Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Economic Theory 111 (2003) 49–87
JOURNAL OF Economic Theory
http://www.elsevier.com/locate/jet

Probabilistic representation of complexity

Nabil I. Al-Najjar,[a,*] Ramon Casadesus-Masanell,[b] and Emre Ozdenoren[c]

[a] Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA
[b] Harvard Business School, Boston, MA 02163, USA
[c] Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220, USA

Received 4 January 2000; final version received 25 March 2002

Abstract

We introduce a framework to study individuals' behavior in environments that are deterministic, but too complex to permit tractable deterministic representations. An agent in these environments uses a probabilistic model to cope with his inability to think through all contingencies in advance. We interpret this probabilistic model as embodying all patterns the agent perceives, yet allowing for the possibility that there may be important details he had missed. Although the implied behavior is rational, it is consistent with an agent who believes his environment is too complex to warrant precise planning, foregoes finely detailed contingent rules in favor of vaguer plans, and expresses a preference for flexibility.

JEL classification: D81; D80; L14

Keywords: Complexity; Cognitive uncertainty; Bounded rationality; Incomplete contracts

1. Introduction

A remarkable aspect of human behavior is individuals' ability to cope with complexity. The role of complexity in everyday life is almost self-evident: people engage in complex social and economic interactions where it is pointless to examine every possible contingency. Yet individuals somehow manage to formulate coherent plans of action in situations they recognize as too complex to permit a full analysis.

*Corresponding author. Fax: +1-847-467-1220.
E-mail addresses: al-najjar@nwu.edu (N.I. Al-Najjar), rmasanell@hbs.edu (R. Casadesus-Masanell), emreo@umich.edu (E. Ozdenoren).

doi:10.1016/S0022-0531(03)00075-9

Understanding this behavior is a key problem in cognitive sciences and is of obvious importance in economic and strategic contexts.

While its importance and pervasiveness is beyond dispute, complexity is a difficult and elusive concept to formally model. For complexity considerations to matter, the agent's cognitive abilities must be limited relative to his environment. One way to introduce such gap is to assume "bounded rationality," preventing the agent from identifying the optimal course of action in a given environment. The environments considered are usually simple enough that the optimal course of action can be easily figured out, leaving the "boundedly rational" agent vulnerable to arbitrage opportunities and money-pumps that are incompatible with equilibrium analysis. The challenge facing such approach is reconciling "bounded rationality" with standard economic and game theoretic practice of requiring that behavior does not display exploitable systematic errors.

Models of complexity that preclude systematic errors face the seemingly paradoxical requirement of having to formally model environments that rational agents are unable to fully describe. This paper proposes an approach that produces interesting cognitive limitations due to complexity, yet behavior is rational in the sense of being consistent with learning from data and displays no exploitable systematic errors.

1.1. Outline of the model

We consider an agent facing a "problem," a loose collection of situations or contingencies he views as sharing a common structure. Examples of problems are: taking an appropriate action on behalf of a superior, adjudicating a legal case, categorizing an object, diagnosing a disease, etc.

The agent does not confront the problem at this level of abstraction. Rather, he faces specific *contingencies* or *instances* drawn at random according to a known probabilistic process (e.g., a specific legal case, a specific patient). For simplicity, we focus on "categorization problems." In such problems, the agent places contingencies into a predetermined set of categories, receiving 1 if he matches the correct category and 0 otherwise. Each contingency is characterized by a countable set of objective and observable conditioning variables, representing the data on which the agent's decisions may be based.

Ex ante contingent plans are represented as *rules* that map contingencies into actions. We also allow the agent to retain flexibility through the use of *options*. These are mappings from contingencies to *sets* of actions, capturing the idea that the agent gets to choose an action after observing the specific contingency he ends up facing. The main restriction is that rules and options must be formulated in terms of a finite number of features.

We focus on behavior that displays what we view as central aspects of a rational agent's response to a complex environment: (1) the agent has coherent assessment of the expected payoff of rules and options; (2) ex post, once told which contingency he is actually facing, it is obvious what the optimal action should be; yet (3) the agent is unable to describe ex ante a rule specifying the optimal action in every possible contingency. Far from odd or pathological, this seemingly contradictory behavior is pervasive in decision-making, planning in intertemporal settings, categorization and pattern recognition.[1]

We interpret such behavior in terms of a probabilistic model that represents the agent's theory of how his environment works. We construct a probability space with a complete state space where the agent's utility at each contingency is random. Theorems 1, 2, 6 and 7 roughly state that an agent can consistently evaluate rules and options if and only if he has an essentially unique probabilistic model in which rules and options are evaluated according to their expected utility.

We make considerable effort to isolate complexity as the motive for the agent's choices. Standard models of incomplete information also limit detailed contingent planning and so may appear superficially similar to the complexity-driven behavior studied here. But with incomplete information, the limitation on contingent planning is the result of an exogenous restriction on the sets of events on which choices may depend. This seems very different from problems where the agent has access to a vast amount of conditioning information, but this information is too complex to use in formulating a detailed plan that anticipates all possible contingencies.

The probabilistic models we construct separate complexity from other motives of behavior by displaying two important properties. The first, *instance specific knowledge*, says that the agent has all payoff relevant information. This rules out incomplete information as explanation of the agent's behavior. The second, *exhaustive introspection*, says that the agent optimally uses every possible information available to improve his decisions. This assumption formalizes the idea that the agent's behavior is not due to "bounded rationality," but to the inherent complexity of the problem he is facing.

In Section 6.2 we provide a formal sense in which the agent's probabilistic model represents his "steady-state' theory of the environment after all useful learning has taken place. In particular, the choices implied by that model would persist even with continued arrival of new data. Finally, in Section 8 we provide an application of our model to a simple delegation problem between a principal and an agent.

1.2. Interpretation

We now provide an informal intuition for our model. An agent facing a problem may initially be tempted to conduct an exhaustive enumeration of all contingencies and the corresponding optimal action to be taken at each of these contingencies. But there is a vast number of possible contingencies, each occurring with zero probability. This makes a crude representation based on exhaustive enumeration pointless. The agent thus seeks decision rules based on generalizations that can help predict the optimal action based on observable features of contingencies. We have in mind decision rules built up from statements of the form "take action b when

[1] A famous example is Justice Stewart's famous statement confessing that he is unable to come up with a good definition of obscenity, "But I know it when I see it." (Jacobellis vs. Ohio, 378 US 184, 197 (1964); Stewart, J., concurring).

encountering a contingency satisfying a given list of properties.'' Here, ''a list of properties'' refers to a *finite* set of conditions formulated in terms of objectively given features.

The fact that there is a large number of contingencies, by itself, does not imply that the environment is complex. For example, the agent may believe that one particular feature, i, perfectly predicts the optimal action in every contingency. In this case, the agent's model of his environment reduces to a simple similarity judgment that all contingencies with feature i are alike. The optimal decision rule in this case takes an obvious form: ''take action a in every contingency with feature i, and take action b otherwise.'' Although the number of possible contingencies may be large (indeed infinite), the agent in this environment has no difficulty identifying the optimal decision rule, and the problem is not a complex one.

What makes a problem complex is that contingencies have unique, idiosyncratic features that cannot be easily captured by simple generalizations like the one in the last paragraph. In such problems, every rule the agent comes up with has too many exceptions to allow for correct ex ante contingent planning. Our view is that the agent deals with such problems by representing his environment probabilistically, even though he realizes that it is in fact deterministic. The agent's probabilistic model captures all systematic patterns he perceives in his environment in the sense that he views residual unexplained variations as independent and so have no further useful structure. These residual variations reflect the agent's realization that the patterns his probabilistic model embodies miss potentially important variations due to the unique nature of each contingency.

The behavior we describe is that of a fully rational agent in the traditional sense: he optimizes given a coherent model of the environment; he understands and takes advantage of all the implications of this model, unhampered by limitations that prevent him from carrying out reasoning that we, as modelers, can perform.

Despite this, behavior may display many features traditionally considered the hallmark of ''bounded rationality:'' the agent may think the environment is too complex to permit an exhaustive ex ante contingent plan, yet once a specific contingency is known, what to do becomes ''obvious.'' The agent may forego finely detailed planning for future contingencies, choosing to rely instead on coarser and vaguer plans of actions whose details are filled later.

In summary, the agent in our model optimizes, but he does so relative to a theory of his environment that he recognizes to be imperfect. Since optimization is not sacrificed, closed equilibrium analysis is possible without ad hoc restrictions on agents' abilities (see Section 8 for an illustration). This is a considerable advantage in multi-agent settings because it allows predictions that do not hinge on requiring agents to be dumb, ignore potentially useful information, or use ad hoc rules that the modeler or a rival could easily exploit.

1.3. Related literature

This paper relates to several strands of literature. The issues addressed in this paper have the flavor of those appearing in the literature on bounded rationality (see

[13] for survey of this vast literature). An important point to stress is that we do not develop a procedural model of decision-making; we make no attempt to answer questions like: "how do people think?' Rather, we provide a model of behavior displaying as much interesting cognitive limitations as consistent with the level of rationality needed for standard equilibrium analysis.

To derive interesting cognitive limitations, we need to restrict the set of decision rules available. Here we impose the mild (arguably, the mildest) restriction that decision rules be computable, i.e., correspond to procedures that can be effectively carried out. This follows several authors, including Binmore [3], Gilboa and Schmeidler [7], Anderlini and Felli [1] and Anderlini and Sabourian [2], among others. The closest of these to the model of this paper is by Anderlini and Felli [1] who consider contracting problems.

Another strand of literature is that emerging from Kreps's [11,12] interpretation of preference for flexibility in terms of an agent's inability to foresee future contingencies. Works in this vein include [4,14,15]. We discuss the link with this literature in Section 6.1.

Complexity, of course, is a central issue in cognitive sciences.[2] Our formal model of individuals facing contingencies of a problem, each identified with a potentially large collection of features, shares many similarities with learning models in the pattern recognition literature (e.g., [5]). Our focus, however, is quite different. Much of this literature is concerned with procedural models, and especially on the convergence properties of various classes of algorithms. Here we take the behavior of a fully rational agent as given, and try to infer this agent's model of his environment, what he considers complex and simple, and how this impacts his decisions.

