Econ Theory
https://doi.org/10.1007/s00199-018-1122-z

RESEARCH ARTICLE

Stable behavior and generalized partition

David Dillenberger[1] · **Philipp Sadowski**[2]

Received: 18 May 2017 / Accepted: 19 April 2018

Abstract Behavior is *stable* if the ex ante ranking of two acts that differ only on some event I coincides with their ex post ranking upon learning I. We identify the largest class of information structures for which the behavior of a Bayesian expected utility maximizer is stable. We call them *generalized partitions* and characterize the learning processes they can accommodate. Often, the information structure is not explicitly part of the primitives in the model, and so becomes a subjective parameter. We propose a way to identify how the individual plans to choose contingent on learning an event, and establish that for a Bayesian expected utility maximizer, stable behavior—formulated in terms of this indirectly observed contingent ranking—is a tight characterization of subjective learning via a generalized partition.

Keywords Bayesian updating · Stable behavior · Subjective expected utility · Subjective learning · Dynamic consistency · Generalized partition

JEL Classification D80 · D81

Some of the results in this paper previously appeared in Dillenberger and Sadowski (2012). We thank David Ahn, Brendan Daley, Laura Doval, Haluk Ergin, Itzhak Gilboa, Faruk Gul, Peter Landry, Wolfgang Pesendorfer, Todd Sarver, Andrei Savochkin, Pablo Schenone, Marciano Siniscalchi, and Roee Teper for their comments and suggestions.

✉ David Dillenberger
ddill@sas.upenn.edu

Philipp Sadowski
p.sadowski@duke.edu

1 Department of Economics, University of Pennsylvania, Philadelphia, USA

2 Department of Economics, Duke University, Durham, USA

Published online: 23 April 2018

1 Introduction

Consider a standard dynamic problem of decision making under uncertainty, where given a space of payoff-relevant states of nature, S, a Bayesian expected utility maximizer anticipates obtaining some information (in the form of signals that are statistically related to the true state) before taking an action.[1] It is well known that if the underlying information structure is a partition of S, where a signal corresponds to an event I that contains the true state $s \in S$, then the individual will update his prior by excluding all states that are not in I, keep the relative weights of any two states in I fixed, and choose according to the resulting posterior. In this case, the ex ante ranking of acts that differ only on event I coincides with their ex post ranking contingent on actually learning I. We will refer to this behavioral property as *stable behavior*.

An important property of Bayesian decision making is that it is dynamically consistent, in the sense that the ex ante comparison of alternatives contingent on receiving a certain signal is the same as the ex post comparison of those alternatives after the realization of the signal in question. Stable behavior is commonly associated with dynamic consistency and is often referred to as such (see, for example, Ghirardato 2002). The implicit assumption in identifying the two as equivalent is that the following two rankings coincide: (i) the ex ante comparison of alternatives contingent on learning that $s \in I$ and (ii) the ex ante ranking of acts that differ only on event I. But there are many information structures on S for which the two differ. For example, if for two states $s, s' \in S$ the information structure is such that conditional on the true state being s the decision maker (henceforth DM) will be told that event $\{s, s'\}$ occurs, but if it is s' then either $\{s'\}$ or $\{s, s'\}$ will be randomly reported, then Bayes' law implies that beliefs should shift toward s upon learning $\{s, s'\}$. Consequentially, if initially the DM is indifferent between betting on s versus betting on s', then conditional on learning $\{s, s'\}$ he will strictly prefer betting on s.

In this paper, we first identify the largest class of information structures for which the behavior of a Bayesian expected utility maximizer is stable (Theorem 1), meaning that (ii) is indeed an appropriate substitute for (i), in case the latter is unobserved. We call this class *generalized partitions* and show that it can accommodate many plausible learning processes.

Second, we analyze contexts where information is not explicitly stated as part of the primitives. This may be because the analyst does not know how information is generated. Alternatively, the analyst may know, but the true information structure has not been communicated to the DM. In either situation, the information structure the DM perceives is subjective, so that the analyst is unaware which information the DM reads into different signals. We propose a way to identify—from observable behavior—how the DM plans to choose contingent on subjectively learning event I (i.e., contingent on ruling out all states outside of I and retaining all those in I). Our main result, Theorem 2, shows that for a Bayesian expected utility maximizer, stable behavior is a tight characterization of subjective learning via a generalized partition.

[1] We assume that the space of payoff-relevant states is exogenously given to the analyst. For example, when analyzing a specific data set, the collection of alternative actions accounted for, and with it the space of payoff-relevant states, is given. See also the remark at the end of Sect. 3.

The literature often imposes stable behavior for all possible events $I \subset S$, without explicitly mentioning the underlying information structure and without discussing how one could elicit the collection of ex post preferences for all $I \subset S$.[2] Our results suggest how (at least for any event I the DM can foresee learning) the analyst can elicit preferences contingent on learning I, denoted $\succeq_I$, from ex ante behavior without any knowledge of the underlying information structure. Our main theorem then implies that the behavior of a Bayesian expected utility maximizer is stable for this family of contingent preferences, if and only if it is generated by a generalized partition. Importantly, the class of generalized partitions includes information structures that support all events,[3] so that it may be possible to elicit $\succeq_I$ for all $I \subset S$. In other words, stable behavior for all $I \subset S$ is indeed compatible with Bayesian decision making when the information structure is subjective, but restricts it to be a generalized partition with full support.

In order to identify $\succeq_I$ from observed behavior, we follow the approach in Dillenberger et al. (2014, henceforth DLST) and take as primitive a preference relation over sets (or menus) of acts, defined over a given space of payoff-relevant states.[4] In Theorem 2, we derive a *generalized-partition representation*, which is interpreted as follows: the DM behaves as if he (i) has prior beliefs over the state space and (ii) perceives a particular type of stochastic rule, namely a generalized partition, that determines which event he will learn contingent on the true state. Upon learning an event, the DM calculates posterior beliefs using Bayes' law, which leads him to exclude all states that are not in that event, keeping the relative weights of the remaining states fixed. He then chooses from the menu the act that maximizes the corresponding expected utility. The prior and the generalized partition are uniquely identified.

As the name suggests, generalized partition extends the notion of a set partition, according to which the DM learns which cell of a partition contains the true state. In the case of a set partition, signals are deterministic; that is, for each state there is only one possible event that contains it.[5] Another example of a generalized partition is a random partition, where one of multiple partitions is randomly drawn and then an event in it is reported. A situation that may give rise to a random partition is an experiment with uncertainty about its precision. Alternatively, DM may be unsure about the exact time at which he will have to choose an act from the menu, and different partitions may simply correspond to the information the DM expects to have at different points in time. A sequential elimination of candidates, say during a recruiting process, may

[2] An experimenter in the laboratory might have at his disposal different partitions and be able to commit prior to the realization of the state which one of those to use for generating a signal. In that case, eliciting $\succeq_I$ for all $I \subset S$ may be possible. In many situations, the analyst cannot commit to different partitions, in particular if the information structure is unknown to him or has not been communicated to the DM.

[3] While obviously there is no partition that supports all $I \subset S$, a randomization over all bi-partitions $\left\{I, I^C\right\}$ is a generalized partition with this property.

