms contributes to the conclusions. The more advanced Theorem 2 is devised for this purpose.

The rest of this paper is organized as follows. Section 2 introduces the twofold uncertainty framework and the preference conditions for Theorem 1 via a motivating example. Section 3 states Theorem 1 formally, and section 4 does the same for Theorem 2. Section 5 is devoted to technical comparisons within Bayesian decision theory and section 6 to conceptual comparisons within the philosophy of probability. An appendix gives proof details on the two theorems.

2 A motivating example

Given any probability space $(\Omega, \mathcal{A}, P)$, two events $A, B \in \mathcal{A}$ are said to be *stochastically independent* if

$$P(A \cap B) = P(A).P(B).$$

Building on this elementary definition, probability theory also defines what it means for sets of events, partitions or random variables to be stochastically independent. Here we will approach stochastic independence (SI) by specializing the state set Ω to be a *product set*, a standard move in the theory when it comes to working with this property (see, e.g., Halmos, 1974, p. 191-192). The simple example of this section illustrates this framework and the main decision-theoretic concepts of the paper.

Suppose that a corn producer must decide how much land to farm while not knowing what the climatic conditions and the demand for corn will be at the time of the harvest; supppose also this producer evaluates each farming policy in terms of monetary proceeds and no other criterion. Following the basics of decision theory, we can reexpress his example symbolically as follows. There is a set of states of the world, which takes the form of a product set $\Omega = S \times T$, where S represents the set of unknown climatic conditions and T the set of unknown values for demand. There is a set of consequences, which we take to be the set of real numbers $\mathbb{R}$ to represent monetary proceeds. There is a set of uncertain prospects, i.e., mappings from the states of the world to consequences, each of which represents a farming policy, which we take to be $\mathbb{R}^{S \times T}$, the set of real functions on Ω. Finally, there is a binary relation $\succsim$ on the last set of prospects to capture the producer's preferences among cultivation policies.

Now suppose that this preference relation obeys *the EU rule*, i.e., there exists a probability function π on $\Omega = S \times T$ and a utility function u on $\mathbb{R}$ such that for all uncertain prospects $\mathbf{X}$, $\mathbf{Y}$,

$$\mathbf{X} \succsim \mathbf{Y} \text{ iff} \sum_{s,t} \pi(s,t) u(\mathbf{X}(s,t) \geq \sum_{s,t} \pi(s,t) u(\mathbf{Y}(s,t),$$

and moreover that π satisfies the *stochastic independence* (SI) property with respect to S and T, i.e.,

$$\pi(s,t) = p(s)q(t),$$

where p and q are probability functions on S and T respectively. This equation, also written as $\pi = p \otimes q$, determines p and q uniquely; these are the *marginals* of π on S and T, respectively. The EU rule and SI property are our intended conclusions; in this section, we reason heuristically, working backwards from them to a set of preference conditions that could be proposed as axioms.

Assume for simplicity that $S = \{s_1, s_2\}$ and $T = \{t_1, t_2\}$. Then, the probabilities are given by the table:

	t_1	t_2
s_1	$p_{s_1} q_{t_1}$	$p_{s_1} q_{t_2}$
s_2	$p_{s_2} q_{t_1}$	$p_{s_2} q_{t_2}$

and an uncertain prospect $\mathbf{X}$ is represented by the following table, which gives the consequences of this prospect in each state:

$\mathbf{X}$	t_1	t_2
s_1	x_{11}	x_{12}
s_2	x_{21}	x_{22}

The EU formula for $\succsim$:

$$V(\mathbf{X}) = p_{s_1} q_{t_1} u(x_{11}) + p_{s_1} q_{t_2} u(x_{12}) + p_{s_2} q_{t_1} u(x_{21}) + p_{s_2} q_{t_2} u(x_{22}).$$

can be restated either as:

$$(*)\ V(\mathbf{X}) = p_{s_1} \left[q_{t_1} u(x_{11}) + q_{t_2} u(x_{12}) \right] + p_{s_2} \left[q_{t_1} u(x_{21}) + q_{t_2} u(x_{22}) \right],$$

or as:

$$(**)\ V(\mathbf{X}) = q_{t_1} \left[p_{s_1} u(x_{11}) + p_{s_2} u(x_{21}) \right] + q_{t_2} \left[p_{s_1} u(x_{12}) + p_{s_2} u(x_{22}) \right].$$

Observe that the bracketed sums in $(*)$ are numerical representations for conditional preferences on the possible values of s, and those in $(**)$ are numerical representations for conditional preferences on the possible values of t. Thus, the equations entail that (i) *conditional preferences are orderings.* Since the same functional form $q_{t_1} u(.) + q_{t_2} u(.)$ appears in the two bracketed sums of $(*)$, and similarly the same functional form $p_{s_1} u(.) + p_{s_2} u(.)$ appears in the two bracketed sums of $(**)$, the equations also entail that (ii) *conditional preferences are the same for different s, and the same for different t.* Lastly, from the same equations, if the conditional orderings for both s_1 and s_2, or the conditional orderings for both t_1 and t_2, agree to rank prospect $\mathbf{X}$ above prospect $\mathbf{Y}$, then the overall preference $\succsim$ ranks $\mathbf{X}$ above $\mathbf{Y}$. Thus, it also holds that (iii) *preferences over prospects are increasing with respect to either family of conditional preferences.*

We have stated (i), (ii) and (iii) in the preference language and by abstracting entirely from the specifics of EU. Each of the three can indeed be satisfied by more general theories, and in particular, the *dominance* property (iii) is well known to apply to most existing alternatives to EU theory (like rank-dependent theory, see, e.g., Wakker, 2010).