2. The model

We develop a formal framework of an agent facing a "problem," a collection of situations, or contingencies, he views as sharing a common structure. Examples of problems are: taking an appropriate action on behalf of a superior, adjudicating a legal case, categorizing an object, etc. Formally, a problem consists of: a set of *contingencies*, or *instances*, X; a *probabilistic process* generating data, λ; a finite set of *actions*, B; a finite set of *ex post utilities*, V; an algebra of *conditioning events*, $\mathscr{A}$. We describe each in detail below.

2.1. Contingencies, actions and utilities

A problem is a large collection of *contingencies* or *instances*. Each such contingency x consists of an exhaustive set of objective data, or conditioning information, the agent may use in his decisions. We formalize this by defining the set of contingencies in terms of *features*, with the ith feature, x^i, taking values in a finite

[2] See, for example, [8].

set X^i. For notational simplicity, and without loss of generality, we assume that the features are binary, i.e., $X^i = \{0, 1\}$.

We do not want to rule out an agent who believes the problem he is facing is subject to a "curse of dimensionality:" for any set of features considered, there may be other relevant dimensions he may have failed to take into account. We therefore allow a countable number of features. Situations with a predetermined, finite number of relevant features may be incorporated as a special case.[3] The set of contingencies is then the product $X = \prod_{i=1}^{\infty} X^i$; with each contingency corresponding to a sequence $(x^1, x^2, \ldots)$ of values of the features.[4]

The agent chooses an action $b \in B = \{b_1, \ldots, b_K\}$ upon observing a contingency $x \in X$. There is a finite set V such that each $v \in V$ is a utility function of the form $v : B \to \mathbb{R}$. We later interpret such v's as state-dependent, ex post rankings of actions.

Throughout the paper we shall focus on the special class of *categorization problems* in which the agent places contingencies into a predetermined set of categories, receiving 1 if he matches the correct category and 0 otherwise.[5] In a categorization problems the set of actions B is a set of K categories, with the interpretation of action b_k at x as "x is of category k," and payoffs satisfy:

A.0. For every $v_k \in V$, $v_k(b_k) = 1$ and $v_l(b_k) = 0$ for all $l \neq k$.

We shall assume A.0 throughout the paper, unless we specifically indicate otherwise.

Categorization is one of the most basic and pervasive aspects of decision-making. Examples range from mundane tasks in pattern recognition (e.g., "Is a given object a chair or a table?") to more complex decision-making and problem-solving activities. In a sense, categorization appears in virtually every decision problem, since decisions are often made based on concepts and categories, rather than raw data.[6] In economic contexts, managerial decisions may be viewed as attempts at categorizing or classifying situations into predetermined, prototypical classes. See [9] for discussion and references.

2.2. Ex ante vs. ex post decisions

We distinguish between the *instance-specific*, or *ex post* decision in which the agent has to take an action while facing specific contingency, and the *ex ante* decision in which he formulates a state-contingent plan before knowing the specific contingency he will be facing later on.

Our goal is to formalize the idea that a problem is complex if deciding ex ante (which requires considering all contingencies of the problem in advance) is

[3] The case in which X is finite will turn out to be trivial from complexity point of view given our other assumptions.

[4] It is sometimes useful to think of X as the set of binary expansions (sequences of 0's and 1's) of numbers in the interval $[0, 1]$, though this may be misleading. We attach no particular meaning to the ordering of features, so the "location" of an x on the interval is of no relevance.

[5] Section 8 illustrates that the model generalizes to broader contexts.

[6] This is the prevailing view in cognitive sciences; see, for example, [8].

potentially much harder than deciding ex post, once a specific contingency is known. For example, designing a contract that anticipates all future contingencies is harder than identifying the optimal ex post action given a specific contingency. Another example is that formulating rules to define such mundane categories of objects as "chairs" and "tables" can be surprisingly difficult, even though most people would have no difficulty identifying the correct category once they see a specific object.

We capture these ideas formally through the concepts of rules and options which we now introduce.

2.2.1. Rules

A rule is an explicit list of instructions that assigns to each contingency x an action $f(x)$ to be taken should x arise. For a function f to be a rule, it should be possible, at least in principle, for f to be codified, written down, communicated, instructed to a subordinate, and re-produced. The intuitive idea of a "rule" should thus preclude whim, gut-feeling, oracles, or other subjective criteria that cannot be made explicit.

It is helpful to think of the agent as having a "language" that enables him to refer to (and condition his actions on) elementary sets of "contingencies that have (do not have) a particular feature i."[7] Formally, an elementary set is of the form $\{x : x^i = j\}$, $i \in \{1, 2, \ldots\}$ and $j \in \{0, 1\}$. A set $A \subset X$ is *finitely-defined* if it belongs to the algebra $\mathscr{A}$ generated by all elementary sets. Roughly, any $A \in \mathscr{A}$ can be defined in terms of finite conjunctions and disjunctions of elementary sets.[8]

A *rule* is any function $f : X \to B$ measurable with respect to $\mathscr{A}$. In words, a rule is a conditional statement that describes the action to be taken at each contingency x as a function of its features.[9]

2.2.2. Options

We formulate the agent's ex post choices using an idea due to Kreps [12] which consists of offering the agent a state-contingent *option* to make a decision after a specific contingency is realized.[10] Formally, an *option* is any function $g : X \to 2^B \backslash \emptyset$ measurable with respect to $\mathscr{A}$.[11] The interpretation of an option g is: "you may wait

[7] For example, if feature 1 corresponds to whether the color of an object is blue or not, then $\{x : x^1 = 1\}$ corresponds to the sentence "objects that are blue", and $\{x : x^1 = 0\}$ corresponds to "objects that are not blue."

[8] It can be shown that $A \in \mathscr{A}$ if and only if it is *algorithmic*, i.e., there exists a Turing machine, or a computer program, α with input x and output in $\{0, 1\}$ such that for every x, $\alpha(x) = 1$ if $x \in A$ and 0 otherwise. The idea that agents are limited to procedures (decision rules, contracts, etc.) that can be carried out "effectively" (and so must be algorithmic) is found in, among others, [1,3,7]. (See [1,7] for a detailed formalism of computability). That every set in $\mathscr{A}$ is algorithmic is obvious. The argument that an algorithmic set A must be in $\mathscr{A}$ hinges on the assumption that the algorithm defining A always halts—so for any input x it never looks beyond a predetermined finite set of features. The compactness of X then ensures that one can find a uniform bound over all inputs.

[9] For example, a contract to undertake an action depending on feature i takes the general form: "in contingencies x that have feature i take action b, otherwise take action b'."

[10] Kreps does not use the term "option," but the idea is the same.

[11] That is, $2^B \backslash \emptyset$ is the set of all non-empty subsets of B.

until a contingency x is drawn, then choose any action b you like subject to the constraint $b \in g(x)$." Note that options can be used to represent varying degrees of flexibility (e.g., consider an option allowing full flexibility ($g(x) = B$) on some set A, but much less flexibility outside it).

Let $\mathscr{F}$ and $\mathscr{G}$ denote the sets of rules and options, respectively. Note that $\mathscr{F} \subset \mathscr{G}$: rules are trivial sort of options requiring rigid, precise planning at each contingency. Rules therefore preclude the vagueness or incompleteness characteristic of many decision processes.

2.3. *The space of problems*

Let $\overline{\mathscr{A}}$ denote the σ-algebra generated by $\mathscr{A}$. Then $(X, \overline{\mathscr{A}})$ is a standard measure space whose properties are characterized in Appendix A.

Contingencies are drawn from X according to a probability distribution λ. We assume that λ is atomless and satisfies $\lambda(A) > 0$ for all non-empty $A \in \mathscr{A}$. (In what follows, to avoid redundancy, we shall limit attention to non-empty subsets of $\mathscr{A}$.) The reader may assume, for concreteness, that λ is the uniform distribution on X.

It is often more convenient to work with $\bar{\mathscr{F}}$ and $\bar{\mathscr{G}}$, the sets of rules and options measurable with respect to $\overline{\mathscr{A}}$. Although a rule $f \in \bar{\mathscr{F}} \backslash \mathscr{F}$ is not necessarily finitely-defined, it is always possible to find a finitely-defined rule $f' \in \mathscr{F}$ that differs from f on a set of arbitrarily small measure.

3. Describing the agent's choices

We are interested in providing a formal model of an agent whose behavior displays a coherent assessment of the expected payoff of rules and options; ex post, once told which contingency he is actually facing, it is obvious what the optimal action should be; yet he is unable to describe a full ex ante rule specifying what the optimal action should be in every possible contingency.

Let $\Delta^{\mathbb{R}}$ denote the set of all probability distributions on the real line. We shall take as a primitive an agent's *utility distribution*:

$$U : \mathscr{G} \times \mathscr{A} \to \Delta^{\mathbb{R}}.$$

That is, we imagine that we can get the agent to reveal his payoff for any option $g \in \mathscr{G}$ and any set $A \in \mathscr{A}$. Note that we allow the agent to report a random payoff, i.e., $U(g, A)$ need not be a degenerate distribution. For example, when asked about the payoff of the constant action b_1, say, over X, the agent may reply that it is either 1 or 0 with equal probability. We do not collapse the distribution $U(g, A) \in \Delta^{\mathbb{R}}$ to its expectation (in the case of this example, we do not collapse the 50/50 lottery on 0 and 1 to its expectation of 0.5). Why we use distributions will become clearer in the sequel; for now, we just note that a distribution $U(g, A) \in \Delta^{\mathbb{R}}$ conveys more information about the agent's theory of how his environment works than the expected value $EU(g, A) \in \mathbb{R}$. In subsequent sections this additional information turns out to be essential to identifying a *unique* such theory, and to separate complexity as an explanation of behavior from other forces such as incomplete information, bounded rationality, and so on.

1. *The agent has coherent assessment of the average performance of contingent plans of action*: We formalize this by the following two conditions:

A.1*. For every $A \in \mathscr{A}$ there are weights $l = (l_1, \ldots, l_K) \in \Delta^V$ (depending on A) such that, for every $b_k \in B$,[12,13]

$$EU(b_k, A) = l_1 v_1(b_k) + \cdots + l_K v_K(b_k).$$

A.2*. For every $f \in \mathscr{F}$, $A \in \mathscr{A}$ and partition $\{A_1, \ldots, A_N\}$ of A by sets in $\mathscr{A}$,

$$EU(f, A) = \sum_{i=1}^{N} \lambda(A_i \mid A) EU(f, A_i).$$

In the context of categorization problems, this says that the agent can evaluate the frequencies of various categories over any subset $A \in \mathscr{A}$, and that he does so coherently across subsets. Note that these conditions do not say anything about how options are evaluated.