[4] The interpretation is that the DM initially chooses among menus and subsequently chooses an act from the menu. If the ultimate choice of an act takes place in the future, then the DM may expect information to arrive prior to this choice. Analyzing preferences over future choice situations allows one to identify the anticipated future choice behavior without observing ex post preferences and to interpret it in terms of the information the DM expects to receive.

[5] The special case of partitional learning is analyzed in DLST.

also lead to learning via a generalized partition; if k candidates out of n are to be eliminated in the first stage, then the resulting collection of events the DM might learn is the set of all $(n-k)$-tuples. Theorem 3 characterizes all types of learning processes that can be accommodated by a generalized partition.

2 Stable behavior and generalized partition

Let $S=\{s_1,\ldots,s_k\}$ be a finite state space. An act is a mapping $f:S\rightarrow[0,1]$. We interpret payoffs in $[0,1]$ to be in utils; that is, we look directly at utility acts, assuming that the cardinal utility function over outcomes is known and payoffs are stated in its units.[6] For any number $c\in[0,1]$, we simply denote by c the constant act that yields c in every state. Let $\mathcal{F}$ be the set of all acts.

We model a DM who expects to receive a signal prior to choosing an act from some feasible set. Let Σ be the set of all signals and let $r:S\times\Sigma\rightarrow[0,1]$, with $\sum_{\sigma\in\Sigma}r_s(\sigma)=1$ for all $s\in S$, be the information structure the DM faces, where $r_s(\sigma)$ denotes the probability of learning signal $\sigma\in\Sigma$ given $s\in S$. The following class of information structures will play a central role in our analysis.

Definition 1 A function $\rho:2^S\rightarrow[0,1]$ is a *generalized partition* of S if for any $s\in S$ and $I\in 2^S$, ρ_s defined by $\rho_s(I)=\begin{cases}\rho(I) & \text{if } s\in I\\ 0 & \text{if } s\notin I\end{cases}$ satisfies $\sum_{I\subseteq S}\rho_s(I)=1$.

A generalized partition induces an information structure where $\Sigma\subset 2^S$ and $\rho_s(I)$ is the probability of event $I\in\Sigma$ being reported contingent on the state being s. (When there is no danger of confusion, we simply refer to ρ itself as an information structure.) The special case of a set partition corresponds to ρ taking only two values, zero and one. In that case, for every $s\in S$ there exists a unique $I_s\in 2^S$ with $s\in I_s$ and $\rho_s(I_s)=1$. Furthermore, $s'\in I_s$ implies that $I_s=I_{s'}$, that is, $\rho_{s'}(I_s)=1$ for all $s'\in I_s$.

We assume throughout that the DM is a Bayesian expected utility maximizer. We divide the analysis below into two parts. In the next subsection, we study the case of observable information where the analyst can observe both ex ante and ex post preferences (after a signal has been received) over acts. We link stable behavior to learning via generalized partition. We then proceed to study the case of subjective learning.

2.1 Observable information

Suppose the analyst is aware of the information structure r faced by the DM. In this section, we assume that the DM's preferences over acts prior to receiving information,

[6] Our analysis can be easily extended to the case where, instead of $[0,1]$, the range of acts is a more general vector space. In particular, it could be formulated in the Anscombe and Aumann (1963) setting. Since our focus is on deriving the DM's subjective information structure, we abstract from deriving the utility function (which is a standard exercise) by looking directly at utility acts instead of the corresponding Anscombe–Aumann acts.

$\succeq^o$, as well as his ex post preferences contingent on learning $\sigma \in \Sigma$, $\succeq^o_\sigma$, are observed. When signals can take values in the space of all events, $\Sigma \subset \mathcal{I} := \{I \subseteq S \,|\, I \neq \phi\}$, stable behavior can be formulated as follows:

Axiom 1 (Stable behavior) *If $I \in \Sigma \subset \mathcal{I}$ and $f, g \in \mathcal{F}$ with $f(s) \neq g(s)$ only if $s \in I$, then*

$$f \succeq^o g \Leftrightarrow f \succeq^o_I g.$$

That is, behavior is stable if the ex ante ranking of acts that differ only on event I coincides with the ex post ranking after I was reported.

Definition 2 The DM is a *Bayesian expected utility maximizer* given the information structure r, if there is a prior μ with $\text{supp}(\mu) = S$,[7] such that prior to the arrival of information he maximizes an expected utility according to μ (i.e., evaluates an act f by $\sum_{s\in S} f(s)\,\mu(s)$), and upon receiving signal $\sigma \in \Sigma$ he replaces μ with its Bayesian posterior, $\Pr(\cdot\,|\sigma)$.

Theorem 1 *For information structure r with $\Sigma \subset \mathcal{I}$, the following statements are equivalent for a Bayesian expected utility maximizer.*

1. *The rankings $\succeq^o$ and $\left\{\succeq^o_I\right\}_{I\in\Sigma}$ satisfy* Axiom 1.
2. *There is a generalized partition ρ such that $r_s \equiv \rho_s$.*

Proof See "Appendix B" □

To see why the theorem holds, note that Bayes' law states that $\Pr(s\,|I) = r_s(I)\,\mu(s)/\mu(I)$. For $s, s' \in I$ this immediately implies that

$$\frac{\Pr(s\,|I)}{\Pr(s'\,|I)} = \frac{r_s(I)\,\mu(s)}{r_{s'}(I)\,\mu(s')}$$

The key observation is that $r_s \equiv \rho_s$ for some generalized partition ρ, if and only if $r_s(I) = r_{s'}(I)$ for $s, s' \in I$. Hence, relative weights of all states in I do not shift upon learning I,

$$\frac{\Pr(s\,|I)}{\Pr(s'\,|I)} = \frac{\mu(s)}{\mu(s')}$$

if and only if $r_s \equiv \rho_s$ for some generalized partition ρ. That is, non-shifting-weights are equivalent to statement (2) in the theorem.

In terms of behavior, changes in relative weights are solely responsible for any changes in the DM's ranking of acts that differ only on event I. Conversely, relative weights do not shift if and only if that ranking remains fixed. That is, relative weights on I do not shift when learning I, if and only if $\succeq^o$ and $\left\{\succeq^o_I\right\}_{I\in\Sigma}$ satisfy Axiom 1, which is statement (1) in the theorem.

[7] Where $\text{supp}(\mu) = \{s\,|\mu(s) > 0\}$.

2.2 Subjective information

We now consider an environment where the information structure r is subjective and ex post preferences are not observable. We adopt the menu choice approach, with the interpretation that preferences over menus of acts reflect the DM's anticipated, signal-contingent preferences. Let $\mathcal{K}(\mathcal{F})$ be the set of all non-empty compact subsets of $\mathcal{F}$. Capital letters denote sets, or menus, and small letters denote acts. For example, a typical menu is $F = \{f, g, h, \ldots\} \in \mathcal{K}(\mathcal{F})$. Let $\succeq$ be a binary relation over $\mathcal{K}(\mathcal{F})$. The symmetric and asymmetric parts of $\succeq$ are denoted by $\sim$ and $\succ$, respectively. Let $fIg \in \mathcal{F}$ be the act that agrees with $f \in \mathcal{F}$ on event I and with $g \in \mathcal{F}$ outside I, that is,

$$fIg(s) = \begin{cases} f(s) & s \in I \\ g(s) & s \notin I \end{cases}$$

2.2.1 Anticipated stable behavior

In order to capture stable behavior in terms of preferences over menus, we start by defining the class of *irreducible events*. According to Definition 3, event $I \in 2^S$ is irreducible if and only if a bet on I, $\{cI0\}$, cannot be perfectly substituted by the option to bet on any of its strict subsets, $\left\{cI'0 \,\middle|\, I' \subset I\right\}$.[8] Intuitively, event I is irreducible if and only if the DM considers all states in I possible and can foresee learning at least one signal σ that does not rule out any of those states.

