Bayesian Decision Theory and Stochastic Independence

Philippe Mongin*†

As stochastic independence is essential to the mathematical development of probability theory, it seems that any foundational work on probability should be able to account for this property. Bayesian decision theory appears to be wanting in this respect. Savage's postulates on preferences under uncertainty entail a subjective expected utility representation, and this asserts only the existence and uniqueness of a subjective probability measure, regardless of its properties. What is missing is a preference condition corresponding to stochastic independence. To fill this significant gap, the article axiomatizes Bayesian decision theory afresh and proves several representation theorems in this novel framework.

1. Introduction. The property of stochastic (or statistical or probabilistic) 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. 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,

Received April 2018; revised January 2019.

*To contact the author, please write to: CNRS and HEC Paris, 1 rue de la Libération, F-78350 Jouy-en-Josas; e-mail: mongin@greg-hec.com.

†A first version of this article was presented at a seminar at the Munich Center for Mathematical Philosophy and at the Theoretical Aspects of Rationality and Knowledge (TARK) 2017 conference. The current version has particularly benefited from detailed comments made by Richard Bradley, Donald Gillies, Robert Nau, Marcus Pivato, Burkhard Schipper, Peter Wakker, Paul Weirich, two anonymous TARK referees, and two anonymous referees of this journal. I thank the Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) for supporting this research.

Philosophy of Science, 87 (January 2020) pp. 152–178. 0031-8248/2020/8701-0008$10.00
Copyright 2020 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

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, 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 (SI), but this is not the case. In this article, I investigate Bayesian decision theory, one of the most influential among these theories and at the same time a curious 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 typically stop their work when they have completed this demonstration. By not explaining how it can be extended to recover the definition of SI, they are open to the objection that they have not yet fully connected the probability calculus with rational belief. That is, they can be criticized for having handled only the basics of this calculus instead of 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. After Ramsey's and de Finetti's sketches, this strategy was implemented in full axiomatic detail by Savage (1972). At one point, Savage (1972, 44) extends the representation theorem to obtain a posterior probability measure, that is, 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 I report below, however, he acknowledges that a treatment of SI should have come next, but his admission of unfinished business was generally lost to later Bayesian decision theorists. The rare exceptions will be discussed below.

What is missing from Savage's axiom system and its variants is a condition put in the preference language of these systems that would account for SI specifically. The current article formulates one such condition. I do not use Savage's system but a more accessible one, and it is another contribution of the article to introduce this alternative system. The desired condition turns out to be highly simple and intuitive. I set up a framework in which there

https://doi.org/10.1086/706083 Published online by Cambridge University Press

are two distinct sources of uncertainty S and T, and states of nature thus have the form of two-component vectors $(s, t) \in S \times T$. In this framework, the condition relates to the agent's *conditional preferences* and stipulates that 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 effectively means is that if the agent knows t but not s, or knows s but not t, then this partial knowledge cannot improve the decisions made under the residual uncertainty. This recovers in axiomatic preference terms one of the standard informal justifications of SI: 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. When Bayesian decision theory is extended so as to encompass SI, there is no danger of confusion since the information carried by events matters only if it enters the agent's decision process and, 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 extended Bayesian theory is to give the informational reading of SI a foundation that clearly sets it apart from other possible readings. Apart from this semantic contribution, the theory casts some light on the reflective discussion of that property in 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 this literature has sometimes called into question.

I offer two representation theorems in succession. Both are adapted to the multiple uncertainty framework sketched above, and both deliver an expected utility formula in which the subjective probability is multiplicative in the values of two sources and, hence, satisfies SI. The system of theorem 1 has very few axioms, which deliver the two desired outputs jointly. This condensed treatment is both an advantage and a disadvantage. To derive a subjective expected utility formula and the multiplicative property in two successive 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 article is organized as follows. Section 2 introduces the twofold uncertainty framework and the preference axioms for theorem 1 via a motivating example. Section 3 states this result formally, and section 4 does the same for theorem 2. Section 5 is devoted to comparisons within Bayesian decision theory, including the earlier sketches of a preference analysis of SI.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

Section 6 draws connections with the philosophy of probability. Three appendixes collect the more specialized material. Appendix A gives proof details on the two theorems. Appendix B generalizes theorem 1 by allowing for zero-probability events. Appendix C pursues some of the comparisons of section 5 in more detail.

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

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

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 I will approach SI by specifying that the state set Ω is a *product set*, a standard move in the theory when it comes to working with this property (see, e.g., Halmos 1974, 191–92). The simple example of this section illustrates this framework and the main decision-theoretic concepts of the article.

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; suppose 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 this 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, that is, mappings from the states of the world to consequences, each of which represents a farming policy, which we take to be the set $\mathbb{R}^{S \times T}$ of all real functions on Ω. Finally, there is a binary relation $\gtrsim$ on the last set of prospects to capture the producer's preferences among cultivation policies.

Now suppose that this preference relation obeys the *expected utility* (EU) rule; that is, 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 $\mathbf{X}$, $\mathbf{Y}$,

$$\mathbf{X} \gtrsim \mathbf{Y} \text{ if and only if } (\text{iff}) \sum_{s,t} \pi(s,t) u(\mathbf{X}(s,t)) \geq \sum_{s,t} \pi(s,t) u(\mathbf{Y}(s,t));$$

and suppose moreover that π satisfies the SI property with respect to S and T; that is,

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

https://doi.org/10.1086/706083 Published online by Cambridge University Press

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 with an axiomatically derived subjective probability (what will be referred to as *subjective expected utility*, SEU) and the SI property of this subjective probability are the intended conclusions of the theorems of this article. In the current section, I reason heuristically, working backward 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 as follows:

	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 $\succsim$, that is,

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

can be restated either as

$$V(\mathbf{X}) = p_{s_1}[q_{t_1}u(x_{11}) + q_{t_2}u(x_{12})] + p_{s_2}[q_{t_1}u(x_{21}) + q_{t_2}u(x_{22})] \tag{1}$$

or as

$$V(\mathbf{X}) = q_{t_1}[p_{s_1}u(x_{11}) + p_{s_2}u(x_{21})] + q_{t_2}[p_{s_1}u(x_{12}) + p_{s_2}u(x_{22})]. \tag{2}$$

Observe that the bracketed sums in (1) are numerical representations for conditional preferences on the possible values of s, and those in (2) are numerical representations for conditional preferences on the possible values of t. Thus, the equations entail that (i) *these 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 (1), and similarly the same functional form $p_{s_1}u(\cdot) + p_{s_2}u(\cdot)$ appears in the two bracketed sums of (2), the equations also entail that (ii) *these conditional preferences are the same for different s and the same for different t*. Finally, 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

https://doi.org/10.1086/706083 Published online by Cambridge University Press

also holds that (iii) *preferences over prospects are increasing with respect to either family of conditional preferences*.