2. *Ex post*, *once told which contingency he is actually facing*, *it is obvious what the optimal action should be*. We formalize this by strengthening A.1* and A.2* to cover options:

A.1. For every $A \in \mathscr{A}$ there are weights $l = (l_1, \ldots, l_K) \in \Delta^V$ (depending on A) such that, for every $C \in 2^B \backslash \emptyset$,

$$EU(C, A) = l_1 \max_{b \in C} v_1(b) + \cdots + l_K \max_{b \in C} v_K(b).$$

A.2. *Coherence across sub-problems*: For every $g \in \mathscr{G}$, $A \in \mathscr{A}$ and partition $\{A_1, \ldots, A_N\}$ of A by sets in $\mathscr{A}$,

$$EU(g, A) = \sum_{i=1}^{N} \lambda(A_i \mid A) EU(g, A_i).$$

This says that once given a subset C to choose from, the agent evaluates it as if he knew the true payoff function v_i. For example, suppose there are two categories and the weights implied by A.1* over X are $l_1 = l_2 = 0.5$. We interpret this to mean that the agent believes that the two categories are equally likely over X. In particular, he knows he will face contingencies where b_1 is optimal half of the time, and

[12] For categorization problems the RHS of the following equation reduces to l_k. We maintain the redundant terms to facilitate comparison with Condition A.1 below.

[13] Δ^V denotes the set of probability distributions on the finite set V; we make the usual identification of V with the set vertices of Δ^V—i.e., the set of degenerate distributions on V.

contingencies where b_2 is optimal in the other half. If he is constrained to choose an uncontingent action (a constant rule) b_1, say, then he recognizes that this action will be wrong half of the time, and indeed under A.1* we must have $U(b_1, X) = 0.5$. Contrast this with the uncontingent option $C = B = \{b_1, b_2\}$. Condition A.1 implies that once the agent sees the actual contingency he is facing, he finds out its "type" (i.e., whether it is v_1 or v_2), then decides which action to take. Indeed $U(B, X) = 1$ in this case.

3. *Although the agent has thought through the problem, he is still unable to describe a rule specifying in advance the optimal action in every contingency.* This has two parts; the second part, stating that the agent "is unable to describe a rule specifying in advance the optimal action in every contingency," may be formalized as:

$$\sup_{f \in \mathscr{F}} EU(f, X) < 1.$$

To interpret this, let us first note that Condition A.1 always guarantees that the agent's expected payoff from the "grand option" $g(x) = B$ for all x is always 1 (i.e. $EU(B, X) = 1$). If the agent can replicate this ex post flexibility by an ex ante contingent rule, then he should be able to find a rule, $f \in \mathscr{F}$, such that $EU(f, X) = 1$. Thus, $\sup_{f \in \mathscr{F}} EU(f, X) < 1$ says that the agent is unable to replicate the flexibility of the grand option by identifying the optimal action at each contingency ex ante. The intuition is that the optimal action depends on the details of the contingencies in a way too complicated to be reduced to a rigid ex ante rule.

The more subtle problem is to formalize the first part of the statement, namely that the agent has "thought through the problem" yet still finds it hard.[14] Making this intuition formal requires a formal description of the agent's model of his environment. The problem is that a utility distribution U does not provide *direct* clues of what the agent's model may be. On the other hand, the assumption that the agent has "thought through the problem" should be somehow reflected in U, at least indirectly. We thus propose the following condition to capture this idea:

A.3. For every $A \in \mathscr{A}$ and $g \in \mathscr{G}$, $U(g, A)$ is degenerate.[15]

We defer the interpretation of this condition in terms of the agent having "thought through the problem" until Section 4 where we introduce probabilistic models. This will be our formal way to describe the "agent's model of his environment." In particular, Sections 4.3.2 and our main theorems links A.3 to a condition on the agent's probabilistic model stating that he has conducted exhaustive introspection about the implications of his model.

[14] Where "hard" here means that the agent still faces a discrepancy between $\sup_{f \in \mathscr{F}} EU(f, X)$ and the payoff from the grand option $EU(B, X) = 1$.

[15] That is, $U(g, A)$ puts unit mass on some point $r \in \mathbb{R}$).

4. Probabilistic models of complexity

How are we to think of an agent with choices satisfying A.1–3? In this section, we introduce a framework where these choices are derived from a *probabilistic model* which we will interpret as the agent's "theory" or "model" of how his environment works. Within this framework, it will be easy to identify those situations where the agent (behaves as if he) views his environment as deterministic but too complex to have a tractable deterministic representation.

4.1. Probabilistic models

Definition 1. A *probabilistic model* consists of a probability space (Ω, Σ, P), a σ-algebra $\mathscr{M} \subset \Sigma$ representing the agent's knowledge, and random utility vectors $\tilde{v}(x) : \Omega \to V,\ x \in X$, such that

B.1. *Complete state space*: The state space is $\Omega = V^X$, the set of all functions from contingencies to payoffs; Σ is generated by the events $\{\omega : \tilde{v}(x) = v\},\ v \in V$ and $x \in X$.

B.2. *Measurable correlation*: For every n, and every $(z_1, \dots, z_n) \in V^n$, the probability $P\{\tilde{v}(x_1) = z_1, \dots, \tilde{v}(x_n) = z_n \mid \mathscr{M}\}$ depends measurably on $(x_1, \dots, x_n) \in X^n$.[16]

Since the state space (Ω, Σ) is held fixed, we refer to $(P, \mathscr{M})$ as the agent's probabilistic model.

We shall interpret $\tilde{v}(x, \omega)$ as the payoff function at x given state ω (for a categorization problem, this is just the category of x in state ω). Condition B.2 requires that correlation in category membership is well-behaved. B.1 is the more substantial condition. It requires that Ω be the obvious *complete state space* corresponding to this decision problem. In particular, our model precludes behavior based on missing states or unforeseen contingencies (see Section 6.1). Given this state space, Σ is simply the minimal set of events rich enough to make the payoff functions $\tilde{v}(x)$ well-defined random vectors. In summary, B.1 ensures that no v is a priori ruled out as the payoff function at x; if the agent considers such v impossible at x, this should be a property of his probabilistic assessment P rather than part of the state space.

Notational conventions. We use $\tilde{v}(x) : \Omega \to V$ to denote the random vector of payoffs at contingency x (one entry in the vector for each action $b \in B$). We use $\tilde{v}(x, \omega) \in V$ to denote the value of this random vector at a state ω. Finally, $\tilde{v}(b, x, \omega) \in \mathbb{R}$ denotes the payoff of action b under the vector $\tilde{v}(x, \omega)$.

[16] Relative to the product σ-algebra $\overline{\mathscr{A}}^n$.

4.2. Expected utility

Consider an agent facing a specific contingency x. His expected payoff at x given $\mathscr{M}$ and the constraint $b \in g(x)$ is:

$$\max_{b \in g(x)} E[\tilde{v}(b, x)|\mathscr{M}](\omega).$$

Although $\mathscr{M}$ enters this expression very much like information does in an incomplete information model, it plays a very different role here. Assumption B.3 below will require that the agent has complete information at each and every contingency, eliminating the possibility that $\mathscr{M}$ reflects incomplete information, and supporting the interpretation of our model as a model of complexity.

Ex ante the agent faces a randomly drawn contingency from X. Since each $x \in X$ has probability zero of being drawn under λ, the agent is not concerned about any specific x, but about evaluating the average performance of rules and options against all contingencies. The following definition makes this formal:[17]

Definition 2. The utility of an option $g \in \bar{\mathscr{G}}$ on a set $A \in \overline{\mathscr{A}}$ relative to $\mathscr{M}$ is

$$\tilde{U}(g, A) = \frac{1}{\lambda(A)} \int_A \max_{b \in g(x)} E[\tilde{v}(b, x)|\mathscr{M}]\, d\lambda. \tag{1}$$

The integral in Eq. (1) is the Pettis integral of the random variables $\max_{b \in g(x)} E[\tilde{v}(b, x)|\mathscr{M}]$, $x \in A$, a well-known integral first used in economics by Uhlig [18]—Appendix A provides background and references.[18,19] From the definition of the integral, $\tilde{U}(g, A)$ is an equivalence class of random variables, all of which agree P-almost everywhere. As detailed in Appendix A, using L^2 to denote the linear space of (equivalence classes of) square integrable random variables on Ω, $\tilde{U}$ is a function[20]

$$\tilde{U} : \mathscr{G} \times \mathscr{A} \to L^2.$$

We will frequently use the fact that the *expectation* of the Pettis integral is the integral of the expectations:[21]

$$E\tilde{U}(g, A) = \frac{1}{\lambda(A)} \int_A E \max_{b \in g(x)} E[\tilde{v}(b, x)|\mathscr{M}]\, d\lambda(x). \tag{2}$$

The next two examples illustrate some of the main issues in this paper.

[17] Although $\mathscr{A}$ and $\mathscr{G}$ are our primary objects of interest, some of the definitions are formulated more generally to cover $\overline{\mathscr{A}}$ and $\bar{\mathscr{G}}$ as these are useful in the proofs and in Section 8.

[18] This integral is needed to deal with the fact that $E[\tilde{v}(b, x)|\mathscr{M}](\omega)$ may be non-measurable in X for a fixed ω.

[19] In the case of a rule f, (1) reduces to $\tilde{U}(f, A) = \frac{1}{\lambda(A)} \int_A E[\tilde{v}(f(x), x)|\mathscr{M}]\, d\lambda(x)$.

[20] We will loosely talk about $\tilde{U}(g, A)$ as a random variable and write $\tilde{U}(g, A)(\omega)$. Whenever we do this, we are referring to a representative of the equivalence class $\tilde{U}(g, A)$. As noted above, the choice of a different representative must yield the same answer with P-probability 1.

[21] See Lemma A.1. The integral in the RHS is the usual Lebesgue integral.

