

Bayesian Decision Theory and Stochastic Independence

Philippe Mongin

CNRS & HEC Paris

mongin@greg-hec.com

Stochastic independence 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. Bayesian decision theorists such as Savage can be criticized for being silent about stochastic independence. From their current preference axioms, they can derive no more than the definitional properties of a probability measure. In a new framework of twofold uncertainty, we introduce preference axioms that entail not only these definitional properties, but also the stochastic independence of the two sources of uncertainty. This goes some way towards filling a curious lacuna in Bayesian decision theory.

1 Introduction and preview

The property of stochastic (or statistical) independence occupies a rather special place in the mathematical theory of probability. It does not belong to the properties that this theory singles out to define a probability measure axiomatically. It is indeed a property of given events for a given probability measure, and its adoption can only result from a modelling choice to fit the particular situation. At the same time, probability theory obviously uses independence assumptions extensively; they are needed for such major results as the Laws of Large Numbers, various theorems on stochastic processes, and some central results of statistical theory. For Kolmogorov [14] himself, the inventor of the axiomatic definition, this property occupies a "central position in the theory of probability" (1933-1950, p. 8). One would thus expect all theories of the foundations of probability to pay careful attention to stochastic independence, but curiously, this is not the case with Bayesian decision theory, one of the most influential among these theories.

Bayesian decision theorists claim that an agent's uncertain beliefs should be represented by a probability measure and ground this claim on a pragmatic argument. They formally show 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, and in particular have nothing to add on stochastic independence. Thus, they can be criticized for falling short of justifying the probability calculus as it actually works and stopping too early in their foundational work.

More technically, Bayesian decision theorists prove 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 detail by Savage (1954) [19]. In a subsequent simplification of Savage's system, Anscombe and Aumann (1963) [1] took some probability values for granted in order to obtain the remaining ones more easily. All these authors derive a *prior* probability measure to represent initially uncertain beliefs. Savage (1954-1972, p. 44) extends this argument to obtain a *posterior* probability measure, i.e., one that represents beliefs after a partial resolution of uncertainty, and he shows that this posterior obeys Bayes's rule of

revision; literally, the "Bayesian" label becomes fully justified only at this stage. This is also where Savage stops. He however acknowledges that a treatment of stochastic independence should have come next.

Two comments in Savage make this point clearly. 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, he reiterates the complaint differently: "it would be desirable, if possible, to find a simple qualitative personal description of independence between events" (p. 91). (Savage prefers the expression of "personal probability" to the more received one of "subjective probability".) In today's Bayesian theory, the first comment is not justified anymore. There now exist richer systems of qualitative probability than Savage's, in which a special relation serves to express the stochastic independence of two events or two random variables (see Domotor, 1969 [5], Fine, 1971 [6], Kaplan and Fine, 1977 [12], Luce and Nahrens, 1978 [15], to cite but the early papers). However, the second comment is still topical. We understand it as referring to preferences over uncertain prospects, i.e., the ultimate primitive in Savage's construction. To the best of our knowledge, Bayesian decision theory has not yet explicated stochastic independence in terms of this overarching concept.

The present paper is an attempt to do so. We assume that there are *two distinctive sources of uncertainty*, and accordingly that states of nature have the form of two-component vectors. As in Savage and in Anscombe and Aumann, we define uncertain prospects to be mappings from states of the world to consequences, and take the agent's preferences over these prospects to be the only axiomatic primitive. Our construction leans towards Anscombe and Aumann by considering a finite number of states and by making a structural assumption on consequences (they are real numbers). However, it also leans towards Savage because we eschew any numerical data, hence Anscombe and Aumann's questionable trick of taking some probability values for granted. The twofold uncertainty framework has recently been introduced by Mongin and Pivato (2016) [17] with a theoretical purpose different from the present one.¹