I have stated i, ii, and iii in pure preference language, thus abstracting entirely from the specifics of a numerical representation. In theorem 1 below, I assume no more than i, ii, and iii, plus some background conditions, and derive a SEU rule with the SI property. Given the definition of a conditional, which is restated below, it is actually possible to reduce iii to i and obtain an even more condensed system. One may wonder how such apparently weak conditions can do the mathematical work. One key point of the formal argument is that the weak ordering property, when applied to both s and t as in i, makes it possible to decompose the preference $\succsim$ not only in terms of s- and t-conditionals but also in terms of conditionals based on any other partitions of Ω and, from there, to get an additive representation for $\succsim$. The other key point is that the invariance property, when applied to both s and t as in ii, permits not only turning this additive representation into an EU formula but also giving the probabilities in this formula the multiplicative form.[1]

3. A First Representation Theorem Involving Subjective Expected Utility and 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 sources of uncertainty as in the motivating example, but this is only for mathematical simplicity. The next section will suggest how a larger number can be handled. We take the set of consequences to be $\mathbb{R}$ and the set of prospects to be all mappings $S \times T \to \mathbb{R}$, which is identified with $\mathbb{R}^{S \times T}$.[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 $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_{|S|})$, where each component is a row vector $\mathbf{x}_s \in \mathbb{R}^T$, or $\mathbf{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^{|T|})$, where each component is a column vector $\mathbf{x}^t \in \mathbb{R}^S$.

We endow the agent with an ex ante preference relation $\succsim$ to compare prospects and specifically assume that

(**A**) $\succsim$ is a continuous weak ordering.

1. These two heuristic features underlie the proof of theorem 3 by Mongin and Pivato (2015), which is the main technical tool in this article.

2. Again, these two assumptions are only for mathematical simplicity. Using the full technology of Mongin and Pivato (2015), the theorems of this article could be proved for smaller domains than $\mathbb{R}$ and $\mathbb{R}^{S \times T}$. This would permit paying attention to feasibility constraints on what counts as a consequence and what counts as a prospect.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

Hence, $\succsim$ is representable by a continuous utility function $U(\mathbf{X})$.[3] The agent's other preference relations are obtained from $\succsim$ as conditionals. There are three families of conditionals of interest, that is, $\{\succsim_s\}_{s\in S}$, $\{\succsim_t\}_{t\in T}$, 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 conditionals are defined from the master relation $\succsim$.[4] Take any subset of states $I \subseteq S \times T$. The *conditional* of $\succsim$ on I is the relation $\succsim_I$ on $\mathbb{R}^I$ defined by the property that for all $\mathbf{x}_I, \mathbf{y}_I \in \mathbb{R}^I$, $\mathbf{x}_I \succsim_I \mathbf{y}_I$ iff $\mathbf{X} \succsim \mathbf{Y}$ for some $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{S\times T}$ such that $\mathbf{X}$ is $\mathbf{x}_I$ on I, $\mathbf{Y}$ is $\mathbf{y}_I$ on I, and $\mathbf{X}$ and $\mathbf{Y}$ are equal outside I. Taking $I = \{s\} \times T$, $I = S \times \{t\}$, and $I = \{(s, t)\}$ in this definition, we obtain the three families of conditionals mentioned above.[5]

In terms of the matrix representation of prospects, the definitions of $\succsim_s$ and $\succsim_t$ read as follows:

- for all $\mathbf{x}_s, \mathbf{y}_s \in \mathbb{R}^T$, $\mathbf{x}_s \succsim_s \mathbf{y}_s$ iff $\mathbf{X} \succsim \mathbf{Y}$ for some $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{S\times T}$ such that $\mathbf{x}_s$ is the s-row of $\mathbf{X}$, $\mathbf{y}_s$ is the s-row of $\mathbf{Y}$, and $\mathbf{X}$ and $\mathbf{Y}$ are equal outside their s-row,
- for all $\mathbf{x}^t, \mathbf{y}^t \in \mathbb{R}^S$, $\mathbf{x}^t \succsim_t \mathbf{y}^t$ iff $\mathbf{X} \succsim \mathbf{Y}$ for some $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{S\times T}$ such that $\mathbf{x}^t$ is the t-column of $\mathbf{X}$, $\mathbf{y}^t$ is the t-column of $\mathbf{Y}$, and $\mathbf{X}$ and $\mathbf{Y}$ are equal outside their t-column.

Importantly, the definition of conditionals does not by itself make them weak orderings. By a well-known fact of decision theory, $\succsim_I$ is a weak ordering iff the choice of $\mathbf{X}$, $\mathbf{Y}$ in the definition of $\succsim_I$ is immaterial, or more precisely, iff

$$\mathbf{X} \succsim \mathbf{Y} \Leftrightarrow \mathbf{X}' \succsim \mathbf{Y}',$$

for all $\mathbf{X}'$, $\mathbf{Y}'$ that also satisfy the condition stated for $\mathbf{X}$, $\mathbf{Y}$ in this definition. In this case, the quantifier "for some" in the definition can be replaced by "for all." When this holds, $\succsim$ is said to be *separable in I*. By another well-known fact, separability in I is equivalent to the property that $\succsim$ is *monotonically increasing* with the family of conditionals $\{\succsim_i, i \in I\}$. We formally illustrate

3. A weak ordering (i.e., a reflexive, transitive, and complete binary relation) $\succsim$ on some Euclidean space $\mathbb{R}^n$ is continuous if for all $\mathbf{x} \in \mathbb{R}^n$, the upper and lower contour sets $\{\mathbf{x}' \in \mathbb{R}^n \mid \mathbf{x}' \succsim \mathbf{x}\}$ and $\{\mathbf{x}' \in \mathbb{R}^n \mid \mathbf{x} \succsim \mathbf{x}'\}$ are closed sets. This definition and the accompanying representation theorem are familiar since Debreu (1954) established them.

4. The formalism below is standard in decision theory. See, e.g., Fishburn (1970) and Wakker (1989) or, at a less technical level, Keeney and Raiffa (1993).

5. Throughout, I use the standard identification of s with $\{s\} \times T$, t with $S \times \{t\}$, and (s,t) with $\{(s,t)\}$.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

this property on the example of $\{\succsim_s, s \in S\}$. Then, if $\succsim$ is separable in each $s \in S$ (or equivalently, each $\succsim_s$ is a weak ordering), the following statement holds: 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 $\mathbf{X} \succ \mathbf{Y}$.[6]

Combining the two basic facts just said, we see that the monotonicity condition iii of the last section actually follows from the weak ordering condition i. This will permit reducing the assumptions of the theorem to a bare minimum.