Example 1. Let $V = \{v_1, v_2\}$; $\mathscr{M} = \Sigma$; and P_i be the distribution such that for every x, $P_i\{\tilde{v}(x) = v_i\} = 1$. Define $P = 0.5P_1 + 0.5P_2$.

Here, either action b_1 is the optimal action at *all* contingencies, or action b_2 is. To see this formally, for $i = 1, 2$, let ω_i denote the state defined by $\tilde{v}(x, \omega_i) = v_i$ for every $x \in X$. Then $\tilde{U}(b_i, X)$ is any random variable such that $\tilde{U}(b_i, X)(\omega_i) = 1$ and $\tilde{U}(b_i, X)(\omega_j) = 0$, for $i, = 1, 2$ and $i \neq j$.[22] Example 1 shows that $\tilde{U}$ may itself be random. Later, we shall argue that the implied behavior in this example fails to use the knowledge embedded in $(P, \mathscr{M})$.

We will find it useful to define the *utility distribution* associated with $(P, \mathscr{M})$ as the function

$$U : \mathscr{G} \times \mathscr{A} \to \Delta^{\mathbb{R}}, \tag{3}$$

where $U(g, A)$ is the distribution of the random variable $\tilde{U}(g, A)$. Define the *expected utility* of g over A as $E\tilde{U}(g, A)$ (which is of course the same as $EU(g, A)$). In Example 1, the utility distribution $U(b_i, X)$ is one that puts 0.5 mass on 1 and 0.5 mass on 0 so $E\tilde{U}(b_i, A) = 0.5$.[23]

Example 2. Let $V = \{v_1, v_2\}$; $\mathscr{M} = \Sigma$; and P be the i.i.d. distribution with $P\{\tilde{v}(x) = v_i\} = \frac{1}{2}$ for every x.

In this example $\tilde{U}(b_i, X)$, $i = 1, 2$, is the degenerate random variable that takes the value 0.5 with probability 1.

If we interpret the probabilistic model as the agent's theory of how his world works, then Examples 1 and 2 present drastically different theories. In Example 1, one observation enables the agent to eliminate all uncertainty, while no such learning can occur in Example 2.[24] Despite this difference, since $E\tilde{U}(b_i, X) = \frac{1}{2}$ in both examples, taking the expectations of $\tilde{U}$ cannot separate them.

4.3. Knowledge and introspection

We introduce two restrictions that enable us to relate an agent's probabilistic model to his view of the complexity of his environment (as opposed to other factors such as incomplete information or bounded rationality):

[22] $\tilde{U}$ is defined arbitrarily otherwise. Note that there are infinitely many such $\tilde{U}$'s, but they must all be equivalent.

[23] Of course, U and $\tilde{U}$ depend on the probabilistic model $(P, \mathscr{M})$, and so strictly speaking we should write $U_{P,\mathscr{M}}$. However, whenever $(P, \mathscr{M})$ is clear from the context we suppress references to it to avoid unnecessarily cumbersome notation.

[24] Later we relate the presence of correlation in these examples to introspection and learning.

4.3.1. Instance-specific knowledge

Randomness in payoffs has two potential sources: (1) *objective uncertainty* due to lack of information about the determinants of the true category of x; and (2) *cognitive uncertainty* due to the complexity of the relationship between ex post payoffs and the features defining contingencies, rather than lack of information about these features. Although one would expect both types of uncertainty to exist in each problem, in this paper we focus exclusively on modeling cognitive uncertainty. The following definition captures this restriction:

B.3. *Instance-specific knowledge*: $\mathscr{M} = \Sigma$.

Recall that Σ is generated by events of the form $\{\omega : \tilde{v}(x) = v\}$ for some $x \in X$ and $v \in V$. Informally, any such event corresponds to a statement: "the category of contingency x is v." Stating that the agent knows all events of the form $\{\omega : \tilde{v}(x, \omega) = v\}$ means that once he faces x all uncertainty at x is resolved. Since Σ contains all such events, B.3 makes formal the intuition that the agent "knows-it-when-he-sees-it."

Note that B.3 eliminates as motive for behavior uncertainty due to lack of knowledge of payoff-relevant states. In particular, under B.3, any randomness in payoffs cannot be the result of incomplete or imperfect information. An example that violates B.3 may help illustrate this point:

Example 3. Let $V = \{v_1, v_2\}$; $\mathscr{M} = \{\emptyset, \Omega\}$; and P be any distribution such that $P\{\tilde{v}(x) = v_i\} = \frac{1}{2}$ for every x.

In this example, even when given complete freedom to choose ex post any action he likes in the form of the "grand option" $g(x) = B$ for all x, the best the agent can hope for is $E\tilde{U}(g, X) = \frac{1}{2}$. On the other hand, B.3 would have required the grand option to have expected value 1.[25] Example 3 reflects objective uncertainty about payoff-relevant states, rather than any complexity considerations—indeed, there may be nothing complex about the environment in this example. This is very different from our intuition that complexity has to do with the agent's inability to think in advance of every possible contingency.

Condition B.3 is extreme in that it eliminates any possibility of objective uncertainty. However, we find this abstraction useful in isolating complexity as motive for behavior.

4.3.2. Introspection

Suppose an agent "knows" whether an event S occurred (that is, $S \in \mathscr{M}$). For example, S may be the event "x is of category v," for some x and v. We would expect a rational agent to incorporate this knowledge in evaluating expected payoffs. Our next condition formulates the idea that the agent has incorporated all the implications of what he knows about the problem through exhaustive introspection.

[25] We have $\max_{b \in B} E[\tilde{v}(b, x)|\mathscr{M}] = 1$ for every x, so Eq. (2) implies $E\tilde{U}(g, X) = 1$.

For any $S \in \mathscr{M}$ with $P(S) > 0$, let P_S denote the updated probability measure given by Bayes' rule.[26] We use $E_{P_S}[\tilde{v}(b,x)|\mathscr{M}](\omega)$ to denote the conditional expectations given $\mathscr{M}$ under P_S. We may then define the *conditional utility*,

$$\tilde{U}(g, A \mid S) = \frac{1}{\lambda(A)} \int_A \max_{b \in g(x)} E_{P_S}[\tilde{v}(b,x)|\mathscr{M}]\, d\lambda(x).$$

Our next condition is:

B.4. *Exhaustive introspection:* For any $g \in \mathscr{G}$, $A \in \mathscr{A}$, and event $S \in \mathscr{M}$, with $P(S) > 0$,

$$E\tilde{U}(g, A) = E\tilde{U}(g, A \mid S).$$

The idea is that the agent's expected utility $E\tilde{U}(g, A)$ has already taken into account every possible introspection given $\mathscr{M}$.

As an illustration of the failure of B.4, consider Example 1: since any event of the form $S \equiv \{\omega : \tilde{v}(x) = v_1\}$ belongs to Σ (hence to $\mathscr{M}$), if the agent introspects about any contingency x and finds out that S is true, his updated model P_S would have $E\tilde{U}(b_1, X \mid S) = 1$, while without such introspection $E\tilde{U}(b_1, X) = \frac{1}{2}$, violating B.4. In Example 1 the agent's theory does not incorporate in P knowledge he already has as part of $\mathscr{M}$.[27]

By ruling out such apparent contradictions, B.4 essentially requires that the agent has thought through his probabilistic model of the environment, leaving no potentially useful inferences unexploited.

5. Representation and implications

5.1. Representation

Theorem 1. *Suppose that an agent has a probabilistic model* $(P, \mathscr{M})$ *satisfying* B.3 *and* B.4 *and let* $U_{P,\mathscr{M}}$ *denote its utility distribution. Then* $U_{P,\mathscr{M}}$ *satisfies* A.1–3.

All proofs are in Appendix A.

Next we turn to the converse: an agent who makes choices satisfying A.1–3 behaves as if he maximizes expected utility with respect to a probabilistic model of his environment. First we need the following definition:

---

[26] That is, P_S is the probability measure on (Ω, Σ) defined by $P_S(Z) = \frac{P(Z \cap S)}{P(S)}$ for every $Z \in \Sigma$.

[27] In Example 1 there is a logical contradiction between the probabilistic model and the agent's introspection. The following is a variant of that example in which introspection only gradually changes the agent's evaluations: Let $V = \{v_1, v_2\}$ and assume $\mathscr{M} = \Sigma$. For $i = 1, 2$, define P_i to be a distribution where the $\tilde{v}(x)$ are i.i.d. with mean $p_i \in (0, 1)$, with $p_1 \neq p_2$. Define P to be a 50/50 lottery over P_1 and P_2. That is, P is generated in a two-stage process where, first $i \in \{1, 2\}$ is drawn, then the $\tilde{v}(x)$'s are generated according to P_i.

Definition 3. A probabilistic model $(P, \mathscr{M})$ rationalizes a utility distribution U if for every $g \in \mathscr{G}$ and $A \in \mathscr{A}$,

$$U(g, A) = U_{P,\mathscr{M}}(g, A) \tag{4}$$

The next result provides the desired converse:

Theorem 2. *Every utility distribution U satisfying* A.1–3 *can be rationalized by a probabilistic model $(P, \mathscr{M})$ satisfying* B.3 *and* B.4.

An outline of the proof may be helpful. We first use a martingale argument to establish the following result:

Lemma 1. A.1* *and* A.2* *imply that there exists a function $l: X \to \Delta^V$, measurable with respect to $\overline{\mathscr{A}}$, such that for every $A \in \mathscr{A}$ with $\lambda(A) > 0$*

$$EU(b_k, A) = \frac{1}{\lambda(A)} \int_A l_k(x)\, d\lambda.$$

The function l is essentially unique: if l' is any other such function then l and l' must agree on almost every x.

The lemma derives (essentially unique) local weights $l(x)$ which represent the probability of each category at x. We use these local weights to define a probability distribution P on (Ω, Σ) using Kolmogorov's extension theorem. The construction ensures that B.3–4 hold.

5.2. Complexity: definition and characterization

We are interested in an agent who is unable to describe a rule specifying in advance the optimal action in every conceivable contingency. The intuition is that the optimal action depends on the details of the contingencies in a way too complicated to be reduced to a rigid ex ante rule. First we introduce our criterion for determining the complexity of a problem:

Definition 4. A problem is *complex over A* if

$$\sup_{f \in \mathscr{F}} E\tilde{U}(f, A) < 1,$$

and *simple over A* if $\sup_{f \in \mathscr{F}} E\tilde{U}(f, A) = 1$.