Our axioms entail that there exists an EU representation for the agent's preferences, and that the probability measure in this representation decomposes *multiplicatively* on the two sources of uncertainty, which establishes their stochastic independence. The same uniqueness conditions hold as in standard EU representation theorems. A heuristic argument indicates where to locate the stochastic independence property in our preference axioms. They state that, for either source of uncertainty, there exist preferences conditional on each value this source can take, and moreover that these conditional preferences are invariant across possible values. This heuristically means that the realization of one of the two uncertainty components does not affect the agent's preferences over those prospects which only depend on the other, still unknown component. In a betting interpretation, if the initial bets relate, say, to tomorrow's weather and tomorrow's economic conditions, and the agent somehow comes to know what tomorrow's weather will be, this does not affect the agent's preferences amongst bets on tomorrow's economic conditions, and vice-versa.

We offer two representation theorems along the lines just explained. The first adapts a result already obtained in Mongin and Pivato (2015) [16]. We develop it here because it neatly exemplifies how Bayesian decision theory can approach stochastic independence. Although this theorem implements the heuristics of last paragraph, it is, in a subtle sense to be explained, not yet entirely satisfactory. Hence we propose a second representation theorem, which is the technical novelty of this paper. Section 2 adds further motivations. Sections 3 and 4 state the two representation theorems respectively. Section 5 returns

¹That paper aims at offering a solution to the classic problem of defining a normatively compelling notion of social preference under uncertainty.

to conceptual comments and comparisons, while also sketching directions for future work.

2 Further motivating the approach

A corn producer must decide how much land to cultivate while not knowing what the climatic conditions and the state of demand for corn will be at the time of the harvest. Each cultivation policy can be analyzed as an uncertain prospect, i.e., a mapping from the unknown states to the possible consequences, here monetary proceeds. We will develop this example along a Bayesian theorist's line, and heuristically reason backwards, taking for granted what the Bayesian theorist would conclude, plus the target property of stochastic independence. Recall the textbook definition since Kolmogorov: given a probability space (Ω, \mathcal{A}, P) , two events A, B in \mathcal{A} are said to be stochastically (or statistically) independent if $P(A \cap B) = P(A) \cdot P(B)$. From this definition, others, which are equally standard, follow concerning collections of events or random variables.

Assuming for simplicity that climate and demand take two values, we fix two sets $S = \{s_1, s_2\}$ and $T = \{t_1, t_2\}$ and define a state of nature to be any element of the product set $S \times T$. We now apply stochastic independence to each possible pair of events (subsets) $\{s\} \times T$ and $S \times \{t\}$, or by an obvious identification, to each possible pair (s, t) . The producer's probabilities are thus given by the matrix:

$$\begin{array}{cc} & \begin{array}{cc} t_1 & t_2 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \end{array} & \begin{array}{cc} p_{s_1}q_{t_1} & p_{s_1}q_{t_2} \\ p_{s_2}q_{t_1} & p_{s_2}q_{t_2} \end{array} \end{array}$$

where (p_{s_1}, p_{s_2}) and (q_{t_1}, q_{t_2}) are probability vectors on S and T , respectively. Now, a policy \mathbf{X} for the producer can be represented by its monetary proceeds in the various states:

$$\mathbf{X} \quad \begin{array}{cc} t_1 & t_2 \\ s_1 & x_{11} \quad x_{12} \\ s_2 & x_{21} \quad x_{22} \end{array}$$

Denote by \succsim the producer's preferences over policies \mathbf{X} . The EU representation for \succsim is

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

This can be restated either as:

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

or as:

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

The bracketed sums in $(*)$ contain utility representations for conditional preferences on the possible values of s , and those in $(**)$ contain utility representations for conditional preferences on the possible values of t . Thus, the overall conclusions entail that (i) *conditional preferences are orderings*. Since the same functional form $q_{t_1}u(\cdot) + q_{t_2}u(\cdot)$ appears in the two bracketed sums of $(*)$, and similarly, the same functional form $p_{s_1}u(\cdot) + p_{s_2}u(\cdot)$ appears in the two bracketed sums of $(**)$, these conclusions 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, the conclusions also entail that (iii) *preferences over prospects are increasing with respect to either family of conditional preferences*.