In Theorem 1 below, we assume (i), (ii) and (iii), plus some background conditions, thus showing that these conditions are not only necessary, but also sufficient for our desired conclusions. Given the definition of a conditional, which is restated below, it is actually possible to fuse (i) with (iii) and obtain an even more condensed system. One may wonder how such apparently weak conditions can do the mathematical work. The key point is that they apply to s and t *at the same time*, and this creates the possibility of representing the preference $\succsim$ both in terms of s-conditionals and t-conditionals. Identifying these representations leads to the results. The equivalence between them thus transpires here from the algebraic equivalence between the two factorizations $(*)$ and $(**)$.

3 A first EU representation theorem involving stochastic independence

Formally, there are two variables of interest, $s \in S$ and $t \in T$, and a state of the world is any pair $(s,t) \in \Omega = S \times T$. By assumption, S and T are finite with cardinalities $|S|$, $|T| \geq 2$. We keep the same number of factors in the product set Ω as in the motivating example, but this is only for mathematical simplicity. The next section will illustrate how a larger number of factors can be handled. We take the set of consequences to be $\mathbb{R}$ and the set of prospects to $\mathbb{R}^{S\times T}$.[2] The sets of all probability functions on S, T and $S \times T$ are denoted by Δ_S, Δ_T and $\Delta_{S\times T}$, respectively.

It is convenient to represent prospects $\mathbf{X}$ as $|S| \times |T|$ matrices, with each s standing for a row and each t standing for a column. We will thus write $\mathbf{X} = [x_s^t]_{s\in S}^{t\in T}$, but sometimes also $\mathbf{X} = (\mathbf{x}_1,...,\mathbf{x}_{|S|})$, where each component is a row vector $\mathbf{x}_s \in \mathbb{R}^T$, or $\mathbf{X} = (\mathbf{x}^1,...,\mathbf{x}^{|T|})$, where each component is a column vector $\mathbf{x}^t \in \mathbb{R}^S$.

By assumption, the agent compares prospects in terms of an *ex ante* preference relation $\succsim$. As a maintained assumption, we take this relation to be a continuous weak ordering, hence representable by a continuous utility function. The other preference relations are obtained from $\succsim$ as conditionals. There are three families of conditionals to consider, i.e., $\{\succsim_s\}_{s\in S}$, $\{\succsim_t\}_{t\in T}$ and $\{\succsim_{st}\}_{s\in S,t\in T}$. The last family represents *ex post* preferences, and the first two represent *interim* preferences, since each relation in these families depends on fixing one variable and letting the other vary, which amounts to resolving only part of the uncertainty.

We now formally define how these conditionals are obtained from their master relation $\succsim$. For example, $\succsim_s$, the *conditional of* $\succsim$ *on* $s \in S$, is the relation

[2] These two assumptions are only for mathematical simplicity. The theorem below could be proved for smaller domains than $\mathbb{R}$ and $\mathbb{R}^{S\times T}$, so as to pay attention to *feasibility constraints* on what counts as a consequence and what counts as a prospect.

$\succsim_s$ on $\mathbb{R}^T$ defined by the property that for all $\mathbf{x}_s, \mathbf{y}_s \in \mathbb{R}^T$,

$$\begin{aligned} \mathbf{x}_s &\succsim_s \mathbf{y}_s \text{ iff } \mathbf{X} \succsim \mathbf{Y} \\ \text{for some } \mathbf{X}, \mathbf{Y} &\in \mathbb{R}^{S\times T} \text{ s.t. } \mathbf{x}_s \text{ is the } s\text{-row of } \mathbf{X},\ \mathbf{y}_s \text{ is the } s\text{-row of } \mathbf{Y}, \\ &\text{and } \mathbf{X} \text{ and } \mathbf{Y} \text{ are equal outside their } s\text{-row.} \end{aligned}$$

The *conditional of* $\succsim$ *on* $t \in T$ and the *conditional of* $\succsim$ *on* $(s,t) \in S \times T$ are defined in similar ways. Importantly, the definition of conditionals does not by itself make them weak orderings. By a well-known fact of decision theory, $\succsim_s$ is a weak ordering if and only if the choice of $\mathbf{X}, \mathbf{Y}$ in the definition of $\succeq_s$ is immaterial, or more precisely, if and only if $\mathbf{X} \succsim \mathbf{Y} \Longleftrightarrow \mathbf{X}' \succsim \mathbf{Y}'$ when $\mathbf{X}', \mathbf{Y}'$ also satisfy the condition stated for $\mathbf{X}, \mathbf{Y}$ in this definition. When this holds, $\succsim$ is said to be *weakly separable* in s. By another well-known fact, weak separability in a factor (or product of factors) is equivalent to the property that $\succsim$ is *increasing* with the conditional on this factor (or the conditionals on the product of factors). Thus, for all $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{S\times T}$, if $\mathbf{x}_s \succsim_s \mathbf{y}_s$ for all s, then $\mathbf{X} \succsim \mathbf{Y}$; and if moreover $\mathbf{x}_s \succ_s \mathbf{y}_s$ for some s, then $\mathbf{X} \succ \mathbf{Y}$.[3] Combining the two facts just stated, we see that conditions (i) and (iii) of the previous section can be fused into the single requirement that all $\succsim_s$ and all $\succsim_t$ are weak orderings.[4]

Since conditionals $\succsim_{st}$ compare real numbers, it makes sense to identify them with the natural order of these numbers. This assumes that they represent desirable quantities, be they money values, as in the producer example, or something else. Thus, as another maintained assumption, we require that for all $(s,t) \in S \times T$ and all $x_s^t, y_s^t \in \mathbb{R}$,

$$x_s^t \succsim_{st} y_s^t \text{ iff } x_s^t \geq y_s^t.$$

Since this equivalence turns the $\succsim_{st}$ into orderings, $\succsim$ is increasing with each of them, hence also with each entry x_s^t of $\mathbf{X}$.