Since conditionals $\succsim_{st}$ compare real numbers, it makes sense to identify them with the natural order of these numbers. This amounts to assuming that, given any realized state (s,t), they are desirable quantities, be they monetary values, as in the producer example, or something else. Thus, we also assume that

(**B**) for all $(s, t) \in S \times T$ and all $x_s^t, y_s^t \in \mathbb{R}$, $x_s^t \succsim_{st} y_s^t$ iff $x_s^t \geq y_s^t$.

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

Let us now say that the conditionals $\succsim_s$ (respectively $\succsim_t$) are an *invariant family* if $\succsim_s = \succsim_{s'}$ for all $s, s' \in S$ (respectively $\succsim_t = \succsim_{t'}$ for all $t, t' \in T$). These two requirements capture condition ii of last section. None is needed for the conditionals $\succsim_{st}$, since they are identical relations by construction.

We are now ready for the first representation theorem.

Theorem 1. Given assumptions (**A**) on $\succsim$ and (**B**) on $\succsim_{st}$, 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 both families of conditionals are invariant.
- There are increasing, continuous functions $u : \mathbb{R} \to \mathbb{R}$, and strictly positive probability functions $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$, such that $\succsim$ is represented by the function $V : \mathbb{R}^{S \times T} \to \mathbb{R}$ that computes the $\mathbf{p} \otimes \mathbf{q}$-expected value of u, that is, by the function thus defined: 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.

6. By $\succ$, $\succ_s$, $\succ_t$, and $\succ_{st}$, we denote the *strict* preference relation associated with the weak preference relations $\succsim$, $\succsim_s$, $\succsim_t$, and $\succsim_{st}$. We denote the corresponding *indifference* relations by $\sim$, $\sim_s$, $\sim_t$, and $\sim_{st}$.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

The conclusions state both that the ex ante preference relation can be represented by a SEU formula and that the two sources of uncertainty satisfy SI. Thus, this theorem extends Bayesian decision theory up to the point that, from the argument made in the introduction, it ought to have reached. The conclusion that **p** and **q** have only positive values is restrictive, but it can be relaxed by complexifying the assumptions. I pursue this technical line in appendix B.

4. A Second Representation Theorem Involving Subjective Expected Utility and 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 that are responsible for the existence of the SEU representation from those that 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 and, 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 distinguish between the utility and probability components of the added terms, unlike EU representations. By taking only one of the two families to satisfy the ordering and invariance assumptions, one would get a representation that is separable in that family alone and would be even more remote from the SEU representation.[7]

As it turns out, however, we can obtain a relevant partitioning of assumptions if we enrich the decision-theoretic framework beyond the current 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 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. I return to the interpretation of the added dimension after stating theorem 2.

7. The additively separable representation of the first point reads as $\Sigma_{s\in S,\, t\in T} v_{st}(x^t_s)$, with increasing and continuous $v_{st} : \mathbb{R} \to \mathbb{R}$, for all $s \in S$, $t \in T$. This follows from another theorem of Debreu (1960). As to the next point, if the assumptions only hold for the $\succsim_s$, the separable representation reads as $W(V_1(\mathbf{x}_1), \ldots, V_{|S|}(\mathbf{x}_{|S|}))$, with increasing and continuous $W : \mathbb{R}^S \to \mathbb{R}$ and $V_s : \mathbb{R}^T \to \mathbb{R}$, for all $s \in S$. On this representation, see Blackorby, Primont, and Russell (1978, 108).

https://doi.org/10.1086/706083 Published online by Cambridge University Press

By assumption, i takes its values in a finite set I with cardinality $|I| \geq 2$. The set of prospects is now the set of all mappings $S \times T \times I \to \mathbb{R}$, that is, $\mathbb{R}^{S\times T\times I}$. Prospects can be represented as three-dimensional arrays

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

or as vectors

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

the components of which are matrix valued; that is, $\mathbf{X}_s = [x^i_{st}]^{i\in I}_{t\in T} \in \mathbb{R}^{T\times I}$, $\mathbf{X}_t = [x^i_{st}]^{i\in I}_{s\in S} \in \mathbb{R}^{S\times I}$, and $\mathbf{X}^i = [x^i_{st}]_{s\in S, t\in T} \in \mathbb{R}^{S\times T}$, respectively.

Adapting the formalism of last section, we assume that

> ($\mathbf{A}'$) the agent's ex ante preference relation $\succsim$ on $\mathbb{R}^{S\times T\times I}$ is a continuous weak ordering,

and from this relation, we define four families of conditionals, that is, $\{\succsim_s, s \in S\}$, $\{\succsim_t, t \in T\}$, $\{\succsim^i, i \in T\}$, and $\{\succsim^i_{st}, s \in S, t \in T, i \in I\}$. The $\succsim_s$, $\succsim_t$, and $\succsim^i$ respectively compare matrices $\mathbf{X}_s$, $\mathbf{X}_t$, and $\mathbf{X}^i$, and 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

> ($\mathbf{B}'$) each $\succsim^i_{st}$ coincides with the natural order of real numbers,

which makes it an ordering and furthermore makes the $\succsim^i_{st}$ 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.** Given assumptions ($\mathbf{A}'$) on $\succsim$ and ($\mathbf{B}'$) on $\succsim^i_{st}$, the following conditions are equivalent:
>
> - The conditionals $\succsim^i$ are weak orderings.
> - There are increasing, continuous functions $u^i : \mathbb{R} \to \mathbb{R}$, for all $i \in I$, and a strictly positive probability function $\pi \in \Delta_{S\times T}$, such that $\succsim$ is represented by the function $W : \mathbb{R}^{S\times T\times I} \to \mathbb{R}$ that computes the π-expected value of $\Sigma_{i\in I} u^i$, that is, the function thus defined: for all $\mathbb{X} = [x^i_{st}]^{i\in I}_{s\in S, t\in T}$,
>
> $$W(\mathbb{X}) \coloneqq \sum_{s\in S}\sum_{t\in T}\sum_{i\in I} \pi_{st} u^i(x^i_{st}). \tag{3}$$
>
> In this format of EU representation, π is unique, and the u^i are unique up to positive affine transformations with a common multiplier.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

Moreover, the following items are equivalent:

- The above assumption on the $\succsim^i$ holds, and the $\succsim_s$ are weak orderings and an invariant family.
- The above 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 (3) becomes: for all $\mathbb{X} = [x^i_{st}]^{i \in I}_{s \in S, t \in T}$,

$$W(\mathbb{X}) := \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s q_t u^i(x^i_{st}). \tag{4}$$

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

Unlike theorem 1, theorem 2 is in two parts, corresponding to the SEU formula and the SI property respectively. 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) to be 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' as representing desirable quantities, with $\xi > \xi'$, and the following prospects in matrix form:

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

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 an agent entertains identical $\succsim_s$ and learns which s is realized without yet learning t, this agent would not use the knowledge so obtained to draw information on t.