Importantly, we have stated (i), (ii) and (iii) by abstracting from the EU representation. Each of these properties can indeed be satisfied by more general theories than Bayesian decision theory, and in particular, the *dominance* property (iii) is well-known to apply to most existing alternatives (like rank-dependent theory, see, e.g., Wakker, 2010 [21]).

In the first result, we assume (i), (ii) and (iii), plus some background conditions. Given the formal 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 apparently weak necessary conditions for the representation turn out also to be sufficient for it. The key point is that the conditions 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; comparing these representations leads to the results. Their equivalence shows in the fact that either the p_s or the q_t can be factored out from the same sum - see (*) and (**).

3 A first representation theorem for 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 $(s, t) \in S \times T$; we thus permit the two variables to vary together in any possible way. For technical reasons, we take S and T to be finite with cardinalities $|S|, |T| \geq 2$. Prospects \mathbf{X} are mappings from states (s, t) to real numbers x , and we define the set of prospects to be $\mathbb{R}^{S \times T}$, thus putting no constraint on what counts as a prospect. 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 $\mathbf{X} = [x_{st}]_{s \in S}^{t \in T}$, but sometimes also $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{|S|})$, where each component is a row vector $\mathbf{x}_s \in \mathbb{R}^T$, or $\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^{|T|})$, 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, and this amounts to resolving only part of the uncertainty.

We now formally define the various conditionals in terms of the master relation \succsim . The *conditional of \succsim on $s \in S$* is the relation \succsim_s on \mathbb{R}^T defined by the property that for all $\mathbf{x}_s, \mathbf{y}_s \in \mathbb{R}^T$,

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

Similarly, the *conditional of \succsim on $t \in T$* is the relation \succsim_t on \mathbb{R}^S defined by the property that for all $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^S$,

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

By themselves, these definitions do not make conditionals weak orderings. By a well-known fact of decision theory, \succeq_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 $\mathbf{X} \succsim \mathbf{Y} \iff \mathbf{X}' \succsim \mathbf{Y}'$ when \mathbf{X}', \mathbf{Y}' also satisfy the condition stated for \mathbf{X}, \mathbf{Y} in this definition. When this holds, \succeq is said to be *weakly separable* in s . By another well-known fact, weak separability in a factor (or set of factors) is equivalent to the property that \succsim is *increasing* with the conditional on this factor (or the conditionals of the set of factors). That is to say, for all $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{S \times T}$, if $\mathbf{x}_s \succeq_s \mathbf{y}_s$ for all s , then $\mathbf{X} \succsim \mathbf{Y}$; and if moreover $\mathbf{x}_s \succ_s \mathbf{y}_s$ for some s , then $\mathbf{X} \succ \mathbf{Y}$.² Everything said for s of course applies to t . Combining the two well-known facts, we see that conditions (i) and (iii) of the previous section can be fused into the single requirement that all \succeq_s and all \succeq_t are weak orderings.³

The *conditional of \succeq on $(s, t) \in S \times T$* is defined similarly. Since this conditional \succsim_{st} compares real numbers, it makes sense to identify it with the natural order of these numbers. This amounts to saying that numbers 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 turns the \succsim_{st} into an ordering, \succsim is increasing with each of these conditionals, hence also with each entry x_s^t of \mathbf{X} .

Let us say that the family $\{\succsim_s\}_{s \in S}$ ($\{\succsim_t\}_{t \in T}$) is *invariant* 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. Notice they are not needed for the \succsim_{st} since these are identical relations by construction.

We are now ready for a 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 each family of conditionals is invariant.*
- *There are an increasing, continuous function $u : \mathbb{R} \rightarrow \mathbb{R}$, and strictly positive probability functions $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$, such that \succeq is represented by the function $V : \mathbb{R}^{S \times T} \rightarrow \mathbb{R}$ 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.