Let us say that the conditionals $\succsim_s$ (resp. $\succsim_t$) are an *invariant family* if $\succsim_s = \succsim_{s'}$ for all $s, s' \in S$ (resp. $\succsim_t = \succsim_{t'}$ for all $t, t' \in T$). Such requirements capture condition (ii) of previous section. They are not needed for the $\succsim_{st}$, which are identical relations by construction.

We are now ready for the first representation theorem.

Theorem 1 *The following conditions are equivalent:*

- *The conditionals $\succsim_s$ and $\succsim_t$ are weak orderings for all $s \in S$ and all $t \in T$, and either family of conditionals is invariant.*
- *There are increasing, continuous function $u : \mathbb{R} \longrightarrow \mathbb{R}$, and strictly positive probability functions $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$, such that $\succeq$ is represented by*

[3] By $\succ$, $\succ_s$, $\succ^t$ and $\succ_{st}$ we mean the *strict* preference relation associated with the weak preference relations $\succsim$, $\succsim_s$, $\succsim^t$ and $\succsim_{st}$.

[4] For these definitions and basic facts, see Fishburn (1970) or Wakker (1989).

the function $V : \mathbb{R}^{S\times T} \longrightarrow \mathbb{R}$ *that computes the* $\mathbf{p}\otimes\mathbf{q}$*-expected value of* u*, i.e., by the function defined as follows: for all* $\mathbf{X} = [x_s^t]_{s\in S}^{t\in T}$ *,*

$$V(\mathbf{X}) \ := \ \sum_{s\in S}\sum_{t\in T} p_s\, q_t u(x_s^t).$$

In this format of EU representation, $\mathbf{p}$ *and* $\mathbf{q}$ *are unique, and* u *is unique up to positive affine transformations.*

The conclusions state both the EU formula and that the two sources of uncertainty satisfy SI, so this theorem extends Bayesian decision theory up to the point that, from the argument made in the introduction, it ought to have reached.

4 A second EU representation theorem involving stochastic independence

In Theorem 1, strong results follow from a short list of assumptions, undoubtedly a feature of mathematical elegance, but also a cause for conceptual dissatisfaction. Would it not be better to expand on the assumptions and separate those which are responsible for the existence of the EU representation from those which account for the SI property occurring in this representation? Such a disentangling would be the more justified since SI is an *optional* property of a probability measure, hence in need of a preference condition that should be detachable from those underlying the existence of this measure. However, the assumptions of Theorem 1 cannot be so divided, as the following argument shows. By taking the $\succsim_s$ and $\succsim_t$ to be merely orderings, not invariant orderings, one would get an additively separable representation that does not separate the utility and probability components of the added terms, unlike the EU representation. By taking only one of the two families to satisfy the ordering and invariance assumptions, one would get a representation that is only separable in that family and would be even more remote from the EU representation.[5]

As it turns out, however, we can obtain a relevant partitioning of assumptions if we enrich the decision-theoretic framewok beyond the present two-dimensional

[5] The additively separable representation of the first case reads as

$$\sum_{s\in S, t\in T} v_{st}(x_s^t),$$

with increasing and continuous $v_{st} : \mathbb{R} \to \mathbb{R}$, $s \in S, t \in T$. In the second case, if the assumptions only hold for the $\succsim_s$, the separable representation reads as

$$W(V_1(\mathbf{x}_1), ..., V_{|S|}(\mathbf{x}_{|S|})),$$

with increasing and continuous $W : \mathbb{R}^S \to \mathbb{R}$ and $V_s : \mathbb{R}^T \to \mathbb{R}$, $s \in S$. These conclusions follow from standard results in separability theory; see, e.g., Blackorby, Primont and Russell (1978).

stage. Let us suppose that the agent pays attention not only to the uncertainty dimensions s and t of the final consequences, but also to a third dimension i, so that these consequences are now represented by real numbers x^i_{st}. The added dimension can be thought of in several ways, like time, space or an omitted dimension of uncertainty. Each of these concrete suggestions can fit the motivating example: the added dimension would indicate when, where or under what further unknown circumstance the monetary proceeds of a farming policy accrue to the producer. We will return to the interpretation of the added dimension after stating Theorem 2.

By assumption, i takes its values in a finite set I with cardinality $|I| \geq 2$. Prospects are now defined as mappings from triples (s, t, i) to the real numbers, that is as three-dimensional arrays,

$$\mathbb{X} = \left[x^t_s\right]^{i\in I}_{s\in S, t\in T} \in \mathbb{R}^{S\times T\times I}.$$

These may be rewritten as vectors

$$\mathbb{X} = (\mathbf{X}_1, ..., \mathbf{X}_{|S|}),\ \mathbb{X} =(\mathbf{X}_1, ..., \mathbf{X}_{|T|}) \text{ or } \mathbb{X} = (\mathbf{X}^1, ..., \mathbf{X}^{|I|}),$$

the components of which are now matrix-valued, i.e., $\mathbf{X}_s = (x^i_{st})^{i\in I}_{t\in T} \in \mathbb{R}^{T\times I}$, $\mathbf{X}_t = (x^i_{st})^{i\in I}_{s\in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}\times I}$ and $\mathbf{X}^i = (x^i_{st})_{s\in\mathcal{S}, t\in T} \in \mathbb{R}^{\mathcal{S}\times T}$ respectively.