The converse claim is problematic. Preference reversals may occur between $\succsim_s$ and $\succsim_t$ simply because the agent's preferences are sensitive to which of s or t is realized, and this *evaluative* disposition is logically unrelated to

https://doi.org/10.1086/706083 Published online by Cambridge University Press

the *epistemic* disposition of not using *s*-information to infer *t*-information. That is, the latter disposition can prevail in preference contexts in which the property that the $\succsim_s$ are an invariant family should not be assumed. The only way to secure a converse is to exclude the troublesome evaluative disposition by fiat, which all ordinary axiomatizations of SEU theory actually do, a strategy that has the definite shortcoming of narrowing the application range of the theory. For lack of a better solution, I reproduce this standard move here.

To derive (4), it was unnecessary to assume that both the $\succsim_s$ and the $\succsim_t$ are invariant. The invariance of the latter family follows from the representation (4) itself.[8] If it is sufficient to require one form of invariance, this is because an EU representation (3) holds from the previous stage. Under this assumption, it is impossible to distinguish between *s* bringing no relevant information on *t* and *t* bringing no relevant information on *s*. This is formally shown in appendix A and further discussed in appendix C.

I 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 $\mathbf{X}$ mean *contingent plans*, that is, 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 *noncontingent plans*, since they take the uncertainty to be resolved at each time period.[9]

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, which puts on the $\succsim^i$ the same invariance assumption as was imposed on the $\succsim_s$ and the $\succsim_t$. With this addition, it can be proved that (4) gives way to the following more specific representation: for all $\mathbb{X} = [x^i_{st}]^{i\in I}_{s\in S,t\in T}$,

$$W(\mathbb{X}) \coloneqq \sum_{s\in S}\sum_{t\in T}\sum_{i\in I} p_s q_t r_i u(x^i_{st}), \tag{5}$$

where $\mathbf{r} = (r_i)_{i\in I}$ is a strictly positive probability function on *I*, and the utility function *u* in the SEU formula is now independent of the *i* index. Apart from 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

8. From (4), the $\succsim_t$ are represented by $\Sigma_{s\in S}\Sigma_{i\in I} p_s u^i(\cdot)$, which does not depend on *t*.

9. 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.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

space Ω with any finite number of sources of uncertainty 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. I start this decision-theoretic section with two comments that Savage (1972) makes on SI. 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" (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" (91).[10] The two comments clearly express the need for Bayesian decision theory to complement its derivation of subjective probability with an account of SI, but they 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 complaint 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 SI of two events or two random variables.[11] The second comment does not mention qualitative probability, and I read in it a suggestion to base SI directly on the preference relation. In this respect, Savage's complaint is still topical. The current axiomatic work, which it motivated, appears to have been foreshadowed in only three papers, to be discussed now.

The main connection is with Blume, Brandenburger, and Dekel (1991, 74). These writers use Anscombe and Aumann's (1963) axiom system to explore the preference foundations of lexicographic probability, a topic apparently remote from the current one. However, their construction includes ordinary Kolmogorov probabilities as a particular case, and we may just focus on this application. To account for SI, they impose invariance of relevant conditionals of their ex ante weak ordering (which they have redefined so that it bears on a Cartesian product). Leaving aside the Anscombe-Aumann features (see the critical comments below), their condition is the same as mine.

10. Savage used to say "personal probability" where later theorists say "subjective probability."

11. See Domotor (1969), Fine (1973), Kaplan and Fine (1977), and Luce and Narens (1978), which are only the earliest accounts of SI in terms of qualitative probability. Fine's (1973) classic *Theories of Probabilities* makes interesting comments on SI, and at some point (36–37) it even suggests moving in the direction of a pragmatic, preference-based account of SI.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

Their paper does not contain a proof that this condition delivers SI, but I provide one in appendix C.

In an important follow up, which still uses Anscombe and Aumann's axiom system, Battigalli and Veronesi (1996) push the analysis of Blume et al. in the direction of *conditional probability systems* (CPS). These amount to taking conditional probabilities, rather than absolute probabilities, as being the primitive terms of the probability calculus, a move that philosophers usually associate with the works of Rényi (1955) and Popper (1972). For a suitable definition of what it means for a CPS to satisfy SI, Battigalli and Veronesi (1996, 243) connect this property with a form of conditional invariance. Given the dissimilarity of frameworks, the step from this condition to mine is far from trivial.

A comparison is also in order with a little-known axiomatization of Bayesian decision theory by Bernardo, Fernandez, and Smith (1985). These writers' derivation of SEU hinges on the strong assumption that a preexisting probability measure is available on a subalgebra of events; this will serve to calibrate subjective probability values. They define a preference condition relative to two events E and F that entails the equation $P(E \cap F) = P(E) \cdot P(F)$ at the stage of proving the representation theorem. Although evocative of the informational reading of SI, this condition differs from ours, and this difference seems connected with the technical choice of approaching SI in general probability spaces rather than product spaces.[12]

Theorems 1 and 2 should also be compared with the recent axiomatic work on *second-order expected utility*.[13] This work involves considering two sources of uncertainty; however, unlike the current one, it aims at establishing a logical hierarchy between these sources. By assumption, uncertainty takes both a primary form and a secondary form, which concerns what the primary form may be. This hierarchy can be interpreted epistemically, with the primary form bearing on the realization of natural events and the secondary form bearing on the realization of the primary form considered as an epistemic event, but it also admits of a temporal interpretation, with the primary and secondary sources corresponding to uncertainties of the second and first periods, respectively. In the current notation, if T represents the uncertain states of the primary form, and S those of the secondary form, a second-order EU representation of ex ante preferences is

$$\sum_{s \in S} p_s v\left(\sum_{t \in T} r_{st} u(x_s^t)\right),$$

12. Relatedly, Pfanzagl's (1968, 210–13) text on measurement theory states a preference condition for SI that is defined for a general probability space. Also using a calibration process, but with less elaboration than Bernardo et al. (1985), he derives a SEU formula that satisfies this property.

13. I am grateful to a referee for bringing out this connection.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

where $\mathbf{r}_s = (r_{st})_{t \in T}$ is a conditional probability function for t given s, v is a continuous and increasing function on $\mathbb{R}$, and u is a function on a consequence set that may be, but is not necessarily, $\mathbb{R}$. This representation is axiomatized by Nau (2006, 143). Ergin and Gul's (2009, 912) version uses an unconditional probability for t:[14]

$$\sum_{s \in S} p_s v\left(\sum_{t \in T} q_t u(x'_s)\right).$$

If one takes v to be a positive affine transformation, the two forms of uncertainty play symmetric roles, and both formulas collapse onto a SEU formula, with a subjective probability that satisfies SI in the latter; thus, one gets the representation of theorem 1 as a particular case. How the axioms for second-order EU should be strengthened so as to get this particular case is a relevant question, which I take up in appendix C.[15]

I now compare the axiomatization of the SEU formula in this article with those in current use. Being entirely preference based, the former is like Savage's (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. Also, the highly condensed axiom systems for theorems 1 and 2 do not relate to Savage's seven-postulate system in a transparent way. However, it is well known that given postulate P1 (which requires that the ex ante preference relation $\succsim$ be a weak ordering), postulate P2 (the "sure-thing principle") can be restated as the requirement that the conditional $\succsim_E$ on any event E of the state set be a weak ordering. This restatement facilitates comparison with the current system. I only assume that the conditionals on some distinguished events E of the state set are weak orderings. This weakening is in the spirit of axiomatizations of some non-EU rules that do not entirely relinquish the "sure-thing principle."