As was foreshadowed, the representation theorem combines the conclusions of Bayesian decision theory with the desired property that the probability measure (here a vector) is multiplicative in the two sources. This theorem is Corollary 1(c) in Mongin and Pivato (2015) [16], but recast autonomously and in a different formal language, so as to facilitate the discussion of stochastic independence.

²By \succ , we mean the *strict* preference associated with the weak preference \succeq , and similarly for \succ_s , \succ^t and \succ_{st} .

³For these definitions and basic facts, see Fishburn (1970) [7], Keeney and Raiffa (1976) [13], and Wakker (1989) [22].

4 A second representation theorem for stochastic independence

In Theorem 1, strong results follow from a compact 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 and those which account for the stochastic independence property occurring in this representation? This disentangling would make sense on two counts: stochastic independence is an optional property of probability measures (a logical point) and Bayesian decision theory invests only the existence, not the properties, of such measures with universal rationality significance (a normative point). However, the assumptions of Theorem 1 cannot be divided in the appropriate way. This can be seen as follows. 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. By taking only one of the two families to satisfy the ordering and invariance assumptions, one would only get a representation that is only separable in that family and says nothing on probabilities either.⁴

Fortunately, we can obtain a relevant partitioning of assumptions if we enrich the decision-theoretic framework beyond the present, two-dimensional 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, heuristically unrelated dimension i , so that these consequences are now represented by real numbers x_{st}^i . The added dimension can be interpreted as being the *time* at which these consequences occur, and the corn producer decision problem can easily be reformulated accordingly. For technical reasons, we require i to take its values in a finite set I with cardinality $|I| \geq 2$.

Given this added dimension, alternatives become mappings from triples (s, t, i) to the reals, that is three-dimensional arrays,

$$\mathbb{X} = [x_{st}^i]_{s \in S, t \in T}^{i \in I} \in \mathbb{R}^{S \times T \times I},$$

which may be rewritten as

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

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

With i representing time, alternatives should be viewed as *contingent plans*, i.e., plans 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 just listed are interpreted in terms of *partly contingent plans* (when one dimension of uncertainty is fixed) or *dated prospects* (when the time dimension is fixed). Vector-valued objects can also be interpreted - e.g., the $\mathbf{x}_{st} = (x_{st}^i)_{i \in I} \in \mathbb{R}^I$ as *non-contingent plans*.⁵

⁴The 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 $v_{st} : \mathbb{R} \rightarrow \mathbb{R}$, $s \in S, t \in T$. In the second case, if the assumptions only hold for the \succsim_s , the separable representation reads as

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

with increasing and continuous $W : \mathbb{R}^S \rightarrow \mathbb{R}$ and $V_s : \mathbb{R}^T \rightarrow \mathbb{R}$, $s \in S$. These partial results follow from standard results in separability theory.

⁵It is essential for these interpretations 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.

We assume the agent compares contingent plans in terms of a preference relation \succsim , which is a continuous weak ordering, and this relation gives rise to conditional relations that are defined as in the previous section, *mutatis mutandis*. Among the seven families of conditionals, we pay special attention to $\{\succsim_s\}_{s \in S}$, $\{\succsim_t\}_{t \in T}$, $\{\succsim^i\}_{i \in I}$, $\{\succsim_{st}\}_{s \in S, t \in T}$ and $\{\succsim_{st}^i\}_{s \in S, t \in T, i \in I}$. The \succsim_s and \succsim_t compare plans \mathbf{X}_s and \mathbf{X}_t , respectively; the \succsim^i compare dated prospects \mathbf{X}^i , the \succsim_{st} non-contingent plans \mathbf{x}_{st} and the \succsim_{st}^i real-valued consequences. As before, we assume that each \succsim_{st}^i coincides with the natural order of real numbers, and that all other relations may or may not be weak orderings, and may or may not form an invariant family.

Theorem 2 *The following conditions are equivalent:*

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

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

In this format of representation, π is unique, and the u^i are unique up to positive affine transformations with a common multiplier.