As before, the agent compares prospects in terms of a preference relation $\succsim$, which is a continuous weak ordering, and this relation gives rise to various families of conditional relations. The $\succsim_s$, $\succsim_t$ and $\succsim^i$ respectively compare matrices $\mathbf{X}_s$, $\mathbf{X}_t$, $\mathbf{X}^i$, as defined above, the $\succsim_{st}$ compare vectors $\mathbf{x}_{st} \in \mathbb{R}^I$, and the $\succsim^i_{st}$ compare real numbers. Similarly as before, we assume that each $\succsim^i_{st}$ coincides with the natural order of real numbers, which makes it an ordering, and furthermore turns the $\succsim^i_{st}$ into an invariant family. The other conditional relations may or may not be weak orderings, and may or may not form invariant families, depending on which assumptions are put on them.

Theorem 2 *The following conditions are equivalent:*

- *The conditionals $\succsim^i$ are weak orderings.*
- *There are increasing, continuous functions $u^i : \mathbb{R} \longrightarrow \mathbb{R}$, for all $i \in I$, and a strictly positive probability function $\boldsymbol{\pi} \in \Delta_{S\times T}$, such that $\succeq$ is represented by the function $W : \mathbb{R}^{S\times T\times I} \longrightarrow \mathbb{R}$ that computes the $\boldsymbol{\pi}$-expected value of $\sum_{i\in I} u^i$, i.e., the function thus defined: for all $\mathbb{X} = [x^t_s]^{i\in I}_{s\in S, t\in T}$,*

$$(*)\ W(\mathbb{X}) \ := \ \sum_{s\in S, t\in T} \sum_{i\in I} \pi_{st}\, u^i(x^i_{st}).$$

In this format of representation, $\boldsymbol{\pi}$ is unique, and the u^i are unique up to positive affine transformations with a common multiplier.

Moreover, the following are equivalent:

- *The same assumption on the $\succsim^i$ holds, and the $\succsim_s$ are weak orderings and an invariant family.*

- *The same conclusions hold, and there are strictly positive probability functions $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$ with $\pi = \mathbf{p} \otimes \mathbf{q}$, so that $(*)$ becomes: for all $\mathbb{X} = [x_s^t]_{s \in S, t \in T}^{i \in I}$,*

$$(**) \; W(\mathbb{X}) = \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s \, q_t u^i(x_{st}^i).$$

In this format of representation, $\mathbf{p}$ and $\mathbf{q}$ are unique, while the u^i have the same uniqueness properties as before.

Unlike Theorem 1, Theorem 2 is in two parts, corresponding to the EU formula and the SI property specifically. What appears to be essential to the latter is that one of the two sources (here conventionally taken to be S) gives rise to an *invariant* family of conditionals. We now reinforce the suggestion that invariance is the crucial condition by a heuristic argument.

Considering for simplicity only four states, suppose that the agent takes (s_1, t_1) to be more likely than (s_1, t_2), and (s_2, t_1) less likely than (s_2, t_2). That is, from knowing how the uncertainty on s is resolved, the agent is able to draw an inference on how the uncertainty on t would be resolved. If the agent reasoned probabilistically, the joint probabilities would of course *not* decompose multiplicatively. It is easy to conclude that the conditionals on s cannot be invariant. Take ξ, ξ' representing desirable quantities, with $\xi > \xi'$, and the following prospects in matrix form:

$$\begin{array}{ccc} \mathbf{X} & t_1 & t_2 \\ s_1 & \xi & \xi' \\ s_2 & \xi' & \xi \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathbf{Y} & t_1 & t_2 \\ s_1 & \xi' & \xi \\ s_2 & \xi & \xi' \end{array} .$$

The first line of $\mathbf{X}$, which puts the best consequence on the more likely state, should be preferred to the first line of $\mathbf{Y}$, which puts it on the less likely state; that is, $(\xi, \xi') \succ_{s_1} (\xi', \xi)$. By a similar comparison, the second line of $\mathbf{X}$ should be preferred to the second line of $\mathbf{Y}$; that is $(\xi', \xi) \succ_{s_2} (\xi, \xi')$. Thus, the two conditional preferences differ. Contraposing the argument, we see that, if the $\succsim_s$ are an invariant family, then, were the agent to know the particular s, he would not use this knowledge to draw information on the unknown t. The converse also holds: an agent who is in such an epistemic disposition has no reason for entertaining conditional preferences $\succsim_s$ that are variable with s. This argument connects the invariance property of conditional preferences with the informational rendering of SI mentioned in the introduction.

To derive $(**)$, it is unnecessary to assume that both the $\succsim_s$ and the $\succsim_t$ are invariant. The invariance of the latter relations follows from $(**)$ itself.[6] If it is sufficient to require one direction of invariance, this is because the EU

[6] From $(**)$, the $\succsim_{t \in T}$ are represented by $\sum_{s \in S} \sum_{i \in I} p_s \, u^i(.)$, which does not depend on t.

representation $(*)$ holds from the previous stage. In an EU framework, it is impossible to distinguish between s bringing no preferentially relevant information on t, and t bringing no preferentially relevant information on s. This is formally shown in the appendix.

We now return to the interpretation of the third dimension i introduced in this section. A very natural decision-theoretic account becomes available when i represents time. Then, the alternatives $\mathbb{X}$ mean *contingent plans*, i.e., plans for the future whose consequences in a given period depend on the way the uncertainty - still represented by (s,t) - is resolved in that period. The matrix-valued objects $\mathbf{X}_s$, $\mathbf{X}_t$ and $\mathbf{X}^i$ mean *partly contingent plans* (for the first two, when one dimension of uncertainty is fixed) or *dated prospects* (for the last, when the time dimension is fixed). As to the vector-valued objects $\mathbf{x}_{st}$, they mean *non-contingent plans*, since they take the uncertainty to be entirely resolved.[7]