By contrast, the invariance condition used here is stronger than P3—the "event independence" postulate—because it bears on all possible prospects and not only on constant prospects, as is the case in that postulate. Savage has another important postulate, P4, which is a clear step toward the existence of subjective probability and has no analogue here; it can only be verified from the utility representation itself. My best guess is that P4 is made dispensable by the assumptions that consequences are real numbers and conditional

14. The classic article by Klibanoff, Marinacci, and Mukerji (2005, 1858) has a similar representation with integrals instead of finite sums.

15. Here is a final comparison. As a further development of Joyce's (1999) representation theorem for pairs of credibility and preference relations, Bradley (2017, 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 (1983) theory of decision, which is several steps removed from the Bayesian decision theory of this article and the works just reviewed.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

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. I 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 I am 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. A common rejoinder is that the preexisting probabilities are objective and, hence, of a different nature from the subjective probability to be derived, but this is a free commentary without any basis in Anscombe and Aumann's formal system. I do not deny the practical convenience of this system, but mine is no more complicated, while being perhaps easier to defend theoretically.[16]

6. Connections with Foundational Discussions on Probability. 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 practical 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 grounds that they are absolute measures and as an alternative develop a calculus for conditional probabilities taken as primitive terms. This line has recently been defended by Fitelson and Hájek (2017) in connection with Popper's (1972, apps. 4 and 5) pioneering work in this area. All existing CPS preserve the additivity of probability measures, and an alternative critical line is indeed to question that property. Decision theory has made thorough contributions on this score; see Gilboa (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.[17]

These deep foundational questions arise in connection with the current work but exceed its limited purpose. I meant to fill a gap in Bayesian decision

16. There are alternatives to this system other than the current one for application to finite state sets. An early example is Wakker's (1989, chap. 4).

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

https://doi.org/10.1086/706083 Published online by Cambridge University Press

theory by following its own principles, rather than defend these principles against outside criticism. However, since SI is the current focus, one 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.[18] 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.[19]

The Bayesian decision theory developed here contributes nothing to the first problem. Only a move from absolute to conditional probability could avoid the discrepancy between the provisional and final definitions of SI, and I have not performed this move here. However, on the second problem, the theory has something to say. At the very least, it avoids the equivocations between different informal accounts by firmly opting for informational independence. The invariance property of conditional preferences is the pragmatic criterion by which one can judge that the agent regards s as carrying no information on t.

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

18. Two examples among many of this two-step definitional sequence are to be found in Feller (1968, 125) and Hoel, Port, and Stone (1971, 19).

19. 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, 226). Naturally, one would expect "the knowledge of the other" instead of "the other."

https://doi.org/10.1086/706083 Published online by Cambridge University Press

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, 75–76) explains, citing from *Probabilismo* (de Finetti 1931), de Finetti argued against the application of SI to repeated trials of the same experiment on the grounds 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 in repeated trials 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 from one trial to the other; 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 nontrivial applications 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 possible contribution is to put the symmetry of the definition of SI in perspective. Writers on the foundations of probability have sometimes expressed dissatisfaction toward the fact that asymmetric dependence or independence cannot be formulated within the Kolmogorov axiomatic framework; see Fitelson and Hájek (2017) 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 Hájek concede. In the Bayesian decision theory of this article, one can assume the *s*-component of uncertainty to be informationally irrelevant to the *t*-component 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 SEU are all in place. This result can be understood in two opposite ways—those who take the preference conditions for SEU 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. Conclusion. I have responded to Savage's request to extend the preference apparatus of Bayesian decision theory to the point where it includes an account of SI. To do so, I have reconstructed this preference apparatus and proved representation theorems that contain both a novel derivation of SEU and the desired addition that the subjective probability makes the sources of uncertainty stochastically independent. These theorems call for richer

https://doi.org/10.1086/706083 Published online by Cambridge University Press

variants that need to be pursued elsewhere. One such variant would relax the finiteness assumption put on the set of states of the world. Apart from 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 SI in relation to the nonadditive measures of uncertainty that decision theorists have introduced since they moved away from a primarily Bayesian conception. 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 properties but also to discover those new mathematical definitions that capture SI when probability gives way to weaker notions.

Appendix A

Proofs

In the two theorems of this article, the direction from the SEU representation to the axiomatic conditions is clear. The other direction follows from a result proved by Mongin and Pivato (2015, theorem 1). I restate this result in the current notation and a simplified form adapted to the purpose of deriving them.

Theorem 3. Given assumptions (**A**) on $\succsim$ and (**B**) on $\succsim_{st}$, the following conditions are equivalent:

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

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

In this format of EU representation, **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 over and checking the compatibility

https://doi.org/10.1086/706083 Published online by Cambridge University Press

of the obtained representations. See Mongin and Pivato (2015, corollary 1) for mathematical details.

The first part of the conclusions 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. We first observe that, for every given $s \in S$, the (3) representation of the first part of the conclusions delivers a function $\mathbb{R}^{T\times I} \to \mathbb{R}$,

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

which represents the weak ordering $\succsim_s$. If we define $\pi'_{st} \coloneqq \pi_{st}/\Sigma_{t'\in T}\pi_{st'}$ 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 $s_0 \in S$ and take any $s \in S$. By the invariance of the $\{\succsim_s\}_{s\in S}$ family, there is an increasing function Φ_s on $\mathbb{R}$ such that for all $x^i_{st} \in \mathbb{R}$,

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

Now for all $t \in T$, $i \in I$, define

$$z^i_{st} \coloneqq \pi'_{st} u^i(x^i_{st}),$$

so that the previous equation becomes a Pexider equation:

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

As the u_i are increasing and continuous, and so are their double sums, Φ_s is continuous and open.[20] It follows that the set of possible values for the vector $(z^i_{st})^{i\in I}_{t\in T}$ is connected and open. Thus, we can apply a theorem on Pexider equations and conclude that the Φ_s are positive affine transformations.[21]