Moreover, the following are equivalent:

- *The assumptions made above on the \succsim^i and \succsim_{st} hold, and furthermore the \succsim_s are weak orderings and an invariant family.*
- *The same conclusions hold, and moreover 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 $\mathbb{X} = [x_{st}^i]_{s \in S, t \in T, i \in I}$,*

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

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

Proof. (Sketch). The first part follows from Theorem 1 in Mongin and Pivato (2015) [16]; we leave it for the reader to check that the assumptions of this theorem apply here. For the second part, we first observe that, for every given $s \in S$, the $W(\mathbb{X})$ representation delivers a function $\mathbb{R}^{T \times I} \rightarrow \mathbb{R}$

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

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

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

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 transformation Φ_s s.t.

$$\sum_{t \in T} \sum_{i \in I} \pi'_{s_0 t} u^i(x_{st}^i) = \Phi_s \left(\sum_{t \in T} \sum_{i \in I} \pi'_{st} u^i(x_{st}^i) \right).$$

As the u^i are strictly increasing and continuous, Φ_s is $\mathbb{R} \rightarrow \mathbb{R}$, and we can apply a functional equation argument (Rado and Baker, 1987 [18]) 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.

$$\sum_{t \in T} \sum_{i \in I} \pi'_{s_0 t} u^i(x_{st}^i) = \alpha_s \left(\sum_{t \in T} \sum_{i \in I} \pi'_{st} u^i(x_{st}^i) \right) + \beta_s.$$

After redefining the functions so as to dispense with the constant terms, we see that, for all $s \in S$ and $t \in T$, $\pi'_{s_0 t} = \alpha_s \pi'_{st}$, and in fact (since proportional probability vectors are equal) $\pi'_{s_0 t} = \pi'_{st}$. We thus rewrite (*) as

$$\sum_{s \in \mathcal{S}} \sum_{t \in T} \pi'_{s_0 t} \left(\sum_{i \in I} \pi_{st} \right) (u^i(x_{st}^i)),$$

which is (**) if one takes $\mathbf{p} = (\sum_{t \in T} \pi_{st})_{s \in S}$ and $\mathbf{q} = (\pi'_{s_0 t})_{t \in T}$. The uniqueness of \mathbf{p} and \mathbf{q} in this format of representation is easily established. ■

The two steps of Theorem 2 correspond to the EU representation theorem and the addition made by stochastic independence, respectively. Interestingly, the same assumption - that of invariant conditional orderings - underlies both conclusions, but in the richer framework adopted here, it is possible to apply it twice over, thus separating each step. Note also that it is enough to apply the assumption to the \succsim_{st} and one of the two \succsim_s and \succsim_t families; then, as the representation shows, the other family automatically satisfies this assumption (the \succsim_t can of course be interchanged with the \succsim_s in the theorem statement).

5 Interpretations, comparisons and future directions

The following heuristic argument will help locate the preference ancestor of stochastic independence more precisely. Considering only four states for simplicity, we suppose that the agent considers (s_1, t_1) 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 draws 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. We now check that the s -conditional preferences cannot be invariant. Take ξ, ξ' representing desirable quantities, with $\xi > \xi'$, and the following prospects in matrix form:

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

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. By a similar comparison, the second line of \mathbf{X} should be preferred to the second line of \mathbf{Y} . Thus, the two s -conditional preferences differ. Contraposing the argument, we see that invariant s -preferences express a one-way form of *informational independence*, i.e., that observing s does not bring any information on t . The opposite one-way form, i.e., that observing t does not bring any information on s , would be expressed by invariant t -preferences. The two invariance conditions together convey a sense of *mutual informational independence*.