Definition 3 Event $I \subseteq S$ is *irreducible* if for any $c > 0$,

$$\{cI0\} \succ \left\{cI'0 \,\middle|\, I' \subset I\right\}.$$

Note that contingent on any signal σ with $r_s(\sigma) = 0$, the act $c(I \setminus \{s\})\,0$ is as good as $cI0$. Consequently, if for every foreseen signal σ there is at least one state $s \in I$ with $r_s(\sigma) = 0$, then $\{cI0\} \sim \left\{cI'0 \,\middle|\, I' \subset I\right\}$. That is, $\left\{cI'0 \,\middle|\, I' \subset I\right\}$ is a perfect substitute for $\{cI0\}$. Conversely, if the DM considers all states in I possible and can foresee a signal σ with $r_s(\sigma) > 0$ for all $s \in I$,[9] then contingent on learning σ, the act $cI0$ does strictly better than any act in $\left\{cI'0 \,\middle|\, I' \subset I\right\}$, and thus $\{cI0\} \succ \left\{cI'0 \,\middle|\, I' \subset I\right\}$.

Let $\mathcal{F}_+ := \{f \in \mathcal{F} \,|\, f(s) > 0 \text{ for all } s \in S\}$ be the collection of acts with strictly positive payoffs in all states. Under very mild assumptions, it is sufficient to impose our main axiom only for acts in $\mathcal{F}_+$, since $\mathcal{F}_+$ is dense in $\mathcal{F}$ with respect to the Euclidean metric (viewing acts as vectors in $[0, 1]^{|S|}$). Confining attention to $\mathcal{F}_+$ is convenient in writing Definition 4 and Axiom 2.

[8] Throughout the paper, $\subseteq$ (resp., $\subset$) denotes weak (resp., strict) set inclusion. The support of any function is denoted by supp$(\cdot)$.

[9] Formally, all states in I should not be null in the sense of Savage (1954), i.e., for every s it is not the case that $\{f\} \sim \{g\{s\}f\}$ for all f and g. Clearly, if $s \in I$ is null, then $c(I \setminus \{s\})\,0$ is as good as $cI0$, and hence $\{cI0\} \sim \left\{cI'0 \,\middle|\, I' \subset I\right\}$, that is, $\left\{cI'0 \,\middle|\, I' \subset I\right\}$ is a perfect substitute for $\{cI0\}$.

Definition 4 For an *irreducible* event I and acts $f, g \in \mathcal{F}_+$ with $f(s) \neq g(s)$ only if $s \in I$, define $\succeq_I$ by

$$f \succeq_I g \Leftrightarrow \{f, g\} \sim \{f\} \cup \{0\{s\}g\,|s \in I\}.$$

According to Definition 4, $f \succeq_I g$ describes the property that the DM is willing to commit to choose f over g as long as no state in I has been ruled out. To see this, note that $g \in \mathcal{F}_+$ and hence for any signal σ and $s \in I$ such that $r_s(\sigma) > 0$, g is better than $0\{s\}g$. Therefore, the DM will be indifferent between $\{f, g\}$ and $\{f\} \cup \{0\{s\}g\,|s \in I\}$ upon learning σ only if he expects to choose f from $\{f, g\}$. At the same time, for any signal σ such that $r_s(\sigma) = 0$ for some $s \in I$, the DM will be indifferent between choosing from $\{f, g\}$ and choosing from $\{f\} \cup \{0\{s\}g\}$.

Axiom 2 (Anticipated stable behavior) *For any irreducible event I and acts $f, g \in \mathcal{F}_+$ with $f(s) \neq g(s)$ only if $s \in I$,*

$$\{f\} \succeq \{g\} \Rightarrow f \succeq_I g.$$

Axiom 2 captures the idea of Axiom 1 in the subjective environment: the ranking of two acts that differ only on event I does not change as long as no state in I has been ruled out.[10]

2.2.2 Generalized-partition representation

The following utility function over menus incorporates the notion of a generalized partition.

Definition 5 The pair (μ, ρ) is a *generalized-partition representation* if (i) $\mu : S \to (0, 1]$ is a probability measure with $\text{supp}(\mu) = S$; (ii) $\rho : 2^S \to [0, 1]$ is a generalized partition of S; and (iii)

$$V(F) = \sum_{I \in 2^S} \max_{f \in F} \left[\sum_{s \in I} f(s)\,\mu(s)\right] \rho(I)$$

represents $\succeq$.

According to Definition 5, the DM behaves as if he (i) holds prior beliefs μ with full support on the state space S and (ii) perceives a generalized partition that determines which event he will learn contingent on the true state.[11] Upon learning event I, the DM calculates posterior beliefs using Bayes' law, which leads him not to shift weights; he excludes all states that are not in I and keeps the relative likelihood of the remaining

[10] It is worth mentioning that Axiom 2 is consistent with the possibility that the DM has strict preference for flexibility, in the sense that it does not preclude the ranking $\{f, g\} \succ \{f\} \succeq \{g\}$. That is, while given I the DM would prefer f to g when no states in I have been ruled out, it may well be that for some σ with $r_s(\sigma) = 0$ for some $s \in I$, g is better than f.

[11] We could allow $\text{supp}(\mu) = S' \subset S$, where ρ is a generalized partition of S'. For ease of notation we omit this conceptually straightforward generalization.

states fixed. He then chooses from the menu the act that maximizes the corresponding expected utility.

The two parameters of the generalized-partition representation, the prior μ on S and the generalized partition ρ, are independent, in the sense that the definition places no joint restriction on the two. In other words, a generalized-partition representation accommodates any prior beliefs the DM might have about the objective state space, combined with any information structure that can be described as a generalized partition. It is evident from the description of the model that the information structure in Definition 5 is not objectively given. Instead, the generalized partition should be derived from choice behavior.

To investigate the behavior of a Bayesian expected utility maximizer in this context, we take as a starting point preferences that have a subjective-learning representation, introduced in DLST

Definition 6 A subjective-learning representation is a function $V : \mathcal{K}(\mathcal{F}) \to \mathbb{R}$, such that

$$V(F) = \int_{\Delta(S)} \max_{f \in F} \left(\sum\nolimits_{s \in S} f(s)\, \pi(s) \right) \mathrm{d}p(\pi), \tag{1}$$

where $p(\cdot)$ is a probability measure on $\Delta(S)$, the space of all probability measures on S.