How is complexity reflected in the agent's theory of his environment? If U satisfies A.1–3, Theorem 2 says that the agent's choices are made as if he maximizes expected utility relative to a probabilistic model $(P, \mathscr{M})$ satisfying B.3 and B.4.[28] The next theorem characterizes complexity in terms of $(P, \mathscr{M})$:

[28] Theorems 6 and 7 later show that this representation is essentially unique.

Theorem 3. *Fix U satisfying* A.1–3 *and let* $(P, \mathscr{M})$ *be the probabilistic model constructed in Theorem* 2. *Then a problem is complex over* $A \in \mathscr{A}$ *if and only if there is* $A' \subset A$, $A' \in \overline{\mathscr{A}}$, *with* $\lambda(A') > 0$ *such that* $E\tilde{v}(x) \notin V$ *for every* $x \in A'$.

The theorem characterizes the complexity of a problem in terms of *local* properties of the probabilistic model, namely the non-degeneracy of the random variable $\tilde{v}(x)$ at contingencies $x \in A'$.

To interpret the condition $E\tilde{v}(x) \notin V$, we note that the martingale convergence argument in the proof of Lemma 1 derives $E\tilde{v}(x)$ as the frequency of category membership in vanishingly small neighborhoods around x.[29] On the other hand, condition A.1 on U (and condition B.3 on $(P, \mathscr{M})$) implies that all uncertainty vanishes when the agent knows the specific contingency he is facing. Theorem 3 thus says that complexity arises if (and only if) there is the following discontinuity: although randomness is present in arbitrarily small neighborhoods around any $x \in A'$, it completely vanishes once the agent encounters the specific contingency x. Informally, the agent believes that categories around x are "mixed-up" because small differences in details can have complex, unpredictable effects on outcomes. This is what leads to the discontinuity between the ex ante planning about a large set of contingencies and the ex post choice when a specific contingency is realized.

5.3. Uncertainty without incomplete or imperfect information

A surprising aspect of our construction is that the agent views as random an environment where there is no incomplete or imperfect information. How can we have a non-degenerate randomness (P is not trivial) yet at the same time $\mathscr{M} = \Sigma$? It is useful to highlight the counter-intuitive nature of this conclusion by noting that it *cannot* hold when the set of contingencies is finite:

Example 4. $V = \{v_1, v_2\}$; the space of contingencies X is finite; $\mathscr{A}$ is the set of all subsets; and λ is the normalized counting measure on $(X, \mathscr{A})$.

Our model can be naturally extended to cover the finite case as follows: $\Omega = V^X$; Σ is the set of all subsets of Ω; and $\mathscr{F}$ is the set of all functions $f : X \to B$. A probabilistic model $(P, \mathscr{M})$ then consists of a distribution P on the finite state space (Ω, Σ) and $\mathscr{M} \subset \Sigma$. B.3 implies that $\mathscr{M}$ consists of all subsets of Ω. If $(P, \mathscr{M})$ satisfies B.3–4, then it is easy to see that Theorem 1 still holds. The problem is that this theorem can only hold trivially: any such P must be degenerate, in the sense that $P\{\tilde{v}(x) = v_1\}$ is either 1 or 0 for all x.[30]

Thus, although Theorem 1 is still valid when X is finite, it is inconsistent with any uncertainty. By contrast when X has the structure detailed in Section 2,

[29] $E\tilde{v}(x)$ coincides with the local weights $l(x)$ derived in Lemma 1 almost everywhere.

[30] To see this, suppose that $0 < P\{\tilde{v}(x) = v_1\} = l_1 < 1$, then $\tilde{U}(b_1, x) = \frac{1}{\lambda(x)} \int_x E(\tilde{v}(b_1, x)|\mathscr{M})(\omega)\, d\lambda = E(\tilde{v}(b_1, x)|\mathscr{M})(\omega)$. Since $\mathscr{M} = \Sigma$ and $0 < l_1 < 1$, $E(\tilde{v}(b_1, x)|\mathscr{M})(\omega)$ is a non-degenerate random variable taking values in $\{0, 1\}$, contradicting A.3.

non-degenerate probabilistic models are easy to produce (for instance, Example 2).

So what drives our model of complexity? The answer is that we exploit the special structure on X to restrict the agent to the set of rules, $\mathscr{F}$. In our model, the agent cannot make his actions vary too finely with contingencies because he is limited to rules that must condition on features, rather than the set of all contingent actions, B^X. Of course, we allow for a set of rules rich enough to capture all regularities that can improve the agent's decisions. Given this richness, any details not picked up by rules must look random. Essentially, rules must be measurable with respect to the features, and thus necessarily lump together many contingencies and treat them the same.

5.4. Complexity and independence

Our notion of complexity is intimately related to the idea of statistical independence. We think of our agent as one who has done thorough introspection, uncovering any useful patterns and regularities in the problem. Any remaining randomness in his probabilistic model should contain no exploitable patterns, and should thus represent "independent randomness." This intuition is made formal in the following theorem:

Theorem 4. *Any probabilistic model satisfying* B.3 *and* B.4 *also satisfies*:

B.4*. *For every* $n \geqslant 2$, *there is a set* $A^n \in \mathscr{A}^n$ *with* $\lambda^n(A^n) = 1$ *such that for every* $\{x_1, \ldots, x_n\} \in A^n$, $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n)\}$ *are independent.*[31]

B.4* roughly says that knowing payoffs at one instance conveys no additional information about payoffs at almost any other instance.

6. Unforeseen contingencies, missing states, and statistical consistency

The condition $\sup_{f \in \mathscr{F}} E\tilde{U}(f, A) < 1$ means that the agent strictly prefers options to rules. Options give the agent the flexibility to choose after he knows which contingency he is facing, while rules force him to commit ex ante. Our theorems show that this *preference for flexibility* arises if and only if the problem is complex.

In this section we explore the link with the closely related literature on preference for flexibility and missing-states models introduced by Kreps [12] and studied by Dekel et al. [4], Nehring [14] and Ozdenoren [15]. In particular, we show that in our model, this preference for flexibility is a persistent, "steady-state" phenomenon that does not disappear with learning.

6.1. Comparison with Kreps: missing-states and unforeseen contingencies

Kreps's agent chooses an action in B as a function of the contingency s in a finite set S. The agent has (using our terminology) a preference over rules, defined as

[31] Here, λ^n is the product measure on the product space $(X^n, \bar{A}^n)$.

functions $f : S \to B$, as well as conditional opportunity sets, or options $g : S \to 2^B \backslash \emptyset$. A preference for flexibility arises if the agent strictly prefers an option to any of its refinements by rules. Kreps interprets such preference for flexibility as the agent believing that his model is missing some potentially relevant contingencies he did not foresee. Without going into the details, Kreps [12], Dekel et al. [4], Nehring [14], and Ozdenoren [15] construct expanded state spaces as completions of S, where the agent has a standard preference over the larger state space. The new subjective states are interpreted as the unforeseen contingencies missing in the original state space.

Kreps's agents may value flexibility for a variety of reasons, one of which is an inability to foresee future contingencies. But they may also value flexibility for completely unrelated reasons, such as an intrinsic taste for flexibility, or an exogenous restriction on the set of contingencies they can condition on. The presence of unforeseen contingencies is only one of many possible interpretations of such behavior. As Kreps indicates, it may be impossible to separate it from these other factors [12, p. 268]. In applications, one would expect the source of preference for flexibility to be important when introspection, learning, dynamic choice, and multi-agent considerations are introduced.

The basic idea underlying this paper is similar to that in Kreps [12], namely that the set of allowable rules is too coarse to reflect all the potential variability of the environment. The way we model this intuition, however, is very different: we focus on environments with considerably more structure than Kreps's set of contingencies S—an arbitrary finite set with no special structure. In fact, the state space in our model is the natural *complete* state space V^X, thus ruling out from the outset missing states as an explanation of behavior.[32]

The problem with an arbitrary finite set of contingencies is that there is no obvious or natural restriction on the set of allowable rules (and Kreps does not provide any). In our model, on the other hand, the agent is restricted to rules that belong to $\mathscr{F}$, which is considerably smaller than the set of all contingent actions B^X. Flexibility is valuable because the agent is unable to make his ex ante plans of action depend too finely on contingencies.[33]

In Kreps's model, the missing-states that make flexibility valuable are derived from preferences and interpreted as representing the agent's imperfect perception of the environment. However, no formal explanation is given as to which contingencies are more/less likely to be unforeseen, or why some states were missing and unforeseen in the first place. While this is sufficient as a "static" description of behavior, it can be problematic in incorporating forces that change the agent's perception of his environment, such as learning and introspection. Learning, in particular, can potentially resolve the agent's cognitive uncertainty and thus affects what he does and does not foresee.

[32] Example 4 illustrates that our model uses the special structure of X: when this set is an arbitrary finite set, the only probabilistic model consistent with our assumptions is trivial and displays no preference for flexibility.

[33] "It isn't that the individual does not foresee the contingencies, but only that we don't allow him to have his consumption so finely conditioned as he would like." (Kreps [12, p. 278])

Our model of *complete* state space and *foreseen* contingencies already explicitly incorporates introspection (Section 4.3.2). We now turn to formally show how our model incorporates learning.

6.2. Statistical consistency and learning

This section formalizes the interpretation of the probabilistic model as the agent's "steady-state" theory of his environment *after all useful learning has taken place*.

Consider a statistical experiment in which the agent's choices are observed along an infinite sequence of contingencies $\{x_1, \dots\}$ obtained as the result of independent draws from a subset $A \in \mathscr{A}$ according to the distribution λ. Along this sequence, we imagine the agent finding out what his payoff functions turn out to be at each contingency. So at contingency x_i he finds that his payoff is $\hat{v}_i \in V$, and so on. For a specific draw of a sequence of contingencies, we obtain an infinite sequence of data points of the form $\{x_i, \hat{v}(x_i)\}_{i=1}^{\infty}$.