20. For a proof of continuity, see Fleurbaey and Mongin (2016, lemma 1). Openness is easily established.

21. This functional equation theorem is due to Rado and Baker (1987, theorem 1 and corollary 2). They formally prove it for sums of two terms, but Fleurbaey and Mongin (2016, lemma 2) make the easy generalization to sums of any number of terms.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

That is, there exist a number $\alpha_s > 0$, and for all $t \in T$ and $i \in I$, numbers β^i_{st} such that for all $x^i_{st} \in \mathbb{R}$,

$$\pi'_{s_0t} u^i(x^i_{st}) = \alpha_s \pi'_{st} u^i(x^i_{st}) + \beta^i_{st}.$$

This entails that for all $t \in T$, $\pi'_{s_0t} = \alpha_s \pi'_{st}$, and in fact (since proportional probability vectors are equal) $\pi'_{s_0t} = \pi'_{st}$. We can thus rewrite (3) as

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

which is (4) if one defines $\mathbf{p} := (\Sigma_{t' \in T} \pi_{st'})_{s \in S}$ and $\mathbf{q} := (\pi'_{s_0t})_{t \in T}$. 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 $\succsim^i$ leads to the stronger representation claimed in the main text, that is,

$$W(\mathbb{X}) := \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} p_s q_t r_i u(x^i_{st}), \tag{5}$$

it is enough to reproduce the proof sequence used for (4) mutatis mutandis. QED

Appendix B

Extension to Zero-Probability Events

Theorems 1, 2, and 3 all involve strictly positive probability functions. This restriction is due to the assumption that all of the $\succsim_{st}$ (in theorems 1 and 3) and all of the $\succsim^i_{st}$ (in theorem 2) reproduce the natural order of real numbers. It can indeed be checked that this makes the $\succsim_s$, $\succsim_t$, and $\succsim^i$ nonconstant preference relations, so that there are no "null events" in Savage's (1972, 24) sense and, hence, no zero-probability values either. Given that the state set is finite, it is possible to prune it of the null states without creating side effects on the rest of the framework. This move is often performed in mathematical treatments of decision theory (see, e.g., Debreu 1954, 128; Wakker 1989, 83). One may complain, however, that it is unfit for some game-theoretic applications: different players can differ in their assignments of zero-probability values, and this fact may be invested with strategic relevance. Moreover, and in closer connection with the aims of this article, discarding zero-probability values would leave the current extension of Bayesian theory incomplete.[22] I now

22. I am grateful to a referee for pressing this point.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

show how theorem 1 can be modified so as to include null states and zero-probability values.

In the formal framework of section 3, we define the conditional $\succsim_I$ of $\succsim$ on $I \subseteq S \times T$ to be constant if for all $\mathbf{x}_I, \mathbf{y}_I \in \mathbb{R}^I$, $\mathbf{x}_I \succsim_I \mathbf{y}_I$. When, for a given $s \in S$ (respectively a given $t \in T$), $\succsim_s$ (respectively $\succsim_t$) is constant, we say that s (respectively t) is a *null component*, and when for a given $(s, t) \in S \times T$, $\succsim_{st}$ is constant, we say that (s,t) is a *null state*. We assume that the subsets $S^* \subseteq S$ and $T^* \subseteq T$ of nonnull components satisfy the cardinality restriction $|S^*|$, $|T^*| \geq 2$. In this enriched framework, a more general form of theorem 1 can be proved. In this statement, assumption (**A**) is unchanged, but (**B**) is replaced by

(**B***) for all $(s, t) \in S^* \times T^*$ and all $x_s^t, y_s^t \in \mathbb{R}$, $x_s^t \succsim_{st} y_s^t$ iff $x_s^t \geq y_s^t$.

Theorem 4. Given assumptions (**A**) and (**B***), the following conditions are equivalent:

- For all $s \in S$ and $t \in T$, the conditionals $\succsim_s$ and $\succsim_t$ are weak orderings, and so are the conditionals $\succsim_{st}$ for all $(s, t) \notin S^* \times T^*$. Furthermore, the subfamilies of conditionals $\{\succsim_s\}_{s \in S^*}$ and $\{\succsim_t\}_{t \in T^*}$ are invariant.
- There are increasing, continuous function $u : \mathbb{R} \to \mathbb{R}$, and probability functions $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$, such that $\succsim$ is represented by the function $V : \mathbb{R}^{S \times T} \to \mathbb{R}$ that computes the $\mathbf{p} \otimes \mathbf{q}$-expected value of u, that is, by the function defined as follows: for all $\mathbf{X} = [x_s^t]_{s \in S}^{t \in T}$,

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

with $p_s = 0$ iff $s \in S \backslash S^*$ and $q_t = 0$ iff $t \in T \backslash T^*$, that is, iff s (respectively t) is a null component.

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

Proof. The derivation of the axiomatic conditions from the SEU representation is clear. For the converse derivation, let us first observe that, for all $s \in S$ and all $t \in T$, the conditional $\succsim_{st}$ is constant iff either $s \in S \backslash S^*$ or $t \in T \backslash T^*$. In one direction, if $\succsim_{st}$ is nonconstant, it follows from the weak ordering assumption put on $\succsim_s$ and $\succsim_t$ that these two conditionals are nonconstant. (This assumption ensures that the values outside (s,t) that are used to define $\succsim_{st}$ from $\succsim$ can also be used to define $\succsim_s$ and $\succsim_t$ from $\succsim$, a step needed for the nonconstancy conclusion.) In the other direction, if both $\succsim_s$ and $\succsim_t$ are nonconstant, B* applies, and then $\succsim_{st}$ is nonconstant. Hence, $\Sigma \coloneqq (S^* \times T^*)^c$ is the set of null states. In the next step, we take prospects

https://doi.org/10.1086/706083 Published online by Cambridge University Press

$\mathbf{X}$ in vector form and decompose them into their $S^* \times T^*$ and Σ subsets of entries, denoting the corresponding subvectors by $\mathbf{X}_{S^* \times T^*}$ and $\mathbf{X}_\Sigma$ respectively.