The interpretation of a subjective-learning representation is that the DM behaves as if he expects to learn information that will lead him to update his prior, thereby inducing ex ante beliefs over the possible posterior distributions that he might face at the time of choosing from the menu. For each posterior $\pi \in \Delta(S)$, he expects to choose from the menu the act that maximizes the corresponding expected utility. This DM can always be though of as Bayesian; this is the case since the DM's evaluation of singletons reveals his prior beliefs μ, and, according to (1), $\mu(s) = \int_{\Delta(S)} \pi(s)\, \mathrm{d}p(\pi)$ for any $s \in S$. As established by DLST (Theorem 1 in their paper), the axioms that are equivalent to the existence of a subjective-learning representation are familiar from the literature on preferences over menus of lotteries—*Ranking, vNM Continuity, Nontriviality,* and *Independence*—adapted to the domain $\mathcal{K}(\mathcal{F})$, in addition to *Dominance*, which implies monotonicity in payoffs, and *Set Monotonicity*, which captures preference for flexibility. The prior μ has support S if all states $s \in S$ are *non-null* (see Footnote 9). DLST show that the function $p(\cdot)$ in (1) is unique.

Theorem 2 *Suppose that the relation* $\succeq$ *admits a subjective-learning representation [as in* (1)*]. Then* $\succeq$ *satisfies* Axiom 2 *if and only if it has a generalized-partition representation,* (μ, ρ)*. Furthermore, the pair* (μ, ρ) *is unique.*

Proof See "Appendix C". □

For a Bayesian expected utility maximizer, stable behavior tightly characterizes learning via a generalized partition. Obviously, generalized partition is consistent with the familiar set partition where signals are deterministic contingent on the state, but it also accommodates many other plausible learning processes. In the next section, we provide examples, as well as a characterization, of those learning processes that give rise to generalized partitions. In terms of behavior, the following is a simple example of a pattern which our axioms and representation accommodate, but would be precluded if anticipated signals were deterministic. Consider the state space $\{s_1, s_2\}$ and the menu $\{(1,0), (0,1), (1-\varepsilon, 1-\varepsilon)\}$, which contains the option to bet on either state, as well as an insurance option that pays reasonably well in both states. A DM who is uncertain about the information he will receive by the time he has to choose from the menu may strictly prefer this menu to any of its subsets (for ε small enough). For instance, an investor may value the option to make a risky investment in case he understands the economy well, but also value the option to make a safe investment in case uncertainty remains unresolved at the time of making the investment choice. Our model accommodates this ranking. In contrast, such a ranking of menus is ruled out if signals are deterministic. If the DM expects to learn the true state, then preference for flexibility stems *exclusively* from the DM's prior uncertainty about the true state and the insurance option is irrelevant, that is, $\{(1,0), (0,1), (1-\varepsilon, 1-\varepsilon)\} \sim \{(1,0), (0,1)\}$. And if the DM does not expect to learn the true state, then, for ε small enough, he anticipates choosing the insurance option with certainty, that is, $\{(1,0), (0,1), (1-\varepsilon, 1-\varepsilon)\} \sim \{(1-\varepsilon, 1-\varepsilon)\}$.[12]

3 A characterization of generalized partitions

Compared to the information structures permitted by a subjective-learning representation [as captured by the function $p(\cdot)$ in (1)] a generalized partition (Definition 1) rules out information structures in which two distinct states can generate the same signal but with different probabilities. For example, consider a defendant who is on a trial. Generalized partition is inconsistent with a setting in which there are two states of nature, guilty (G) or innocent (I), and two signals, guilty (g) and innocent (i), such that $\Pr(g\,|G) > \Pr(g\,|I) > 0$ and $\Pr(i\,|I) > \Pr(i\,|G) > 0$.

In this section, we characterize the types of learning processes that can give rise to a generalized partition, by identifying all sets of events that support such an information structure. Formally, we characterize the set

$$\left\{\Psi \subseteq 2^S \,\middle|\, \text{there is a generalized partition } \rho : 2^S \to [0,1] \text{ with } \operatorname{supp}(\rho) = \Psi \right\}.$$

It is worth mentioning that the analysis in this section does not have any decision theoretical aspect and it also does not depend on the distinction between observable information and unobservable information.

Definition 7 A collection of events $\Psi \subseteq 2^S$ is a *uniform cover* of S, if (i) $S = \bigcup_{I\in\Psi} I$ and (ii) there is a function $\beta : \Psi \to \mathbb{Z}_+$ and a constant $k \geq 1$, such that $\sum_{I\in\Psi|s\in I} \beta(I) = k$ for all $s \in S$.

To better understand the notion of a uniform cover, consider the following example. Suppose $S = \{s_1, s_2, s_3\}$. Any partition of S, for example $\{\{s_1\}, \{s_2, s_3\}\}$, is a

[12] "Appendix A" provides a formal behavioral comparison between our model and one with deterministic signals.

uniform cover of S (with $k = 1$). A set that consists of multiple partitions, for example $\{\{s_1\}, \{s_2, s_3\}, \{s_1, s_2, s_3\}\}$, is a uniform cover of S (in this example with $k = 2$). The set $\Psi = \{\{s_2, s_3\}, \{s_1, s_2, s_3\}\}$ is *not* a uniform cover of S, because $\sum_{I|s_1 \in I} \beta(I) < \sum_{I|s_2 \in I} \beta(I)$ for any $\beta : \Psi \to \mathbb{Z}_+$. The set $\{\{s_2, s_3\}, \{s_1\}, \{s_2\}, \{s_3\}\}$, however, is a uniform cover of S with

$$\beta(I) = \begin{cases} 2 & \text{if } I = \{s_1\} \\ 1 & \text{otherwise} \end{cases}.$$

Lastly, the set $\{\{s_1, s_2\}, \{s_2, s_3\}, \{s_1, s_3\}\}$ is a uniform cover of S (with $k = 2$), even though it does not contain a partition.

An empirical situation that gives rise to a uniform cover consisting of two partitions is an experiment that reveals the state of the world if it succeeds and is completely uninformative otherwise. For a concrete example that gives rise to a uniform cover that does not contain a partition, consider the sequential elimination of n candidates, say during a recruiting process. If k candidates are to be eliminated in the first stage, then the resulting uniform cover is the set of all $(n - k)$-tuples.

Theorem 3 *A collection of events Ψ is a uniform cover of S if and only if there is a generalized partition $\rho : 2^S \to [0, 1]$ with $\text{supp}(\rho) = \Psi$.*[13]

Proof See "Appendix D" □

The 'only if' part in the proof of Theorem 3 amounts to finding a solution to a system of linear equations, showing that any uniform cover can give rise to a generalized partition. To illustrate the idea, consider the collection $\{\{s_1\}, \{s_2, s_3\}, \{s_1, s_2, s_3\}\}$ discussed above, which consists of multiple partitions and is a uniform cover with $k = 2$. An information structure r that can be described as a generalized partition

$$\rho : \{\{s_1\}, \{s_2, s_3\}, \{s_1, s_2, s_3\}\} \to (0, 1]$$

should satisfy two conditions. First, $r_s(I)$—the probability of receiving the signal I given state s—should be positive and identical for any $s \in I$, that is,

$$\begin{aligned} &r_{s_1}(\{s_1\}) =: \rho(\{s_1\}) > 0 \\ &r_{s_2}(\{s_2, s_3\}) = r_{s_3}(\{s_2, s_3\}) =: \rho(\{s_2, s_3\}) > 0 \\ &r_{s_1}(\{s_1, s_2, s_3\}) = r_{s_2}(\{s_1, s_2, s_3\}) = r_{s_3}(\{s_1, s_2, s_3\}) =: \rho(\{s_1, s_2, s_3\}) > 0 \end{aligned}$$

and 0 otherwise. Second, each $r_s(\cdot)$ must be a probability measure, that is,

$$\begin{aligned} r_{s_1}(\{s_1\}) + r_{s_1}(\{s_1, s_2, s_3\}) = 1 \\ r_{s_2}(\{s_2, s_3\}) + r_{s_2}(\{s_1, s_2, s_3\}) = 1 \\ r_{s_3}(\{s_2, s_3\}) + r_{s_3}(\{s_1, s_2, s_3\}) = 1 \end{aligned}$$

[13] The notion of uniform cover is closely related to that of balanced collection of weights, which Shapley (1967) introduces in the context of cooperative games. Similarly, a generalized partition formally resembles a cooperative game with unit Shapley value.