We interpret $E\tilde{U}(b, A)$ as the agent's prediction, based on his probabilistic model, of the average payoff from taking action b on A. The agent may test this prediction by checking it against the data, $\{x_i, \hat{v}(x_i)\}_{i=1}^{\infty}$. To make this formal, call $(P, \mathscr{M})$ *compatible with* $\{x_i, \hat{v}(x_i)\}_{i=1}^{\infty}$ if for every $b \in B$ and $A \in \mathscr{A}$,[34]

$$E\tilde{U}(b, A) = \lim_{N \to \infty} \frac{1}{\#\{x_i \in A\}} \sum_{i=1}^{N} \chi_A(x_i)\hat{v}(b, x_i).$$

That is, the agent's prediction $E\tilde{U}(b, A)$ is consistent with the *empirical frequency* of the event that b is indeed the optimal action. Although there may be more sophisticated tests of consistency, this simple test will suffice in illustrating our main point.

The following theorem examines compatibility with data for a "typical" sequence of contingencies $\{x_1, \dots\}$:

Theorem 5. *Any probabilistic model satisfying* B.3–4 *is compatible with data almost surely.*

Formally, let λ^{∞} denote the product distribution on the set of sequences from X. Then for every $b \in B$ and $A \in \mathscr{A}$, for λ^{∞}-a.e. sequence of contingencies,[35]

$$E\tilde{U}(b, A) = \lim_{N \to \infty} \frac{1}{\#\{x_i \in A\}} \sum_{i=1}^{N} \chi_A(x_i)\tilde{v}(b, x_i, \omega), \quad P\text{-a.s.}$$

[34] **Remarks**. (1) χ_A denotes the indicator function of the set A; (2) if $\#\{x_i \in A\} = 0$, then the RHS in the following equation is infinite and equality is violated (since the LHS is finite); and (3) the equality means that the limit exists and that it is equal to $E\tilde{U}(b, A)$.

[35] To define this expectations, fix $\{x_1, \dots\}$, and let $\Omega_{\{x_i\}} = V^{\mathbb{N}}$ denote the set of all infinite sequences in V, $\Sigma_{\{x_i\}} \subset \Sigma$ is the sub-σ-algebra generated by the random variables $\{\tilde{v}(x_1), \dots\}$.

That is, the agent does not expect even an *infinite* amount of data to falsify his prediction of the expected payoff $E\tilde{U}(b, A)$ based on his probabilistic model. This is one way to model the intuition that the agent views his probabilistic model as a steady state model that cannot be further refined by data.

It is useful to consider the case of a probabilistic model that is *not* statistically consistent. In Example 1 for $b \in \{b_1, b_2\}$, $E\tilde{U}(b, X)$ is a degenerate random variable that takes the value 0.5 with probability 1. On the other hand, for any sequence $\{x_i\}$ of contingencies the agent perfectly learns whether P_1 or P_2 is true, so $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \tilde{v}(b, x_i, \omega)$ is a random variable that takes value 0 or 1 with equal probability.

In this example, although the agent faces an initial uncertainty, he expects to learn enough from the data to eliminate this uncertainty. The fact that the agent values flexibility in this example is not robust to learning, as it quickly disappears upon learning whether P_1 or P_2 is the true distribution.

In summary, the value of flexibility in Example 1 is due to neither complexity nor unforeseen contingencies, but to the fact that the agent has not exhausted all learning opportunities. What Theorem 5 tells us is that when the agent values flexibility in our model, he does so as a "steady-state" behavior that already took into account all learning possibilities.

7. Uniqueness of the representation

Theorem 2 only shows the possibility of finding one probabilistic model satisfying B.3–4 and that rationalizes the agent's choices. It does not rule out that these same choices may be consistent with other probabilistic models that do not satisfy B.3–4.

Our next result shows that one *cannot* substantially weaken conditions B.3 and B.4.

Theorem 6. *Any probabilistic model rationalizing a utility distribution U satisfying* A.1–3 *must satisfy*

B.3*. *For almost every* $x \in X$, *there is a partition of* Ω, $\mathcal{M}_x = \{M_v : v \in V\} \subset \mathcal{M}$, *such that for every* $v \in V$, $P\{M_v \,\Delta\, \{\omega : \tilde{v}(x) = v\}\} = 0$.[36]

B.4*. *For every* $n \geqslant 2$, *there is a set* $A^n \in \mathcal{A}^n$ *with* $\lambda^n(A^n) = 1$ *such that for every* $\{x_1, \ldots, x_n\} \in A^n$, $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n)\}$ *are independent.*

Conditions B.3* and B.4* in the previous theorem slightly weaken conditions B.3 and B.4 used in Theorem 1. Essentially we need the weaker assumptions, since we cannot observe differences confined to measure zero events from the behavior of the

[36] Δ denotes symmetric difference of two sets.

agent. Therefore from A.1–3, we can only hope to get versions of B.3 and B.4 that hold up to sets of measure zero.[37]

Each probabilistic model provides an interpretation of how the agent views his environment. Our next task is to provide an unambiguous interpretation of the agent's behavior in terms of his perception of the complexity of the environment. The following theorem establishes a uniqueness result of the probabilistic representation to support this interpretation:

Theorem 7 (Uniqueness). *Let* $(P, \mathscr{M})$ *and* $(P', \mathscr{M}')$ *be two probabilistic models rationalizing utility distribution* U *satisfying* A.1–3. *Then*:

(i) Equivalent knowledge: *For almost every* $x \in X$, *there are partitions of* Ω, $\mathscr{M}_x = \{M_v : v \in V\} \subset \mathscr{M}$ *and* $\mathscr{M}'_x = \{M'_v : v \in V\} \subset \mathscr{M}'$, *such that for every* $v \in V$, $P\{M_v \,\Delta\, M'_v\} = 0$.

(ii) Equivalent distributions: *For every n, there is* $A^n \subset X^n$ *with* $\lambda^n(A^n) = 1$ *such that for every* $(x_1, \ldots, x_N) \in A^n$, *the random payoffs* $(\tilde{v}(x_1), \ldots, \tilde{v}(x_N))$ *have identical distributions under* P *and* P'.

8. Application: optimal delegation contracts

We provide an example illustrating how our model may be used in a contractual setting. We describe a simple delegation problem between a principal and an agent and characterize the optimal contract. We show, in particular, that the optimal contract may be incomplete depending on the complexity of the environment.

8.1. Contracts as options

A principal decides on how much discretionary control over an action he should delegate to an agent. One example is that of an owner of a firm (the principal) delegating to an agent the task of managing the day-to-day operation of a firm. Another example is a legislature that introduces laws but delegates the adjudication of legal cases to judges. These and other settings are typically characterized by "contractual incompleteness," in the sense that the contracts (or laws in the legislature example) do not specify a precise action in every conceivable contingency. Instead, they tend to contain vaguer clauses that may determine boundaries of

[37] It is immediate that the difference between B.3* and B.3 is only on sets of measure zero. Thus B.3* requires almost instance specific knowledge for almost all instances.
To see the difference between B.4* and B.4 consider the following "full independence" assumption:

B.4**. For every $n \geq 2$, and every $\{x_1, \ldots, x_n\} \in X^n$, $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n)\}$ are independent.

It is easy to show that B.4** implies B.4, and in Appendix A we show that B.4 implies B.4*. So loosely speaking the difference between B.4 and B.4* is less than the difference between B.4** and B.4*. Note that the difference between the latter two assumptions are only on measure zero sets, thus the difference between B.4 and B.4* must also be of this nature.

permissible actions, leaving the rest to the agent's discretion. We show that the complexity of the environment is a major source of such incompleteness. We also perform comparative static analysis that explains how complexity shapes the *form* of incompleteness.

We imagine an agent choosing an action $b \in B$ at each contingency x. A contract is a constraint on the agent's choice of action. One way to model this is to view contracts as options: at each contingency x, the agent chooses an action b subject to the constraint that $b \in g(x)$. Here we expand the set of contracts to include the closure $\bar{\mathscr{G}}$ of the set of options $\mathscr{G}$. We do this primarily for technical reasons: the space $\mathscr{G}$ is not closed, so an optimal contract may not necessarily exist in $\mathscr{G}$ but will always exist in $\bar{\mathscr{G}}$. We do not view the difference between $\mathscr{G}$ and $\bar{\mathscr{G}}$ to be of substantive importance for the application considered here: If $\bar{g} \in \bar{\mathscr{G}}$ is optimal, then for any $\varepsilon > 0$ we can always find a finitely defined option $g \in \mathscr{G}$ such that the expected payoff from using g is ε-close to that of $\bar{g}$, and g and $\bar{g}$ coincide except on a set of measure at most ε.[38] The following definition provides a useful taxonomy of the different types of contracts:

Definition 5. A *contract* is an option $g \in \bar{\mathscr{G}}$.

- A contract g is *complete* if it belongs to $\bar{\mathscr{F}}$ (i.e., a rule, or a single-valued option);
- A contract g is *incomplete at* x if $g(x)$ is not a singleton (otherwise it is *complete at* x);
- g is essentially *incomplete* if it is incomplete on a set of positive measure of x's (otherwise it is *essentially complete*).

To see how a contract g can capture the degree of discretion given to the agent, note first that when g is single-valued, the agent is left with no scope for discretion. Such contract represents a rigid rule that specifies ex ante the agent's actions in every conceivable contingency. At the other extreme, the contract "$g(x) = B$ for all x" gives the agent full discretion. The model allows a rich set of intermediate cases in which the agent is given partial discretion ($g(x)$ is neither a singleton nor equal to all of B), and allows the degree of discretion to vary across contingencies.

8.2. Optimal contracts

We use our representation of complexity to identify the optimal level of discretion in terms of a trade-off between giving the agent greater flexibility and controlling the potential conflicts of incentives. We characterize the optimal delegation contracts which may be incomplete depending on the complexity of the environment.

[38] Anderlini and Felli [1] make a similar point.

To model this, we assume that the principal and the agent share a common probabilistic model $(P, \mathscr{M})$ which is assumed to satisfy B.1–4.

The only difference with the setup of Section 3 is that we let $\tilde{v}(x) : \Omega \to V \times V$, and write $\tilde{v}(x) = (\tilde{v}_1(x), \tilde{v}_2(x))$, with $\tilde{v}_1(x)$ and $\tilde{v}_2(x)$ denoting the principal's and agent's payoffs, respectively.[39] We also drop assumption A.0 and allow V to be any finite set of payoff functions $v : B \to \mathbb{R}$ where each function induces a strict ranking of actions.