This concept of independence actually underlies one of the two main informal explications of stochastic independence, the other relying on the very different concept of *mutual causal independence*⁶. Today's

⁶The causal explication of stochastic independence is actually the older and more established of the two. On the problematic linkage of stochastic independence with causality, see Spohn (1980) [20].

probability texts sometimes entangle the two independence concepts, and it is perhaps one virtue of the above axiomatization to bring out the informational concept in a way that precludes any confounding with its causal competitor. Our betting agents may or may not be influenced by what they perceive of causal relations, but this is irrelevant, given that only their decisions matter to the analysis. However, the informational concept and associated informal explication can be pursued in various ways, and our axiomatization more specifically contributes to giving both a pragmatic slant. Generally, when it comes to the epistemic aspects of probability, a divide appears between theorists who would like to explore these aspects *per se*, and Bayesian decision theorists, who absorb them into their practical rationality concerns.⁷

The introduction sketched a comparison with two such theorists that can now be made precise. We share with Anscombe and Aumann (1963) [1] not only the two assumptions of a finite set of states and a structured set of consequences, but also that of two distinctive sources of uncertainty. However, unlike them, we do not suppose that one of these sources already has a probabilistic representation, a question-begging assumption from the perspective of Bayesian decision theory, since this would require a preference derivation for *every* kind of probability. Our completely preference-based approach likens it to Savage's (1954) [19] despite the technical dissimilarities concerning the set of states and set of consequences, as well as an axiomatic difference we now clarify. The assumptions in Theorems 1 and 2 that certain conditionals are orderings amount to replacing his postulate P2, i.e., the notorious "sure-thing principle", by a dominance principle, which is weaker and more generally accepted. However, to make good for this loss, we had to revise his other postulates, and this was done by the way in which we state the invariance of the orderings in question. Savage's P3 requires that conditional preferences be invariant across all possible conditioning events, but only when these preferences compare constant prospects. We require invariance only for some events, but - crucially - for any comparison of prospects, whether these are constant or not. The other event-invariance condition of Savage, P4, has no role to play here, because it serves to order an unstructured consequence set, while ours inherits the order structure of real numbers.⁸

The results of this paper may be extended in several directions. One of them is *conditional probability*, and the corresponding definition of stochastic independence in terms of this concept rather than that of joint probability. Conditional prospects can be introduced into the preference apparatus with relevant preference axioms. Such a variation is unlikely to make much difference to the conclusions, but it would be judged preferable by those probability theorists and philosophers who, unlike Kolmogorov, regard conditional probability as the genuine primitive of the probability calculus.⁹ Another, conceptually more problematic direction is *one-sided stochastic independence*. Such a concept appears rarely, if ever, in probability theory. In our preference framework, it is possible to express the idea that s does not bring any information on t whereas t may bring information on s ; it is enough to assume that the s -conditionals are invariant, while not assuming that the t -conditionals are. However, Theorem 2 teaches us in effect that, if the preference assumptions endow (s, t) with a probabilistic representation, one-sided informational independence automatically entails mutual informational independence. This suggests that the issue of one-sided stochastic independence may be impossible to pursue in the present framework; which other preference framework would facilitate its investigation is unclear. Last but not least, stochastic indepen-

⁷Joyce (1998) [11] has made this divide very clear, while adopting the former position.

⁸The difference with earlier results in Bayesian decision theory is also one of mathematical techniques. The proofs of Theorems 1 and 2 depend on additive separability arguments; see Mongin and Pivato (2015 [16], 2016 [17]), and for the source papers, Debreu (1960) [4] and Gorman (1968) [10].

⁹This is a common line to take, especially among philosophers of probability; in connection with stochastic independence, see Fitelson and Hajek (2014) [8].

dence has been reconsidered in the currently active work on multiple (or "imprecise") probabilities, and it would be an interesting project to connect one or more of the definitions given to it in this work with a decision-theoretic apparatus; the latter would of course not be Bayesian in the usual sense. Some steps have been taken in this direction, but much work still remains to be done.¹⁰

Acknowledgements. Many thanks for conceptual and technical comments to Lorraine Daston, Marcus Pivato, the audience of a seminar at the Munich Center for Mathematical Philosophy, and three TARK referees. The author also gratefully acknowledges the hospitality of the Max Planck Institut für Wissenschaftsgeschichte zu Berlin when he developed this project.