The solutions of this system are given by $\rho(\{s_1, s_2, s_3\}) = \alpha$ and $\rho(\{s_1\}) = \rho(\{s_2, s_3\}) = 1 - \alpha$, for any $\alpha \in (0, 1)$.

Theorem 3 characterizes the types of learning that can be accommodated by a generalized partition. To illustrate it, let us consider a specific example. An oil company is trying to learn whether there is oil in a particular location. Suppose the company can perform a test drill to determine accurately whether there is oil, $s = 1$, or not, $s = 0$. In that case, the company learns the uniform cover $\Psi = \{\{0\}, \{1\}\}$ that consists of a partition of the state space, and $\rho(\{0\}) = \rho(\{1\}) = 1$ provides a generalized partition. Now suppose that there is a positive probability that the test may not be completed (for some exogenous reason, which is not indicative of whether there is oil or not). The company will either face the trivial partition $\{\{0, 1\}\}$, or the partition $\{\{0\}, \{1\}\}$, and hence $\Psi = \{\{0, 1\}, \{0\}, \{1\}\}$. Suppose the company believes that the experiment will succeed with probability q. Then $\rho(\{0, 1\}) = 1 - q$ and $\rho(\{0\}) = \rho(\{1\}) = q$ provides a generalized partition.

We can extend the previous example and suppose the company is trying to assess the size of an oil field by drilling in l proximate locations, which means that the state space is now $\{0, 1\}^l$. As before, any test may not be completed, independently of the other tests. This is an example of a situation where the state consists of l different attributes (i.e., the state space is a product space), and the DM may learn independently about any of them. Such learning about attributes also gives rise to a uniform cover that consists of multiple partitions and can be accommodated.

To find a generalized partition based on (i) a uniform cover Ψ of a state space S, for which there is a collection Π of partitions whose union is Ψ and (ii) a probability distribution q on Π, one can set $\rho(I) = \sum_{\mathcal{P} \in \Pi | I \in \mathcal{P}} q(\mathcal{P})$. We refer to the pair (q, Π) as a random partition.

A different situation in which the DM effectively faces a random partition, (q, Π), is one where learning is governed by a given filtration of S, but the speed of learning is uncertain. In that case, distinct partitions in Π simply correspond to the information the DM might have at the time of choice and should thus be ordered by fineness. In that case, $q(\mathcal{P})$ captures the probability of arriving at $\mathcal{P}$ by the time of choice.

Lastly, reconsider the example of sequential elimination of candidates outlined above. Suppose that one out of three candidates will be selected. Write the corresponding state space as $S = \{s_1, s_2, s_3\}$. If one candidate will be eliminated in the first round, then the uniform cover of events the DM might learn is given by $\Psi = \{\{s_1, s_2\}, \{s_2, s_3\}, \{s_1, s_3\}\}$. Suppose that, contingent on person i being the best candidate, the DM considers any order of elimination of the other candidates as equally likely. This corresponds to the generalized partition with $\rho(I) = 0.5$ for all $I \in \Psi$ and $\rho(I) = 0$ otherwise.

Each of the generalized partitions discussed in this section can be coupled with any prior beliefs μ to generate a generalized-partition representation (μ, ρ).

Remark 1 If the state space is defined via the value of all random variables the DM might observe, then it gives rise to an information structure that is a partition. Conversely, any information structure can always be described via a partition, if the state space is made sufficiently large. To attain a state space that is surely large enough, one could follow Savage and postulate the existence of a grand state space that describes all *conceivable* sources of uncertainty. Identification of beliefs on a larger state space, however, generally requires a much larger collection of acts, which poses a serious conceptual problem, as in many applications the domain of choice (the available acts) is given. In that sense, acts should be part of the primitives of the model.[14] Our approach instead identifies a behavioral criterion for checking whether a given state space (e.g., the one acts are naturally defined on in a particular application) is large enough: behavior satisfies a version of stable behavior (either Axiom 1 or Axiom 2) if and only if the resolution of any subjective uncertainty corresponds to an event in the state space. Our results demonstrate that this does not require a state space on which learning generates a partition. To emphasize our point, reconsider the drilling example, with $S = \{0, 1\}$ and a probability q for the test to be completed successfully. This is a random partition with $\Pr(\{0\}, \{1\}) = q$ and $\Pr(\{0, 1\}) = 1 - q$. Suppose we enlarge the state space to be $S \times X$, where $X = \{\text{success, failure}\}$. While on this state space the DM's learning is described by a partition, acts that condition on X may not be available: it is plausible that the payoff of drilling rights does not depend on the success or failure of the test drill, but only on the presence of oil. Under our assumptions, the domain of acts that are defined on S is sufficient to allow the description of expected information as events.

4 Related literature

Our model builds on a general subjective-learning representation of preferences over menus of acts, which identifies signals with posterior distributions over states (Definition 6 in Sect. 2.2.2). Such representation was first derived in DLST. They further derived a partitional learning representation, in which signals correspond to events that partition the state space. Partitional learning is also studied in De Oliveira et al. (2016), who adopt the same menu choice approach but, unlike us, allow the DM to choose his information structure, and in Lu (2016), who uses the random choice approach to elicit information. A dynamic version of partitional learning appears in Dillenberger et al. (2017).

In terms of generality, the generalized-partition representation lies in between the two extreme cases studied in DLST. For a Bayesian DM, a salient feature of partitional learning is that the relative weights of states in the learned event do not shift. In this paper, we show that this non-shifting-weights property corresponds to stable behavior, which links ex ante and (anticipated) ex post rankings of acts. We then identify the *least restrictive* model with this property. Sections 2.2.2 and 3 demonstrate the merit of the generality permitted by our model (over partitional learning). In "Appendix A,"

[14] Gilboa et al. (2009, 2012) point out the problems involved in using an analytical construction, according to which states are defined as functions from acts to outcomes, to generate a state space that captures all conceivable sources of uncertainty. First, since all possible acts on this new state space should be considered, the new state space must be extended yet again, and this iterative procedure does not converge. Second, the constructed state space may include events that are never revealed to the DM, and hence, some of the comparisons between acts may not even be potentially observable. A related discussion appears in Gilboa (2009, Section 11.1).

we show how one can strengthen our main axiom to give a different characterization of partitional learning than the one provided in the aforementioned papers.