Define the local weights $l : X \to \Delta^{V \times V}$ by $l(x)(v_1, v_2) = P\{v(x) = (v_1, v_2)\}$. While we maintain the independence assumption (B.4) across contingencies, we allow payoffs to be correlated at each contingency (so v_1 and v_2 may be correlated). Such correlation would capture the (possibly stochastic) relationship between the principal's and the agent's interests. For example, if l puts unit mass on the diagonal of $V \times V$ (i.e., $v_1 = v_2$ with probability 1), there is always complete alignment between the principal's and agent's ranking of actions. On the other hand, off-diagonal utilities represent potential conflict of interest that may lead the principal to transfer less than full discretion to the agent (in the extreme case, leaving him with no discretion at all).

For a given agent utility $v_2 \in V$ and subset of actions $C \in 2^B \backslash \emptyset$, define $b(C, v_2) = \operatorname{argmax}_{b \in C} v_2(b)$, which is single valued by our assumption that payoffs strictly rank actions. In a given contingency x, and state ω, the agent's optimal action is given by $b(C, \tilde{v}_2(x, \omega))$, generating a principal's (random) payoff $\tilde{v}_1(b(C, \tilde{v}_2(x, \omega)), x, \omega)$. His expected utility from a contract g is thus:

$$\begin{aligned} U(g) &= E \int_X \tilde{v}_1(b(g(x), \tilde{v}_2(x, \omega)), x, \omega)\, d\lambda \\ &= \int_X \sum_{v_1, v_2} l(x)(v_1, v_2) v_1(b(g(x), v_2))\, d\lambda. \end{aligned}$$

The second equality, which may be proven analogously to Lemma A.1, asserts that one can evaluate g by computing its value locally then aggregate the resulting payoffs.

8.3. Comparative statics

We want to characterize the optimal contract in terms of the primitives of the environment, in particular, complexity and the conflict of interest between the principal and the agent. Call $g^+ \in \bar{\mathscr{G}}$ *globally optimal* if $U(g^+) \geqslant U(g)$ for every $g \in \bar{\mathscr{G}}$. That is, g^+ is the optimal level of discretion to be transferred to the agent.

A simpler optimality criterion requires g to be *locally optimal*: for almost every x

$$g(x) \in \underset{C \in 2^B \backslash \emptyset}{\operatorname{argmax}}\ E_{l(x)} v_1(b(C, v_2)). \tag{5}$$

This expression may be interpreted as follows: at each contingency x, the principal sets C optimally, assuming that payoffs are drawn according to

[39] Assuming a common set V is without loss of generality because we can always take this set to be the union of the support of the two players' utilities.

the weights $l(x)$ and the agent picks his action after the payoff functions are drawn.

Theorem 8. (i) *A locally optimal contract exists.*
(ii) *A contract is globally optimal if and only if it is locally optimal.*

Part (i) guarantees that local optima can be threaded together into a well-defined option, while (ii) says that global optimality is equivalent to optimality at local problems that involve no complexity considerations.

To illustrate the theorem, consider the following two polar cases. First, suppose that the $l(x)$'s are non-degenerate with support contained in the diagonal of $V \times V$. It is easy to see that the solution of local problems, as determined by Eq. (5), is to set $g(x) = B$, i.e., to give the agent full discretion. By the theorem, this is also the optimal contract. In this case, although complexity makes it impossible to anticipate every possible future contingency, the principal knows that the interests of the agent will always be aligned with his. Second, suppose that the environment is simple and there is conflict of interest: for every x, $l(x)$ is degenerate and puts unit mass on some point off the diagonal of $V \times V$. In this case, from Eq. (5) we also find that it is optimal to specify the action to be taken in every contingency ($g(x)$ is always single-valued); the optimal contract is complete, leaving the agent with no room for discretion.

The force of the theorem is that it provides a general characterization of optimal contracts that covers not just the two polar cases above, but also more interesting intermediate cases. Essentially, the theorem reduces the problem of searching for an optimal contract in a complex environment to that of solving completely standard, "local" optimization problems. This makes it possible to derive qualitative insights about the way complexity shapes the *form* of incompleteness that arises in applications. In these intermediate cases where l is non-degenerate with support containing off-diagonal elements of $V \times V$, complexity prevents the principal from fully spelling out which action should be taken by the agent in every possible contingency. This creates an incentive to transfer greater discretion to the agent so he can more flexibly adapt to contingencies as they arise. On the other hand, the fact that l puts positive weight on off-diagonal elements of $V \times V$ means that the principal also realizes that there will be contingencies in which the agent's self-interest will conflict with his. The principal then limits the agent's discretion by offering a more narrowly defined contract. The overall arrangement in this case will be one characterized by partial delegation to balance the conflicting goals of greater flexibility to cope with complexity with that of better control over the agent's actions. Theorem 8 ensures that global optimality can be analyzed by examining straightforward local trade-offs.

9. Concluding remarks

While its importance and pervasiveness is beyond dispute, complexity is an elusive concept to formally model. We believe the difficulty arises from modeling complexity in a way that precludes systematic errors and arbitrage opportunities that undermine equilibrium analysis. Our approach is to elicit agents' assessment of the complexity or simplicity of the environment, the world as they see it, rather than prescribe ad hoc computational limitations or thought procedures. We explain an agent's behavior as the result of a coherent probabilistic model representing his attempt at identifying patterns or regularities, while at the same time recognizing that his model cannot explain everything.

The resulting model, as illustrated in the example of Section 8, is consistent with closed equilibrium analysis: agents optimize given an understanding of their environment which is as good as that of the modeler, and there are no arbitrage opportunities or money-pumps through which an agent may be systematically exploited. These requirements, well entrenched in economics and game theory, impose considerable discipline on the model's predictions.

Two other features of our approach are worth emphasizing. First, we show that *rational* agents confronting complex situations may display behavior often associated with bounded rationality. Second, despite the complexity of the environment, the agent's model of it may be remarkably simple. This, in turn, means that we can develop tractable models of behavior, as illustrated by Theorem 8 where complex multi-agent settings may be analyzed using essentially standard tools and concepts.

An important question we hope to address in future work is incorporating learning considerations: our focus has been on a steady state where the agent believes he learned all there is to be learned from the environment. How the agent gets to use data to formulate and refine his model is an obvious next step.

Acknowledgments

We are grateful to an Associate Editor and a referee for valuable comments. We also thank Ehud Kalai, Peter Klibanoff and Tim Van Zandt for helpful discussions. Casadesus-Masanell gratefully acknowledges the support on the Harvard Business School Division of Research.

Appendix A

A.1. Topological and measurable properties of $\overline{\mathscr{A}}$

The next proposition shows that $\mathscr{A}$ and $\overline{\mathscr{A}}$ have standard topological and measurable structures:

Proposition A.1. (i) *$\mathscr{A}$ is a base for the product topology, τ, on X;*

(ii) *τ is a complete, separable metrizable topology;*

(iii) *$\overline{\mathscr{A}}$ coincides with the Borel σ-algebra generated by τ.*

Proof. (i) Let τ' denote the topology generated by taking $\mathscr{A}$ as a base (that is the collection of sets obtained by taking arbitrary unions of sets in $\mathscr{A}$). We need to show that $\tau = \tau'$. Call a set *simple* if it is of the form $\{x : x^i = \alpha\}$, for some feature i, and $\alpha \in X^i$. Note that simple sets are inverse images of projections, and the product topology is the coarsest topology that makes all projections continuous. Thus, τ-open sets are those sets which can be obtained as arbitrary unions of finite intersections of simple sets. Obviously, every simple set is in $\mathscr{A}$, and so is any finite union of simple sets. Consequently, $\tau \subset \tau'$. In the other direction, it is enough to show that $\mathscr{A} \subset \tau$. By definition, any $A \in \mathscr{A}$ is definable in terms of a finite set of features I, say. Then A can be generated by taking finite intersections and unions of simple sets corresponding to features in I. Since every simple set is in τ, A must also be in τ.

(ii) τ is the product topology of a countable collection of compact spaces.[40] Thus, τ itself is compact [16, Theorem 19, p. 166, metrizable, p. 151], and complete and separable [16, Propositions 13–15, pp. 163–164].

(iii) Since $\mathscr{A}$ is countable, the product topology involves only countable unions of sets in $\mathscr{A}$. Thus, the σ-algebra generated by $\mathscr{A}$, namely $\overline{\mathscr{A}}$, coincides with that generated by the product topology. But the latter is just the Borel σ-algebra. □

A.2. The Pettis integral

Here we provide the definition of the Pettis integral as it applies to our setup. The interested reader may consult [6] for more detailed account. We also provide an elementary lemma for manipulating this integral.

Let L^2 to be the linear (Hilbert) space of all random variables on (Ω, Σ, P) with finite mean and variance.[41] The inner product of two points $f, f' \in L^2$ is $(f \mid f') = \int_\Omega f(\omega) f'(\omega)\, dP = cov(f, f') + EfEf'$. If χ_S is the characteristic function of a positive probability event $S \in \Sigma$, then $(\chi_S \mid f) = \int_S f(\omega)\, dP = E(f \mid S)P(S)$.[42]

Consider a function $x \mapsto \tilde{f}(x)$, which maps each contingency $x \in X$ into a random variable $\tilde{f}(x)$. The Pettis integral is a way to integrate such mapping by averaging the random variables $\tilde{f}(x)$ as points in L^2. Formally, $\int_A \tilde{f}(x)\, d\lambda \in L^2$ is the Pettis integral

[40] In fact discrete spaces, as each X_i is assumed to be binary.

[41] More precisely, L^2 consists of equivalence classes of functions because the L^2 norm cannot distinguish between two random variables that differ only on a set of measure zero. Here, we will abuse notation and use the same symbols to denote the random variable and its equivalence class, as the difference plays no role in what follows.

[42] The notation $(x \mid y)$ denotes the inner product of two random variables, while $E(x \mid y)$ denotes the conditional expectation of x given y.

of $x \mapsto \tilde{f}(x)$ over $A \in \overline{\mathscr{A}}$ if for every $z \in L^2$,

$$\left(z \,\middle|\, \int_A \tilde{f}(x)\, d\lambda \right) = \int_A (z \,|\, \tilde{f}(x))\, d\lambda,$$

where the integral on the RHS is the ordinary Lebesgue integral. General results on the existence of the integral $\int_A f(x)\, d\lambda$ are available [6].