We aim at reaching the following equivalence: for all $\mathbf{X} = (\mathbf{X}_{S^* \times T^*}, \mathbf{X}_\Sigma)$, $\mathbf{Y} = (\mathbf{Y}_{S^* \times T^*}, \mathbf{Y}_\Sigma) \in \mathbb{R}^{S \times T}$, and for all $\mathbf{Z}_\Sigma = (z_s^t)_{(s,t) \in \Sigma} \in \mathbb{R}^\Sigma$,

$$\mathbf{X} \succsim \mathbf{Y} \text{ iff } (\mathbf{X}_{S^* \times T^*}, \mathbf{Z}_\Sigma) \succsim (\mathbf{Y}_{S^* \times T^*}, \mathbf{Z}_\Sigma). \tag{B1}$$

The proof uses the property that for all $\mathbf{X}' = (x_s'^t)_{(s,t) \in S \times T}$, $\mathbf{Y}' = (y_s'^t)_{(s,t) \in S \times T} \in \mathbb{R}^{S \times T}$,

$$\text{if } x_s'^t \sim_{st} y_s'^t \text{ for all } s \in S \text{ and } t \in T, \text{ then } \mathbf{X}' \sim \mathbf{Y}'.$$

This monotonicity property results from the fact that the $\succsim_{st}$ are weak orderings. Regarding the $\succsim_{st}$ such that $(s, t) \in \Sigma$, this fact follows from the assumptions and the definition of Σ, and regarding the $\succsim_{st}$ such that $(s, t) \in S^* \times T^*$, it can be proved by a technical argument.[23] Since the $\succsim_{st}$ are constant for all $s \in S \backslash S^*$ and $t \in T \backslash T^*$, monotonicity entails that

$$(\mathbf{X}_{S^* \times T^*}, \mathbf{X}_\Sigma) \sim (\mathbf{X}_{S^* \times T^*}, \mathbf{Z}_\Sigma)$$

and that

$$(\mathbf{Y}_{S^* \times T^*}, \mathbf{Y}_\Sigma) \sim (\mathbf{Y}_{S^* \times T^*}, \mathbf{Z}_\Sigma),$$

whence (B1) follows by the transitivity of $\succsim$. Now, we may let $\mathbf{Z}_\Sigma$ vary in (B1), and from the observation that this does not change the equivalence conclude that the conditional $\succsim_{S^* \times T^*}$ of $\succsim$ on $S^* \times T^*$ is a weak ordering. In words, (B1) means that the particular values taken on the null states make no difference to the ex ante preferences.

To move from (B1) to the representation, we apply theorem 1 to $\succsim_{S^* \times T^*}$, viewed as a new primitive preference relation on the new set of prospects $\mathbb{R}^{S^* \times T^*}$. It can easily be checked that, in this framework, all assumptions needed for theorem 1 hold. It follows that there are increasing continuous functions $u : \mathbb{R} \to \mathbb{R}$, and strictly positive probability functions $\mathbf{p}^* \in \Delta_{S^*}$ and $\mathbf{q}^* \in \Delta_{T^*}$, such that $\succsim_{S^* \times T^*}$ is represented by

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

By (B1), this is also a representation of $\succsim$. It can equivalently be formulated in terms of $\mathbf{p} \in \Delta_S$ and $\mathbf{q} \in \Delta_T$, as in the conclusion of the theorem. The uniqueness of the $\mathbf{p}^*$ and $\mathbf{q}^*$ for the SEU format of representation of

23. This argument involves Gorman's (1968) overlapping separability theorem and is part of the proof of theorem 3 (see Mongin and Pivato 2015, lemma 4). The role of (**B***) in this argument is to ensure that all subsets of $S^* \times T^*$ are essential in Gorman's sense.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

$\succsim_{S^* \times T^*}$ carries through to **p** and **q** for the SEU format of representation of $\succsim$. QED

A noticeable difference between theorems 1 and 4 is that the latter imposes a weak ordering assumption not merely on the $\succsim_s$ and $\succsim_t$ but also on some $\succsim_{st}$. This extra assumption is essential to secure the monotonicity property that drives the proof above.

Appendix C

Technical Connections

Blume et al. (1991, 74) suggest, but do not formally establish, that the SI property of subjective probability follows from their invariance condition, given their choice of an Anscombe-Aumann axiomatization of SEU. A proof for this claim can be obtained by merely adapting that which appendix A gave for the second part of theorem 2. Let us first assume that the Anscombe-Aumann axiomatization has delivered a SEU representation for $\succsim$ of the form

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

and then follow the existing proof by ignoring the complication created by the *i* index. Thus modified, the proof is exactly suitable to Blume et al.'s claim when there are only two factors in the Cartesian product domain. The case of any finite number of factors, as in their explicit statement, can be handled by a recursive argument. Notice that it should be enough to require invariance with respect to $n - 1$ factors, as the representation itself should entail invariance for the *n*th one, in the same way as in the two-factor case. More importantly, the current formal argument involves a functional equation step that depends on having a finite set of states of the world, unlike in Savage.

The conclusion of theorem 1 can be related to the second-order EU representations discussed in section 5. Nau's (2006) framework is most closely related to the current one. He considers two sets of finite cardinality S and T and takes consequences to be real numbers, in his case representing monetary values, so that prospects **X** are elements of $\mathbb{R}^{S \times T}$ (using my notation). At a preliminary stage of his investigation, which will be sufficient for comparative purposes, Nau assumes that the ex ante preference relation $\succsim$ satisfies A and is increasing in each x_s^t, and he imposes two strong separability conditions. One amounts to assuming the full force of Savage's P2 with respect to S-events (i.e., events of the form $S' \times T$, $S' \subseteq S$), and the other, to assuming a conditional form of P2 with respect to T-events (i.e., events of the form

https://doi.org/10.1086/706083 Published online by Cambridge University Press

$S \times T'$, $T' \subseteq T$). In this conditional form, P2 applies to each $\succsim_s$ taken separately rather than the master relation $\succsim$. This asymmetric treatment of the S and T sources of uncertainty is an essential feature of the second-order construction. The following utility representation for $\succsim$ ensues (Nau 2006, 143):

$$W(\mathbf{X}) \coloneqq \sum_{s \in S} v_s \left(\sum_{t \in T} u_{st}(x_s^t) \right),$$

where the v_s and u_{st} are continuous and increasing functions on $\mathbb{R}$. To reach the conclusion of theorem 1 from this formula, one must restrict the v_s to be positive affine transformations. Here is a necessary and sufficient condition for this to obtain: for all $t \in T$, $\succsim_t$ is a weak order. Essentially, when added to the others, this condition secures (P2) on all events of $S \times T$, whence an additive representation follows for $\succsim$, and the result derives from comparing this representation with $W(\mathbf{X})$.[24]

A comparison is also possible with Ergin and Gul (2009), although their framework is more complex than Nau's and the current one. As in Savage, they postulate an infinite set of states of the world and end up with nonatomic probability measures.[25] Barring these and other technical differences, their work also makes it possible to relate SI to second-order EU representations. By itself, their formula with unconditional probability for t, that is,

$$\sum_{s \in S} p_s v \left(\sum_{t \in T} q_t u(x_s^t) \right),$$

does imply that SI holds between s and t (Ergin and Gul 2009, 906). If v does not degenerate to a positive affine transformation, this property manifests itself in the separate use of two probability measures on S and T. If v degenerates, it manifests itself in the SI property of a single probability measure on $S \times T$, as in theorem 1. This simplification can be explained in terms of the agent's second-order risk attitude, thanks to a theorem in Ergin and Gul (2009, 911): v is a positive affine transformation iff the agent is indifferent between risky and nonrisky second-order prospects having the same mean. This establishes an intriguing connection between the SI property and the psychology of the agent of Bayesian decision theory.