Theorem 1 studies stable behavior, which directly links ex ante and actual ex post preferences. Breaking this requirement into two parts, violations of the axiom might emerge from either (i) a violation of dynamic consistency (i.e., a discrepancy between ex ante preferences and anticipated ex post preferences), or (ii) a difference between anticipated ex post preferences and actual ex post preferences. Siniscalchi (2011), for example, explicitly focuses on (ii).[15] In contrast, we start with a Bayesian (and hence dynamically consistent) expected utility maximizer and investigate the discrepancy in (i). Theorem 2 establishes it arises if the underlying information structure is not a generalized partition.

Appendices

A. Comparison with partitional learning

We now show how Axiom 2 can be modified to capture the idea of deterministic signals and thereby to characterize partitional learning. The axiomatization here is different than the one provided in DLST.

Axiom 3 (Partition) *For any two acts $f, g \in \mathcal{F}_+$ and irreducible event $I \subseteq S$,*

$$\{fI0\} \succeq \{gI0\} \Rightarrow \{fI0, gI0\} \sim \{fI0\}.$$

In the words of Kreps (1979), Axiom 3 can be viewed as a requirement of strategic rationality given an event. Recall that if I is irreducible, then the DM can foresee learning an event J that (weakly) contains I. If signals are deterministic and I is irreducible, then there is no $I' \subset I$ the DM also foresees learning. It is thus obvious why for a Bayesian DM partitional learning satisfies this property; $fI0$ and $gI0$ agree outside I, and if $fI0$ does better on I, it will do better on any superset of I. Conversely, if information in the subjective-learning representation of Definition 6 is not partitional, then there are two posteriors $\pi, \pi' \in \text{supp}(p)$, such that $\text{supp}(\pi) = I$, $\text{supp}\left(\pi'\right) = I'$, $I \neq I'$, and $I \cap I' \neq \phi$. Let $f = c$ and $g = c + \varepsilon$. Without loss of generality, suppose $I \cap I' \neq I$. If $\varepsilon > 0$ and small enough, then $\{fI0\} \succeq \left\{g\left(I \cap I'\right)0\right\}$ while, since $g\left(I \cap I'\right)0$ does better than $fI0$ conditional on I', $\left\{fI0, g\left(I \cap I'\right)0\right\} \succ \{fI0\}$.

Since larger sets are better (preference for flexibility), $\{fI0\} \cup \left\{gI'0 \,\middle|\, I' \subset I\right\} \succeq \{fI0\}$. Further, by payoff domination, $\{fI0, gI0\} \succeq \{fI0\} \cup \left\{gI'0 \,\middle|\, I' \subset I\right\}$. Therefore, Axiom 3 implies Axiom 2, while, as is obvious from our results, the converse is false.

B. Proof of Theorem 1

Let $\Sigma \subset 2^S$. We say that r satisfies the *non-shifting-weights* property, if for $s, s' \in I \in \Sigma$,

[15] Siniscalchi (2011) refers to anticipated ex post preferences as *conjectural* ex post preferences.

$$\frac{\mu(s)}{\mu(s')} = \frac{\Pr(s\,|I)}{\Pr(s'\,|I)}. \tag{2}$$

The DM's ex ante preferences on acts, $\succeq^o$, are represented by $V(f) = \sum_{s\in S} f(s)\,\mu(s)$, and his ex post preferences, $\succeq^o_I$, are represented by $V(f\,|I) = \sum\limits_{s\in I} f(s)\Pr(s\,|I)$.

To see that the non-shifting-weights property is equivalent to statement (1) in the theorem, fix $s' \in I$ and set $\kappa := \frac{\Pr(s'|I)}{\mu(s')}$. It follows from Eq. (2) that

$$\sum_s f(s)\Pr(s\,|I) = \kappa \sum_{s\in I} f(s)\,\mu(s).$$

Consider $f, g \in \mathcal{F}$ and $I \in \Sigma$ such that $f(s) \neq g(s)$ only if $s \notin I$. Then

$$\begin{aligned} f \succeq^o g &\Leftrightarrow V(f) \geq V(g) \Leftrightarrow \sum_{s\in S} f(s)\,\mu(s) \geq \sum_{s\in S} g(s)\,\mu(s) \\ &\Leftrightarrow \sum_{s\in I} f(s)\,\mu(s) \geq \sum_{s\in I} g(s)\,\mu(s) \Leftrightarrow \kappa \sum_{s\in I} f(s)\,\mu(s) \geq \kappa \sum_{s\in I} g(s)\,\mu(s) \\ &\Leftrightarrow \sum_s f(s)\Pr(s\,|I) \geq \sum_s g(s)\Pr(s\,|I) \Leftrightarrow V(f\,|I) \geq V(g\,|I) \Leftrightarrow f \succeq^o_I g \end{aligned}$$

To see that the non-shifting-weights property is equivalent to statement (2) in the theorem, note that for information structure r with $\Sigma \subset 2^S$, Bayes' law implies that for any $s, s' \in I$,

$$\frac{\Pr(s\,|I)}{\Pr(s'\,|I)} = \frac{r_s(I)\,\mu(s)\,/\mu(I)}{r_{s'}(I)\,\mu(s')\,/\mu(I)}$$

independently of I. Hence, the non-shifting-weights property holds if and only if $\frac{r_s(I)}{r_{s'}(I)} = 1$ for all $s, s' \in I$. If there is a generalized partition ρ such that $r_s \equiv \rho_s$, then this holds by definition, as $r_s(I) = r_{s'}(I) = \rho(I)$. Conversely, if the information structure r is not a generalized partition, then there exist a signal $\sigma \in \Sigma$ and two states s, s' with $r_s(\sigma) > 0$ and $r_{s'}(\sigma) > 0$, such that $r_s(\sigma) \neq r_{s'}(\sigma)$. For these two states, the non-shifting-weights property does not hold.

C. Proof of Theorem 2

Necessity Suppose $\succeq$ admits a generalized partition representation (μ, ρ), that is,

$$V(F) = \sum_{I\in 2^S} \max_{f\in F} \left[\sum_{s\in I} f(s)\,\mu(s)\right] \rho(I)$$

represents $\succeq$.

Claim 1 *Event $I \subseteq S$ is irreducible if and only if $\rho(J) > 0$ for some $J \supseteq I$.*

Proof This follows immediately from the following two observations:

1. For any J with $J \cap I \neq I$,

$$\max_{f \in \{cI'0 | I' \subset I\}} \left[\sum_{s \in J} f(s)\,\mu(s)\right] = c \sum_{s \in J \cap I} \mu(s) = \sum_{s \in J} cI0(s)\,\mu(s);$$

2. Because $\operatorname{supp}(\mu) = S$, $J \supseteq I$ implies that

$$\max_{f \in \{cI'0 | I' \subset I\}} \left[\sum_{s \in J} f(s)\,\mu(s)\right] < c \sum_{s \in I} \mu(s) = \sum_{s \in J} cI0(s)\,\mu(s).$$

□

Consider an irreducible event I and two acts $f, g \in \mathcal{F}_+$. According to the generalized-partition representation,

$$\{fI0\} \succeq \{gI0\} \Leftrightarrow \sum_{s \in I} \left[f(s)\,\mu(s) \sum_{I \subseteq S} \rho_s(I)\right] \geq \sum_{s \in I} \left[g(s)\,\mu(s) \sum_{I \subseteq S} \rho_s(I)\right]$$