References

- [1] F. J. Anscombe & R. J. Aumann (1963): *A definition of subjective probability*. *The Annals of Mathematical Statistics* 34(1), pp. 199–205, doi:10.1214/aoms/1177704255.
- [2] S. Bade (2008): *Stochastic independence with maxmin expected utilities*. Penn State University, Mimeo.
- [3] F. G. Cozman (2012): *Sets of probability distributions, independence, and convexity*. *Synthese* 186(2), pp. 577–600, doi:10.1007/s11229-011-9999-0.
- [4] G. Debreu (1959): *Topological methods in cardinal utility theory*. In K. J. Arrow, S. Karlin & P. Suppes, editors: *Mathematical Methods in the Social Sciences*, Stanford University Press, pp. 16–26.
- [5] Z. Domotor (1969): *Probabilistic relational structures and their applications*. Technical Report 144, Stanford University, Institute for Mathematical Studies in the Social Sciences.
- [6] T. L. Fine (1973): *Theories of Probability*. Academic Press.
- [7] P. C. Fishburn (1970): *Utility Theory for Decision Making*. Wiley.
- [8] B. Fitelson & A. Hájek (2014): *Declarations of independence*. *Synthese*, doi:10.1007/s11229-014-0559-2.
- [9] P. Ghirardato (1997): *On independence for non-additive measures, with a Fubini theorem*. *Journal of Economic Theory* 73(2), pp. 261–291, doi:10.1006/jeth.1996.2241.
- [10] T. W. Gorman (1968): *The structure of utility functions*. *The Review of Economic Studies* 35(4), pp. 367–390, doi:10.2307/2296766.
- [11] J. M. Joyce (1998): *A Non-Pragmatic Vindication of Probabilism*. *Philosophy of Science* 65, pp. 575–603, doi:10.1086/392661.
- [12] M. Kaplan & T. L. Fine (1977): *Joint orders in comparative probability*. *The Annals of Probability*, pp. 161–179, doi:10.1214/aop/1176995843.
- [13] R. L. Keeney & H. Raiffa (1993): *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge University Press, doi:10.1017/CBO9781139174084.
- [14] A. N. Kolmogorov (1950): *Foundations of the Theory of Probability*. Chelsea Publishing Co. Original as *Grundbegriffe der Wahrscheinlichkeitsrechnung*, in *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 3, 1933.
- [15] R. D. Luce & L. Narens (1978): *Qualitative independence in probability theory*. *Theory and Decision* 9(3), pp. 225–239, doi:10.1007/BF00133452.
- [16] P. Mongin & M. Pivato (2015): *Ranking multidimensional alternatives and uncertain prospects*. *Journal of Economic Theory* 157, pp. 146–171, doi:10.1016/j.jet.2014.12.013.
- [17] P. Mongin & M. Pivato (2016): *Social preference under twofold uncertainty*. Technical Report Research Paper No. ECO/SCD-2016-1154, HEC Paris.

¹⁰See in particular Ghirardato (1997) [9] and Bade (2008) [2]. Cozman (2012) [3] surveys different ways in which stochastic independence can be redefined when beliefs are represented by sets of probabilities.

- [18] F. Radó & J. A. Baker (1987): *Pexider's equation and aggregation of allocations*. *Aequationes Mathematicae* 32(1), pp. 227–239, doi:10.1007/BF02311311.
- [19] L. J. Savage (1954): *The Foundations of Statistics*. John Wiley. Revised edition, 1972.
- [20] W. Spohn (1980): *Stochastic independence, causal independence, and shieldability*. *Journal of Philosophical Logic* 9(1), pp. 73–99, doi:10.1007/BF00258078.
- [21] P. Wakker (2010): *Prospect Theory*. Cambridge University Press, doi:10.1017/CBO9780511779329.
- [22] P. Wakker (2013): *Additive Representations of Preferences: A New Foundation of Decision Analysis*. 4, Springer Science & Business Media, doi:10.1007/978-94-015-7815-8.