However, time considerations are extraneous to uncertainty, which is the concern here, and it may be more appropriate to find an interpretation for i that relates to these concerns. Suppose we declare i to be *a third dimension of uncertainty.* We can then add a third part to Theorem 2, in which puts on the $\succsim^i$ the same invariance assumption as was imposed on the $\succsim_{st}$ and the $\succsim_t$. From this addition, it can be proved that $(**)$ gives way to the following more specific representation: for all $\mathbb{X} = [x_s^t]_{s\in S, t\in T}^{i\in I}$,

$$(***)\ W(\mathbb{X}) = \sum_{s\in S}\sum_{t\in T}\sum_{i\in I} p_s\, q_t r_i u(x_{st}^i),$$

where $\mathbf{r} = (r_i)_{i\in I}$ is a strictly positive probability function on I, and the utility function u in the EU formula is now independent of the i index. Besides having a semantic advantage, this extension of Theorem 2 carries with it a sense of mathematical generalization. To obtain the SI property for a product space Ω with *any finite number of uncertainty factors* is no more difficult than to obtain it for $\Omega = S \times T \times I$, but this would impose a heavy notational burden.

5 Comparisons with decision theory

We start this decision-theoretic section with two comments that Savage makes on SI in his *Foundations of Statistics*. Having axiomatized a qualitative probability relation, he complains that "the notions of independence and irrelevance have ... no analogues in qualitative probability; this is surprising and unfortunate, for these notions seem to evoke a strong intuitive response" (1954-1972, p. 44). Later, at the end of a well-known passage on "small worlds", Savage restates his complaint as follows: "it would be desirable, if possible, to find a simple qualitative personal description of independence between events" (p. 91).[8] The two comments clearly express the need for Bayesian decision theory to

[7] These interpretations assume that each period is uncertain in the same way as any other, i.e., no interaction exists between the resolution of uncertainty and the passing of time.

[8] Savage used to say "personal probability" where later theorists say "subjective probability".

complement its derivation of subjective probability with an account of SI, but point in different directions. The first relates to *qualitative probability*, which is an auxiliary concept in Savage's construction; he uses it as an intermediary between his preference postulates and his final conclusions, in which subjective probability acquires a numerical form. Today, Savage's remark in respect of this concept is no longer justified. There now exist richer systems of qualitative probability than his, which contain a special relation to express the stochastic independence of two events or two random variables.[9] The second comment does not mention qualitative probability and we read in it a suggestion to base SI directly on a preference foundation. In this respect, Savage's complaint of a lacuna is still essentially justified. To the best of our knowledge, there are only three earlier works in decision theory that overlap with the present research.

Blume, Brandenburger and Dekel (1991, p. 74) introduce a preference condition that is akin to our invariance condition and heuristically stress its connection with SI, but do not include it in their representation theorems. Their topic is anyhow the preference foundations of lexicographic probability, not of standard Kolmogorov probability.[10] In an important follow up, Battigalli and Veronesi (1996) push the analysis of Blume, Brandenburger and Dekel up to the stage of representation theorems, but these are also concerned with lexicographic probability or related non-standard notions. Neither Theorem 1 nor Theorem 2 appear in these two works. More directly related is the version of Bayesian decision theory proposed by Bernardo, Ferrandiz and Smith (1985). This includes a preference condition relative to two events E and F that will entail the equation $P(E \cap F) = P(E).P(F)$ at the stage of proving the EU representation theorem. Although evocative of the informational reading of SI, this condition differs from ours, and this difference seems connected with the authors' technical choice of approaching SI in general probability spaces rather than product spaces, as we do here.[11]

We now compare our axiomatization of the EU with those in current use. Being entirely preference-based, the former is like Savage's (1954-1972), but there are important dissimilarities. An obvious one concerns the cardinality of the state set Ω, which is infinite in Savage and finite here. The axiom systems in Theorem 1 and Theorem 2 are highly condensed and do not relate to Savage's seven-postulate system in a transparent way. However, the assumption that

[9] See Domotor (1969), Fine (1971), Kaplan and Fine (1977), Luce and Nahrens (1978), to cite only the earliest accounts of SI in terms of qualitative probability. Fine's 1971 classic, *Theories of Probabilities*, makes interesting comments on SI, and at some point (p. 36-37) even suggests moving in the direction of a pragmatic, preference-based account of SI.

[10] Generally, a lexicographic probability is a finite sequence $(p_1, ..., p_K)$ of probability measures on the same probability space. In the more specific definition privileged by Blume, Brandenburger and Dekel (1991), the supports of the successive p_i are disjoint.

[11] As a further development of Joyce's (1999) representation theorem for pairs of credibility and preference relations, Bradley (2017, p. 104) shows that a weak separability assumption on the credibility relation imposes the SI property on the probability measure representing this relation. Joyce's and Bradley's analyses belong to Jeffrey's (1965) theory of decision, which is several steps removed from the Bayesian decision theory of this paper.

certain conditionals are orderings amounts to replacing his postulate P2 - the "sure-thing principle"- by a dominance principle, which is weaker and more generally accepted. To make good for this loss, the invariance condition, when it applies at all, bears on all possible prospects and not only on constant prospects, as is the case in Savage's P3 - the "event independence" postulate. Savage has another important postulate, P4, which is a clear step towards the existence of subjective probability and has no analogue here. Our best guess is that P4 is made dispensable by the assumption that consequences are real numbers and conditional preferences respect the order of these numbers. By contrast, Savage puts no restriction at all on his consequence set.