Since ρ is a generalized partition, $\sum_{I \subseteq S} \rho_s(I) = 1$ for all $s \in S$, and hence $\{fI0\} \succeq \{gI0\} \Leftrightarrow \sum_{s \in I} f(s)\,\mu(s) \geq \sum_{s \in I} g(s)\,\mu(s)$. Therefore, $\{fI0\} \succeq \{gI0\}$ implies both (i) that for any event $J \supseteq I$ with $J \in \operatorname{supp}(\rho)$ we have

$$\max_{f' \in \{fI0, gI0\}} \left[\sum_{s \in J} f'(s)\,\mu(s)\right] = \max\left[\sum_{s \in I} f(s)\,\mu(s), \sum_{s \in I} g(s)\,\mu(s)\right] =$$
$$= \sum_{s \in I} f(s)\,\mu(s) = \max_{f' \in \{fI0\} \cup \{gI'0 | I' \subset I\}} \left[\sum_{s \in J} f'(s)\,\mu(s)\right];$$

and (ii) that for any event $J \in \operatorname{supp}(\rho)$ with $J \cap I \neq I$ we have

$$\max_{f' \in \{fI0, gI0\}} \left[\sum_{s \in J} f'(s)\,\mu(s)\right] = \max_{f' \in \{fI0\} \cup \{gI'0 | I' \subset I\}} \left[\sum_{s \in J} f'(s)\,\mu(s)\right].$$

Therefore, $\{fI0\} \succeq \{gI0\} \Rightarrow \{fI0, gI0\} \sim \{fI0\} \cup \left\{gI'0 \,\middle|\, I' \subset I\right\}$.

Sufficiency Suppose that $\succeq$ admits a subjective-learning representation and satisfies Axiom 2. For each $s \in S$, let $\mu(s) = \int_{\operatorname{supp}(p)} \pi(s)\,\mathrm{d}p(\pi)$. Since the measure p over $\Delta(S)$ in a subjective-learning representation is unique, so is μ. It is easy to see that μ is a probability measure and that for each $g, f \in \mathcal{F}$, $\{f\} \succ \{g\} \Leftrightarrow \sum_{s \in S} f(s)\,\mu(s) > \sum_{s \in S} g(s)\,\mu(s)$.

Let:

$$\Pi_0 := \left\{\pi \,\middle|\, \exists s, s' \in \operatorname{supp}(\pi) \text{ with } \frac{\pi(s)}{\pi(s')} \neq \frac{\mu(s)}{\mu(s')}\right\}$$

and for $\varepsilon > 0$

$$\Pi_\varepsilon := \left\{\pi \,\middle|\, \exists s, s' \in \operatorname{supp}(\pi) \text{ with } \frac{\pi(s)}{\pi(s')} \geq (1+\varepsilon)\frac{\mu(s)}{\mu(s')}\right\}.$$

Finally, for any $s, s' \in S$ let

$$\Pi_{s,s',\varepsilon} := \left\{\pi \,\middle|\, s, s' \in \text{supp}(\pi) \text{ and } \frac{\pi(s)}{\pi(s')} \geq (1+\varepsilon)\frac{\mu(s)}{\mu(s')}\right\}.$$

Claim 2 *If $p(\Pi_0) > 0$, then there exist $s, s' \in S$ and $\varepsilon > 0$, such that $p\left(\Pi_{s,s',\varepsilon}\right) > 0$.*

Proof First note that there exists $\varepsilon > 0$ such that $p(\Pi_\varepsilon) > 0$. If it was not the case, then there would be $(\varepsilon_n) \downarrow 0$ such that

$$p\left(\left\{\pi \,\middle|\, \exists s, s' \in \text{supp}(\pi) \text{ with } \frac{\pi(s)}{\pi(s')} \notin \left((1-\varepsilon_n)\frac{\mu(s)}{\mu(s')}, (1+\varepsilon_n)\frac{\mu(s)}{\mu(s')}\right)\right\}\right) = 0$$

for all n. But for any s, s' we would have $\left((1-\varepsilon_n)\frac{\mu(s)}{\mu(s')}, (1+\varepsilon_n)\frac{\mu(s)}{\mu(s')}\right) \to \frac{\mu(s)}{\mu(s')}$ and therefore $p(\Pi_0) = 0$, which is a contradiction.

Second, since S is finite we can write $p(\Pi_\varepsilon) = \sum_{s,s' \in S \times S} p\left(\Pi_{s,s',\varepsilon}\right)$. This immediately implies that at least for one pair s, s' we must have $p\left(\Pi_{s,s',\varepsilon}\right) > 0$. □

Now fix s, s' such that $p\left(\Pi_{s,s',\varepsilon}\right) > 0$. Let $f := \frac{1}{2}$, and for $\gamma, \beta > 0$ such that $\frac{\beta}{\gamma} \in \left(\frac{\mu(s)}{\mu(s')}, (1+\varepsilon)\frac{\mu(s)}{\mu(s')}\right)$, define an act g by

$$g(\widehat{s}) := \begin{cases} \frac{1}{2} + \gamma & \widehat{s} = s \\ \frac{1}{2} - \beta & \widehat{s} = s' \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

Note that for any I with $s, s' \in I$

$$\pi \in \Pi_{s,s',\varepsilon} \Rightarrow \frac{1}{2} < \sum_{\widehat{s} \in I} g(\widehat{s})\, \pi(\widehat{s}); \text{ and} \tag{3i}$$

$$\frac{1}{2} > \sum_{\widehat{s} \in I} g(\widehat{s})\, \mu(\widehat{s}). \tag{3ii}$$

Let $\mu_{s,s',\varepsilon}(s) := \int_{\Pi_{s,s',\varepsilon}} \pi(s)\, \mathrm{d}p(\pi)$. We now argue that there exists an irreducible $I \subseteq \text{supp}\left(\mu_{s,s',\varepsilon}\right)$ such that

(a) $f \not\succeq_I g$;
(b) $\{fI0\} \succeq \{gI0\}$.

First, $s, s' \in \text{supp}(\pi)$ for every $\pi \in \Pi_{s,s',\varepsilon}$. Because S is finite, $\Pi_{s,s',\varepsilon}$ can be partitioned into finitely many sets of posteriors, where π and π' are in the same cell of the partition if and only if they have the same support. At least one of these cells must have positive weight under p. Choose I to be the support of the posteriors in such a cell. By (3i), we have $\{fI0, gI0\} \succ \{fI0\} \cup \left\{gI'0 \,\middle|\, I' \subset I\right\}$, or $f \not\succeq_I g$. This establishes part (a). Part (b) is immediately implied by (3ii).