24. In more detail, suppose $\succsim_t$ is a weak ordering for all $t \in T$. Then, by the same argument as in n. 23, $\succsim_A$ is a weak ordering for all subsets $A \subseteq S \times T$. Given that $\succsim$ is increasing with each x_s^t, Debreu's (1960) theorem applies, and $\succsim$ is represented by $\Sigma_{s,t} w_{st}(x_s^t)$, where the w_{st} are continuous and increasing functions. Fixing t, one compares two representations of $\succsim_t$, i.e., $\Sigma_s w_{st}(x_s^t)$ and $\Sigma_s v_s(u_{st}(x_s^t) + k_s)$, where $k_s = \Sigma_{t' \in T} v_s(u_{st}(\bar{x}_s^{t'})$ for arbitrarily fixed values $\bar{x}_s^{t'}$. A functional equation argument in the style of that used to prove theorem 2 leads one to the conclusion that the v_s are positive affine functions.

25. If Ergin and Gul's representation involves finite sums, this is because they restrict attention to prospects with a finite number of distinct consequences.

https://doi.org/10.1086/706083 Published online by Cambridge University Press

REFERENCES

Anscombe, F. J., and R. J. Aumann. 1963. "A Definition of Subjective Probability." *Annals of Mathematical Statistics* 34:199–205.
Battigalli, P., and P. Veronesi. 1996. "A Note on Stochastic Independence without Savage-Null Events." *Journal of Economic Theory* 70:235–48.
Bernardo, J. M., J. R. Fernandez, and A. F. M. Smith. 1985. "The Foundations of Decision Theory: An Intuitive, Operational Approach with Mathematical Extensions." *Theory and Decision* 19:127–50.
Blackorby, C., D. Primont, and R. R. Russell. 1978. *Duality, Separability and Functional Structures: Theory and Applications*. New York: North-Holland.
Blume, L., A. Brandenburger, and E. Dekel. 1991. "Lexicographic Probability and Choice under Uncertainty." *Econometrica* 59:61–79.
Bradley, R. 2017. *Decision Theory with a Human Face*. Cambridge: Cambridge University Press.
Debreu, G. 1954. "Representation of a Preference Ordering by a Numerical Function." In *Decision Processes*, ed. R. M. Thrall, C. H. Combs, and R. L. Davis, 159–65. New York: Wiley. Repr. as chap. 6 in G. Debreu, *Mathematical Economics* (Cambridge: Cambridge University Press, 1983).
———. 1960. "Topological Methods in Cardinal Utility Theory." In *Mathematical Methods in the Social Sciences, 1959*, ed. K. J. Arrow, S. Karlin, and P. Suppes, 16–26. Stanford, CA: Stanford University Press. Repr. as chap. 9 in G. Debreu, *Mathematical Economics* (Cambridge: Cambridge University Press, 1983).
de Finetti, B. 1931. *Probabilismo*. Biblioteca di filosofica. Napoli: Perrella. Repr. in B. de Finetti, *La logica dell'incerto* (Milan: Il Saggiatore, 1989), 3–70. English trans. "Probabilism," *Erkenntnis* 31 (1989): 169–223.
Domotor, Z. 1969. "Probabilistic Relational Structures and Their Applications." Technical Report 144, Institute for Mathematical Studies in the Social Sciences, Stanford University.
Ergin, H., and F. Gul. 2009. "A Theory of Subjective Compound Lotteries." *Journal of Economic Theory* 144:899–929.
Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*. 3rd ed. New York: Wiley.
Fine, T. L. 1973. *Theories of Probability: An Examination of Foundations*. New York: Academic Press.
Fishburn, P. C. 1970. *Utility Theory for Decision Making*. New York: Wiley.
Fitelson, B., and A. Hájek. 2017. "Declarations of Independence." *Synthese* 194:3979–95.
Fleurbaey, M., and P. Mongin. 2016. "The Utilitarian Relevance of the Aggregation Theorem." *American Economic Journal: Microeconomics* 8:289–306.
Gilboa, I. 2009. *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.
Gillies, D. 2000. *Philosophical Theories of Probability*. London: Routledge.
Gorman, W. 1968. "The Structure of Utility Functions." *Review of Economic Studies* 35:367–90.
Halmos, P. R. 1974. *Measure Theory*. New York: Springer.
Hoel, P. G., S.C. Port, and C. J. Stone. 1971. *Introduction to Probability Theory*. Boston: Houghton-Mifflin.
Jeffrey, R. 1983. *The Logic of Decision*. 2nd ed. Chicago: University of Chicago Press.
Joyce, J. M. 1998. "A Non-pragmatic Vindication of Probabilism." *Philosophy of Science* 65:575–603.
———. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Kaplan, M., and T. L. Fine. 1977. "Joint Orders in Comparative Probability." *Annals of Probability* 5:161–79.
Keeney, R. L., and H. Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge: Cambridge University Press.
Klibanoff, P., M. Marinacci, and S. Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica* 73 (6): 1849–92.
Kolmogorov, A. N. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Ergebnisse der Mathematik und ihrer Grenzgebiete 3. Berlin: Springer. Trans. Nathan Morrison as *Foundations of the Theory of Probability*, 2nd ed. (New York: Chelsea, 1956).

https://doi.org/10.1086/706083 Published online by Cambridge University Press

Leitgeb, H., and R. Pettigrew. 2010. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77:236–72.
Luce, R. D., and L. Narens. 1978. "Qualitative Independence in Probability Theory." *Theory and Decision* 9:225–39.
Mongin, P., and M. Pivato. 2015. "Ranking Multidimensional Alternatives and Uncertain Prospects." *Journal of Economic Theory* 157:146–71.
Nau, R. F. 2006. "Uncertainty Aversion with Second-Order Utilities and Probabilities." *Management Science* 52:136–45.
Pfanzagl, J. 1968. *Theory of Measurement*. New York: Wiley.
Popper, K. R. 1972. *The Logic of Scientific Discovery*. 6th ed. London: Hutchinson.
Rado, F., and J. Baker. 1987. "Pexider's Equation and Aggregation of Allocations." *Aequationes Mathematicae* 32:227–39.
Rényi, A. 1955. "On a New Axiomatic Theory of Probability." *Acta Mathematica Academiae Scientiae Hungaricae* 6:268–335.
Savage, L. J. 1972. *The Foundations of Statistics*. 2nd ed. (1st ed. 1954) New York: Wiley.
Wakker, P. 1989. *Additive Representations of Preferences*. Dordrecht: Kluwer.
———. 2010. *Prospect Theory*. Cambridge: Cambridge University Press.

https://doi.org/10.1086/706083 Published online by Cambridge University Press