Another comparison to the point is with Anscombe and Aumann's (1963) popular variation on Savage's system. We share with these authors the assumptions of a finite state set and a highly structured consequence set, but they assume the latter to be *a set of probabilistically defined lotteries*, an assumption we are glad to eschew here. From a Bayesian decision theory perspective, the Anscombe-Aumann system is open to the objection that it is question-begging to derive a subjective probability by supposing that other probabilities already exist. From the perspective of Bayesian decision theory, all probabilistic items require a preference derivation. The rejoinder that the preexisting probabilities are objective, hence of a different nature from the subjective probability to be derived, is a free commentary without any basis in Anscombe and Aumann's formal system. We do not deny the pratical convenience of this system, but ours is no more complicated, while being perhaps easier to defend theoretically.[12]

6 Connections with foundational discussions

Underlying the axiomatic work of Savage and Bayesian decision theorists generally are two major claims on the foundations of probability: probability measures represent uncertain beliefs in the normatively appropriate way, and what makes the measures in question normatively appropriate is that pratical rationality considerations recommend using them in decision making. Both claims have been disputed, with some objections surfacing already before Bayesian decision theory fully took shape. The first claim can be attacked along at least two different lines. One may question the appropriateness of probabilities on the ground that they are *absolute* measures, and as an alternative develop a calculus for conditional probabilities taken as primitive terms. This line is historically associated with Popper's (1959-1972, Appendices *iv and *v) axiomatization of probability and has recently been defended by Fitelson and Hajek (forthcoming). The existing conditional probability systems preserve the *additivity* of probability measures, and an alternative critical line is indeed to question that property. Decision theory has made thorough contributions here; see Gilboa

[12] For convenience, both Blume, Brandenburger and Dekel (1991) and Battigalli and Veronesi (1996) use the Anscombe-Aumann system. There are other alternatives to this system than the present one for applications to finite state sets. An early example is Wakker's (1989, ch. IV).

(2009) and Wakker (2010) for overviews. As to the second foundational claim, it can again be attacked from different sides, one representative example being Joyce's (1998) "nonpragmatic" argument that probabilities are appropriate representations of uncertain beliefs for directly *epistemic* reasons.[13]

These deep foundational questions arise in connection with the present work, but exceed its limited purpose. We meant to fill a gap in Bayesian decision theory by following its own principles, rather than defend these principles against outside criticism. However, since SI is our focus, we should ask whether this theory, as extended here, may contribute to a better understanding of this property.

There is much conceptual tension in the way probability theorists introduce the definition of SI. For one thing, they usually discuss its informal meaning in terms of a provisional definition of SI by the equality of conditional and unconditional probability: for any two events $A, B \in \mathcal{A}$ with $P(B) > 0$,

$$P(A \mid B) = P(A).$$

Once they have made intuitive sense of this equation, they proceed to the multiplicative equation of section 2 as constituting the proper definition of SI, arguing that the latter avoids the sign restriction $P(B) > 0$ and makes SI symmetric. This argument is unconnected with the intuitions supporting the provisional definition, which makes the whole sequence semantically awkward.[14] For another thing, probability theorists informally defend their definitions by resorting to more than one concept of unrelatedness. Prominent examples are logical independence, causal independence (or alternative forms of objective independence), and informational independence. While some accounts are relatively clear on which concept they privilege, many others are equivocal, and some even fall into amazing confusions between them.[15]

The Bayesian decision theory developed here contributes nothing to the first problem because it does not allow for 0-probability events, and thus cannot properly distinguish between the provisional and final definition of SI. This limitation is a price to pay for a handy mathematical apparatus (more on this in the appendix). On the second problem, however, the theory has something to say. At the very least, it avoids the equivocations between the different informal accounts by firmly opting for *informational independence*; the preference condition that Theorem 2 highlights, the invariance property of conditionals, is unquestionably on the pragmatic side of the foundations of probability. The agent's unwillingness to adapt his preferences over t-uncertain prospects to the knowledge of s is the criterion by which we judge that this agent regards information on s as being irrelevant to t.

It remains to be said whether the theory contents itself with endorsing one of the available accounts of SI or adds something significant to that account.

[13] Leitgeb and Pettigrew (2010) have recently pursued this line of purely epistemic justification with a new derivation of probability from an accuracy requirement.

[14] Two examples among many of this two-step definitional sequence are Feller (1950-1968, p. 125) and Hoel, Port and Stone (1971, p. 19).

[15] Here is a curious example due to two otherwise excellent scholars: SI means that "the knowledge of (one event) does not affect the other" (Luce and Narens, 1978, p. 226). Naturally, one would expect "the knowledge of the other" instead of "the other".

We credit the theory with two possible contributions. One is to connect the SI property with the foundations of subjective probability more tightly than is usually done. There has been some vacillation among subjectivists concerning the role of SI assumptions in probability theory. Whereas Savage did not underplay this role, de Finetti considered it with strong reluctance. As Gillies (2000, p. 75-76) explains, citing from *Probabilismo* (1931), de Finetti argued against the application of SI to repeated trials of the same experiment on the ground that this assumption blocked the possibility of learning from the successive results of the trials through Bayes's rule of revision. This argument opened the way to de Finetti's alternative to SI, which is exchangeability. Without entering the rich debate - well covered by Gillies - on the respective merits of the two concepts, we can make the broader point that learning by Bayes's rule is just one particular case to be considered by the subjective theory of probability. It is possible to make perfect subjective sense of the opposite particular case in which no learning occurs; it is actually incumbent on the subject to decide which case is relevant. In other words, there is no logical necessity to associate the subjective theory with a large scope of application of Bayes's rule. Although this point may be clear by itself, it comes out perhaps more clearly after Bayesian decision theory, which is a brand of subjectivism, has offered an account of SI.