The combination of (a) and (b) constitutes a violation of Axiom 2. Therefore, $p(\Pi_0) = 0$, that is, for all but a measure zero of posteriors, $\pi \in \text{supp}(p)$ and $I \subset S$ imply that

$$\pi(s) = \begin{cases} \frac{\mu(s)}{\mu(I)} & s \in I \\ 0 & \text{otherwise} \end{cases}$$

Note that this result implies that if $\pi, \pi' \in \text{supp}(p)$ are such that $\pi \neq \pi'$ then $\text{supp}(\pi) \neq \text{supp}(\pi')$. Therefore, we can index each element $\pi \in \text{supp}(p)$ by its support $\text{supp}(\pi) \in 2^S$ and denote a typical element by $\pi(\cdot\,|I)$, where $\pi(s\,|I) = 0$ if $s \notin I \in 2^S$. There is then a unique measure $\widehat{p}$ such that for all $F \in \mathcal{K}(\mathcal{F})$

$$V(F) = \sum_{I \in 2^S} \max_{f \in F} \left[\sum_{s \in S} f(s)\, \pi(s\,|I)\right] \widehat{p}(I), \tag{3}$$

and $\mu(s) = \sum_{I | s \in I} \pi(s\,|I)\, \widehat{p}(I)$.

We have already established that $\pi(s\,|I) = \frac{\mu(s)}{\mu(I)}$ for all $s \in I \in \text{supp}(\widehat{p})$. Define $\rho(I) := \frac{\widehat{p}(I)}{\mu(I)}$ and substitute $\mu(s)\,\rho(I)$ for $\pi(s\,|I)\,\widehat{p}(I)$ in (3). Bayes' law implies that

$$\rho_s(I) := \begin{cases} \rho(I) & \text{if } s \in I \\ 0 & \text{if } s \notin I \end{cases}$$

is indeed a probability measure for all s.

D. Proof of Theorem 3

(*if*) Let Ψ be a uniform cover of S, where β and k satisfy Definition 7. Set $\rho(I) = \frac{\beta(I)}{k}$ for all $I \in \Psi$.

(*only if*) Suppose that $\rho : 2^S \to [0, 1]$ is a generalized partition, with $\text{supp}(\rho) = \Psi$. In addition to $\rho(I) = 0$ for $I \notin \Psi$, the conditions that ρ should satisfy can be written as $\mathbf{A}\rho_\Psi = 1$, where $\mathbf{A}$ is a $|S| \times |\Psi|$ matrix with entries $a_{i,j} = \left\{\begin{matrix} 1 & s \in I \\ 0 & s \notin I \end{matrix}\right|$, ρ_Ψ is a $|\Psi|$-dimensional vector with entries $(\rho(I))_{I \in \Psi}$, and $\mathbf{1}$ is a $|S|$-dimensional vector of ones.

Suppose first that $\rho(I) \in \mathbb{Q} \cap (0, 1]$ for all $I \in \Psi$. Rewrite the vector ρ_Ψ by expressing all entries using the smallest common denominator, $\xi \in \mathbb{N}_+$. Then Ψ is a generalized partition of size ξ. To see this, let $\beta(I) := \xi\rho(I)$ for all $I \in \Psi$. Then $\sum_{I \in \Psi | s \in I} \beta(I) = \sum_{I \in \Psi | s \in I} \xi\rho(I) = \xi$ for all $s \in S$.

It is thus left to show that if $\rho_\Psi \in (0, 1]^{|\Psi|}$ solves $\mathbf{A}\rho_\Psi = \mathbf{1}$, then there is also $\rho'_\Psi \in [\mathbb{Q} \cap (0, 1]]^{|\Psi|}$ such that $\mathbf{A}\rho'_\Psi = \mathbf{1}$.

Let $\widehat{P}$ be the set of solutions for the system $\mathbf{A}\rho_\Psi = \mathbf{1}$. Then, there exists $X \in \mathbb{R}^k$ (with $k \leq |\Psi|$) and an affine function $f : X \to \mathbb{R}^{|\Psi|}$ such that $\widehat{\rho_\Psi} \in \widehat{P}$ implies $\widehat{\rho_\Psi} = f(x)$ for some $x \in X$. We first make the following two observations:

(i) There exists f as above, such that $x \in \mathbb{Q}^k$ implies $f(x) \in \mathbb{Q}^{|\Psi|}$;
(ii) There exists an open set $\widetilde{X} \subseteq \mathbb{R}^k$ such that $f(x) \in \widehat{P}$ for all $x \in \widetilde{X}$

To show (i), apply the Gauss elimination procedure to get f and X as above. Using the assumption that $\mathbf{A}$ has only rational entries, the Gauss elimination procedure (which involves a sequence of elementary operations on $\mathbf{A}$) guarantees that $x \in \mathbb{Q}^k$ implies $f(x) \in \mathbb{Q}^{|\Psi|}$.

To show (ii), suppose first that $\rho^* \in \widehat{P} \cap (0,1)^{|\Psi|}$ and $\rho^*_\Psi \notin \mathbb{Q}^{|\Psi|}$. By construction, $\rho^*_\Psi = f(x^*)$, for some $x^* \in X$. Since $\rho^*_\Psi \in (0,1)^{|\Psi|}$ and f is affine, there exists an open ball $B_\varepsilon(x^*) \subset \mathbb{R}^k$ such that $f(x) \in \widehat{P} \cap (0,1)^{|\Psi|}$ for all $x \in B_\varepsilon(x^*)$, and in particular for $x' \in B_\varepsilon(x^*) \cap \mathbb{Q}^k$ $(\neq \phi)$. Then $\rho'_\Psi = f\left(x'\right) \in [\mathbb{Q} \cap (0,1]]^{|\Psi|}$. Lastly, suppose that $\rho^*_\Psi \in \widehat{P} \cap (0,1]^{|\Psi|}$ and that there are $0 \leq l \leq |\Psi|$ sets $I \in \Psi$, for which $\rho(I)$ is uniquely determined to be 1. Then set those l values to 1 and repeat the above procedure for the remaining system of $|\Psi| - l$ linear equations.

References

Anscombe, F.J., Aumann, R.J.: A definition of subjective probability. Ann. Math. Stat. **34**, 199–205 (1963)

De Oliveira, H., Denti, T., Mihm, M., Ozbek, M.K.: Rationally inattentive preferences. Theor. Econ. **12**, 621–654 (2016)

Dillenberger, D., Sadowski, P.: Subjective learning. PIER Working Paper Number 12-007 (2012)

Dillenberger, D., Lleras, J.S., Sadowski, P., Takeoka, N.: A theory of subjective learning. J. Econ. Theory **153**, 287–312 (2014)

Dillenberger, D., Krishna, R.V., Sadowski, P.: Subjective Information Choice Processes. Mimeo, New York (2017)

Ghirardato, P.: Revisiting Savage in a conditional world. Econ. Theor. **20**, 83–92 (2002). https://doi.org/10.1007/s001990100188

Gilboa, I.: Theory of Decision under Uncertainty. Cambridge University Press, Cambridge (2009)

Gilboa, I., Postlewaite, A., Schmeidler, D.: Is it always rational to satisfy Savage's axioms? Econ. Philos. **25**, 285–296 (2009)

Gilboa, I., Postlewaite, A., Schmeidler, D.: Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality. Synthese **187**(1), 11–31 (2012)

Kreps, D.: A representation theorem for preference for flexibility. Econometrica **47**, 565–577 (1979)

Lu, J.: Random choice and private information. Econometrica **84**(6), 1983–2027 (2016)

Savage, J.: The Foundations of Statistics. Wiley, New York (1954)

Shapley, L.S.: On balanced sets and cores. Naval Res. Logistics Q. **14**, 453–460 (1967)

Siniscalchi, M.: Dynamic choice under ambiguity. Theor. Econ. **6**(3), 379–421 (2011)