Another contribution is to put the *symmetry* of the definition of SI in perspective. Writers on the foundations of probability have sometimes expressed dissatisfaction towards the fact that asymmetric dependence or independence cannot be formulated within the Kolmogorov axiomatic framework; see Fitelson and Hajek (forthcoming) for a recent example. This is a fair complaint to make in connection with the logical and causal (or more generally objective) readings of SI, but its force as to the informational reading is not so clear, as Fitelson and Hajek concede. In the Bayesian decision theory of this paper, one can assume the s-component of uncertainty to be informally irrelevant to the component t without assuming the converse irrelevance, for this amounts to requiring invariance from the s-conditionals and not from the t-conditionals. However, we have seen that this logical independence of the two assumptions vanishes once the preference axioms for EU are all in place. This result can be understood in two opposite ways - those who take the preference conditions for EU to be normatively compelling will view it as a justification of the postulated symmetry of SI, whereas others, for whom this symmetry is an arbitrary diktat, will turn the result against the allegedly compelling preference conditions.

7 Conclusions

We have responded to Savage's request to extend the preference apparatus of Bayesian decision theory to the point where it includes an account of stochastic independence. To do so, we have reconstructed this preference apparatus and proved representation theorems that contain both a novel derivation of the expected utility formula and the desired specification that the subjective probability in this formula makes the sources of uncertainty stochastically independent. These theorems call for richer variants that need to be pursued elsewhere. One such variant would relax the finiteness assumption put on the set of states of the world and consistently offer the treatment of 0-probability events, the lack of which had previously been made tolerable by this finiteness assumption. Besides absolute probability as in Kolmogorov, this line of research could more ambitiously concern conditional probability taken as an axiomatic primitive, in Popper's sense. Each time, the objective would be to map the features of the probability space onto axiomatic preference counterparts. Another project would be to reconsider stochastic independence in relation to the non-additive measures of uncertainty that decision theorists have introduced since they moved away from a primarily Bayesian outlook. This is the more challenging of the two lines of research, because it requires one not only to find preference counterparts to already defined mathematical features, but also to discover those new mathematical definitions which capture stochastic independence when probability gives way to weaker notions.

8 Appendix

The two theorems of this paper follow from a result proved by Mongin and Pivato (2015, Theorem 1). We restate this result in a simplified form adapted to the purpose of deriving them.

Theorem 3 *Granting the maintained assumptions of section 3 on* $\succsim$ *and the conditionals* $\succsim_{st}$*, the following conditions are equivalent:*

- *The conditionals* $\succsim_s$ *and* $\succsim_t$ *are weak orderings for all* $s \in S$ *and all* $t \in T$*, and the* $\succsim_s$ *are an invariant family.*
- *There are increasing, continuous functions* $u^t : \mathbb{R} \longrightarrow \mathbb{R}$ *for all* $t \in T$*, and there is a strictly positive probability function* $\mathbf{p} \in \Delta_S$*, such that* $\succeq$ *is represented by the function* $V : \mathbb{R}^{S\times T} \longrightarrow \mathbb{R}$ *that computes the* $\mathbf{p}$*-expected value of* $\sum_{t\in T} u^t$*, i.e., by the following function: for all* $\mathbf{X} = [x_s^t]_{s\in S}^{t\in T}$ *,*

$$V(\mathbf{X}) \ := \ \sum_{s\in S}\sum_{t\in T} p_s \, u^t(x_s^t).$$

In this format of EU representation, $\mathbf{p}$ *is unique, and the* u^t *are unique up to positive affine transformations with a common multiplier.*

Theorem 1 requires *both* families of $\succsim_s$ and $\succsim_t$ to be invariant. A proof for it results from applying Theorem 3 twice and checking the compatibility of the obtained representations. See Mongin and Pivato (2015, Corollary 1(c)) for mathematical details.

The first part of Theorem 2 is a direct application of Theorem 3, with i playing the role of s and (s,t) playing the role of t in the statement of the latter. The second part is proved below.

Proof. (Sketch). We first observe that, for every given $s \in S$, the $W(\mathbb{X})$ representation of the first part delivers a function $\mathbb{R}^{T\times I} \to \mathbb{R}$

$$\sum_{t\in T}\sum_{i\in I}\pi_{st}u^i(x^i_{st})$$

that represents the weak ordering $\succsim_s$. If we define $\pi'_{st} := \pi_{st}/\sum_{t\in T}\pi_{st}$ for all $t \in T$, the function

$$\sum_{t\in T}\sum_{i\in I}\pi'_{st}u^i(x^i_{st})$$

is also a representation of $\succsim_s$. Now fix $s_0 \in S$. By the invariance of the $\succsim_s$ family, for every $s \in S$, there is a strictly increasing function Φ_s on $\mathbb{R}$ such that

$$\sum_{t\in T}\sum_{i\in I}\pi'_{s_0t}\,u^i(x^i_{st}) = \Phi_s\left(\sum_{t\in T}\sum_{i\in I}\pi'_{st}u^i(x^i_{st})\right).$$

As the u^i are strictly increasing and continuous, and so are the double sums of them, the Φ_s are continuous, and we can apply a functional equation argument and conclude that the Φ_s are positive affine transformations. I.e., for all $s \in S$, there exist numbers $\alpha_s > 0$ and β_s s.t.

$$\sum_{t\in T}\sum_{i\in I}\pi'_{s_0t}\,u^i(x^i_{st}) = \alpha_s\left(\sum_{t\in T}\sum_{i\in I}\pi'_{st}u^i(x^i_{st})\right) + \beta_s.$$ [16]

After redefining the functions so as to dispense with the constant terms, we see that, for all $s \in S$ and $t \in T$, $\pi'_{s_0t} = \alpha_s\pi'_{st}$, and in fact (since proportional probability vectors are equal) $\pi'_{s_0t} = \pi'_{st}$. We thus rewrite $(*)$ as

$$\sum_{s\in\mathcal{S},t\in T}\sum_{i\in I}\pi'_{s_0t}\,(\sum_{t\in T}\pi_{st})(u^i(x^i_{st}),$$

which is $(**)$ if one takes $\mathbf{p} = (\sum_{t\in T}\pi_{st})_{s\in S}$ and $\mathbf{q} = (\pi'_{s_ot})_{t\in T}$. The uniqueness of $\mathbf{p}$ and $\mathbf{q}$ in this format of representation is easily established.

To show that adding an invariance assumption on the $\succeq^i$ leads to the stronger representation claimed in the text, i.e.,

$$(***)\ W(\mathbb{X}) = \sum_{s\in S}\sum_{t\in T}\sum_{i\in I}p_s\,q_t r_i u(x^i_{st}),$$

it is enough to reproduce the proof sequence used for $(**)$ *mutatis mutandis.* ■

Theorems 1, 2 and 3 all involve *strictly positive* probability functions. This restriction is due to the assumption that the $\succsim_{st}$ (in Theorems 1 and 3) and the $\succsim^i_{st}$ (in Theorem 3) reproduce the natural order of real numbers. It can be checked that this makes the $\succsim_s$, $\succsim_t$ and $\succsim^i$ non-constant preference relations, so there are no "null events" in Savage's (1954-1972, p.24) sense, hence no 0-probability values either. That the mathematics of this paper does not handle these values is a limited shortcoming given that the state set is taken to be finite and there is a single decision-maker to consider. The classic decision-theoretic move in this case is to prune the state set of the states the conditionals of which are constant.[17]

9 References

Anscombe, F.J., & R.J. Aumann (1963), "A Definition of Subjective Probability", *Annals of Mathematical Statistics*, 34, p. 199-205.

Battigalli, P. & P. Veronesi (1996), "A Note on Stochastic Independence without Savage-Null Events", *Journal of Economic Theory*, 70, p. 235-248.

Bernardo, J.M., J.R. Fernandez & A.F.M. Smith (1985), "The Foundations of Decision Theory: An Intuitive, Operational Approach with Mathematical Extensions", *Theory and Decision*, 19, p. 127-150.

Blackorby, C., D. Primont & R.R. Russell (1978), *Duality, Separability and Functional Structures: Theory and Applications*, New York, North Holland.

Blume, L., A. Brandenburger & E. Dekel (1991), "Lexicographic Probability and Choice Under Uncertainty", *Econometrica*, 59, p. 61-79.

Bradley, R. (2017), *Decision Theory with a Human Face*, Cambridge, Cambridge University Press.

de Finetti, B. (1931), *Probabilismo*, Biblioteca filosofica, Perrella, Napoli, p.163-219. Reprinted in B. de Finetti, *La logica dell'incerto*, Milano, Il Saggiatore, 1989, p. 3–70. English tr. as "Probabilism", *Erkenntnis*, 31, p. 169–223, 1989.

Domotor, Z. (1969), "Probabilistic Relational Structures and Their Applications", Technical Report No. 144, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Feller, W. (1950), *An Introduction to Probability Theory and Its Applications*, New York, Wiley (3d ed. 1968).

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

Fishburn, P.C. (1970), *Utility Theory For Decision Making*, New York, Wiley.

Fitelson, B. & A. Hájek (forthcoming), "Declarations of Independence", *Synthese*.

Fleurbaey, M. & P. Mongin (2016), "The Utilitarian Relevance of the Aggregation Theorem", *American Economic Journal: Microeconomics*, 8, p. 289–306.

Gilboa, I. (2009), *Theory of Decision Under Uncertainty*, Cambridge, Cambridge University Press.

Gillies, D. (2000), *Philosophical Theories of Probability*, London, Routledge.

Halmos, P.R. (1974), *Measure Theory*, New York, Springer.

Hoel, P.G., Port, S.C. & C.J. Stone (1971), *Introduction to Probability Theory*, Boston, Houghton Mifflin Company.

[17] Even under finiteness assumptions, this pruning move is not available in a game-theoretic context, since 0 probability moves can be invested there with strategic significance.

Joyce, J.M. (1998), "A Non-Pragmatic Vindication of Probabilism", *Philosophy of Science* 65, p. 575-603.

Joyce, J.M. (1999), *The Foundations of Causal Decision Theory*, Cambridge, Cambridge University Press.

Jeffrey, R. (1965), *The Logic of Decision*, Chicago, Chicago University Press (2nd ed., 1983).

Joyce, J.M. (1999), *The Foundations of Causal Decision Theory*, Cambridge, Cambridge University Press.

Kaplan, M. & T.L. Fine (1977), "Joint Orders in Comparative Probability", *Annals of Probability*, 5, p. 161–179.

Kolmogorov, A.N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, in *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 3. Engl. tr. as *Foundations of the Theory of Probability*, New York, Chelsea (2nd ed., 1956).

Leitgeb, H. & R. Pettigrew (2010), "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy", *Philosophy of Science*, 77, p. 236-272

Luce, R.D. & L. Narens (1978), "Qualitative Independence in Probability Theory", *Theory and Decision*, 9, p. 225–239.

Mongin, P., & M. Pivato (2015), "Ranking Multidimensional Alternatives and Uncertain Prospects", *Journal of Economic Theory*, 157, p. 146-171.

Mongin, P. & M. Pivato (2016), "Social Preference Under Twofold Uncertainty", HEC Paris Research Paper Paper No. ECO/SCD-2016-1154. https://ssrn.com/abstract=2796560

Popper, K.R. (1959), *The Logic of Scientific Discovery*, London, Hutchinson (6th revised ed., 1972).

Rado, F. & J. Baker (1987), "Pexider's Equation and Aggregation of Allocations", *Aequationes Mathematicae*, 32, p. 227-239.

Savage, L.J. (1954), *The Foundations of Statistics*, Wiley, New York (2nd revised ed., 1972).

Wakker, P. (1989), *Additive Representations of Preferences*, Dordrecht, Kluwer.

Wakker, P. (2010), *Prospect Theory*, Cambridge, Cambridge University Press.