The following lemma says that the expectation of the Pettis integral of a collection of random variables is the Lebesgue integral of their expectations:

Lemma A.1. *Suppose that* $S \in \Sigma$ *with* $P(S) > 0$. *Then for every* $A \in \overline{\mathscr{A}}$

$$E\left[\int_A \tilde{f}(x)\, d\lambda \,\middle|\, S \right] = \int_A E(\tilde{f} \,|\, S)\, d\lambda.$$

Proof.

$$\begin{aligned} E\left[\int_A \tilde{f}(x)\, d\lambda \,\middle|\, S \right] &= \frac{1}{P(S)} \left(\chi_S \,\middle|\, \int_A \tilde{f}(x)\, d\lambda \right) \\ &= \frac{1}{P(S)} \int_A (\chi_S \,|\, \tilde{f}(x))\, d\lambda \\ &= \frac{1}{P(S)} \int_A E(\tilde{f} \,|\, S) P(S)\, d\lambda \\ &= \int_A E(\tilde{f} \,|\, S)\, d\lambda. \qquad \square \end{aligned}$$

Note that the Pettis integral itself is a random variable. In Example 1, the integral $\int_X \tilde{v}(b_1, x)\, d\lambda$, say, is a random variable taking the value 0 and 1 with equal probability. This reflects the correlation built into the joint distribution P in that example: with probability 0.5, b_1 is the correct action, yielding a payoff of 1, but with probability 0.5 it is the wrong action so its payoff is 0. In the example in footnote 27, $\int_X \tilde{v}(b_1, x)\, d\lambda$ is p_i and p_j with probability 0.5 each. The interpretation is similar.

On the other hand, under B.3 and B.4 the Pettis integral is in fact constant with probability 1. This is shown in the next subsection.

A.3. Proof of Theorems 1 and 4

We break the proof in three pieces to highlight the role of the different assumptions.

Proposition A.2. *Suppose that* $(P, \mathscr{M})$ *is any probabilistic model* (*i.e.*, *it satisfies* B.1–2), *then* $\tilde{U}$ *must satisfy* A.1*. *If* $(P, \mathscr{M})$ *also satisfies* B.3, *then* $\tilde{U}$ *satisfies* A.1.

Proof. Suppose that B.3 holds and fix $A \in \mathscr{A}$ with $\lambda(A) > 0$. We must show that for every $C \in 2^{B \setminus \emptyset}$,

$$E\tilde{U}(C, A) = l_1 \max_{b \in C} v_1(b) + \cdots + l_K \max_{b \in C} v_K(b).$$

By definition, $E\tilde{U}(C, A) = \int_A E \max_{b \in C} E[\tilde{v}(b, x)|\mathscr{M}]\, d\lambda$. Since $\mathscr{M} = \Sigma$ (condition B.3), we have $E[\tilde{v}(b, x)|\mathscr{M}] = \tilde{v}(b, x)$. Then

$$E \max_{b \in C} \tilde{v}(b, x) = \sum_{k=1}^{K} P\{\tilde{v}(x) = v_k\} \max_{b \in C} v_k(b).$$

Integrating both sides on A and letting $l_k = \int_A P\{\tilde{v}(x) = v_k\}\, d\lambda$ yields the desired conclusion.

To show that A.1* holds even without B.3, note that in A.1* only actions are relevant, so we are only considering $C = \{b_k\}$ for some action b_k. Then the "max" may be deleted in all expressions in the last paragraph, so $E \max_{b \in C} E[\tilde{v}(b, x)|\mathscr{M}]$ reduces to $E\, E[\tilde{v}(b_k, x)|\mathscr{M}] = E\tilde{v}(b_k, x) = P\{\tilde{v}(x) = v_k\}$. Then $E\tilde{U}(b_k, A) = \int_A P\{\tilde{v}(x) = v_k\}\, d\lambda = l_k$. □

Proposition A.3. *If* $(P, \mathscr{M})$ *is any probabilistic model that satisfies* B.3, *then its* $\tilde{U}$ *must satisfy* A.2.

Proof. Fix A, g and a partition $\{A_1, \ldots, A_N\}$ of A as in A.2. Using Eq. (2) and B.3, we may write

$$\begin{aligned} E\tilde{U}(g, A) &= \frac{1}{\lambda(A)} \int_A E \max_{b \in g(x)} \tilde{v}(b, x)\, d\lambda(x) \\ &= \frac{1}{\lambda(A)} \sum_{i=1}^{N} \int_{A_i} E \max_{b \in g(x)} \tilde{v}(b, x)\, d\lambda(x) \\ &= \frac{1}{\lambda(A)} \sum_{i=1}^{N} \lambda(A_i) EU(g, A_i) \\ &= \sum_{i=1}^{N} \lambda(A_i|A)\, EU(g, A_i). \qquad \square \end{aligned}$$

Proposition A.4. *If* $(P, \mathscr{M})$ *is any probabilistic model that satisfies* B.3 *and* B.4, *then its* $\tilde{U}$ *must satisfy* A.3.

It is convenient to first prove Theorem 4. First we introduce some notation: For $n = 1, 2 \ldots$, define the functions:[43]

$$z(x^n, v^n) = P\{\tilde{v}(x_i) = v_i,\ i = 1, \ldots, n\} - \prod_{i=1}^{n} P\{\tilde{v}(x_i) = v_i\},$$

$$z(x^n) = \max_{v^n \in V^n} |z(x^n, v^n)|.$$

[43] In what follows, the notation x^n will refer to the vector $(x_1, x_2 \cdots x_n)$.

These represent the discrepancy from independence (the difference between the joint distribution and the product of the marginals). Note that from B.2 it follows that z is measurable. We also suppress notation indicating that z depends on n as this will always be clear from the context.

Lemma A.2. $(P, \mathscr{M})$ *satisfies* B.4* *if and only if it satisfies*

B.5. *For every* $n \geq 1$, *and almost every* $\{x_1, \ldots, x_n\} \in X^n$ *there is* $A(x_1, \ldots, x_n) \in \overline{\mathscr{A}}$ *with* $\lambda(A(x_1, \ldots, x_n)) = 1$ *such that for every* $x' \in A(x_1, \ldots, x_n)$, $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n), \tilde{v}(x')\}$ *are independent.*

Proof. First suppose that B.4* holds and fix $n \geq 1$. Then the function $z : X^{n+1} \to \mathbb{R}$ given by $z(x, x^n) = \max_{v^{n+1} \in V^{n+1}} |z(x, x^n, v^{n+1})|$ must be zero a.e. on X^{n+1} (here, (x, x^n) is a vector in X^{n+1}). Since z is jointly measurable on X^{n+1}, each of its sections $z_{x^n} : X \to \mathbb{R}$ given by $z_{x^n}(x) = z(x, x^n)$ is measurable. Applying Fubini's theorem we also have that for almost every $x^n \in X^n$, $z_{x^n}(x) = 0$ for a.e. $x \in X$, which is the property required by B.5 given n. Since n is arbitrary, B.5 holds.

In the other direction, suppose that B.5 holds. We prove the claim by induction on n. Clearly, B.5 implies that B.4* holds for $n = 2$ (this may be seen by noting that, using Fubini's Theorem and condition B.5, $\int_{X \times X} z(x, x')\, d\lambda(x)\, d\lambda(x') = 0$, hence $z(x, x') = 0$ for almost every pair x, x').

Suppose that for some n, B.4* holds for all $n' \leq n - 1$; we show that it must also hold for n. Let $\bar{B}^{n-1}$ be a set of λ^{n-1}-measure 1 supplied by the inductive step. By way of contradiction, suppose that there is a subset $A^n \subset X^n$, with $\lambda^n(A^n) > 0$ on which $z(x^n) > 0$ for every $x^n \in A^n$.

Taking the intersection with the λ^n-measure 1 set $\bar{B}^{n-1} \times X$ if necessary, we may assume that A^n is a set of positive measure such that for every $(x_1, \ldots, x_{n-1}, x_n) \in A^n$, by B.5 $z(x_1, \ldots, x_{n-1}, x_n) = 0$. From the definition of A^n, we must have $z(x^n) > 0$, but Fubini's theorem implies

$$\int_{A^n} z(x^n)\, d\lambda^n(x^n) = \int_A \int_{A^{-n}} z(x_1, \ldots, x_{n-1}, x_n)\, d\lambda^{n-1}(x_{-n})\, d\lambda(x_n) = 0.$$

Contradiction. □

Proof of Theorem 4. The proof is by induction on n, beginning with $n = 1$. Suppose that $(P, \mathscr{M})$ satisfies B.3 and B.4, but B.4* fails. Then there must be $\bar{A} \in \overline{\mathscr{A}}$ of positive measure such that for any $x \in \bar{A}$ there is A_x with $\lambda(\mathscr{A}_x) > 0$ such that for every $x' \in A_x$, $\{\tilde{v}(x), \tilde{v}(x')\}$ are not independent. For the remainder of the argument, fix one such $\bar{x} \in \bar{A}$.

The above implies that $z(x, \bar{x}) > 0$ for every $x \in A_{\bar{x}}$. The definition of z and the fact that V is finite imply that there must be $\bar{v},\ v' \in V$, and a set $A \subset A_{\bar{x}}$ with $\lambda(A) > 0$, such that $z(x, \bar{x}; v', \bar{v})$ is either always strictly positive, or strictly negative for all $x \in A$. In either case, this implies $P\{\tilde{v}(\bar{x}) = \bar{v}\} > 0$ (from the definition of z). Note that this implies $l(\bar{x})$ is non-degenerate (otherwise $\tilde{v}(\bar{x})$ would be trivially independent from every $\tilde{v}(x)$).

Assume, without loss of generality, that $z(x, \bar{x}; v', \bar{v})$ is always strictly positive on A. Substituting,

$$0 < \frac{z(x, \bar{x}; v', \bar{v})}{P\{\tilde{v}(\bar{x}) = \bar{v}\}} = P\{\tilde{v}(x) = v' \mid \tilde{v}(\bar{x}) = \bar{v}\} - P\{\tilde{v}(x) = v'\}. \tag{$*$}$$

Let S denote the event $\{\tilde{v}(\bar{x}) = \bar{v}\}$. Then, on the one hand, we have

$$EU(b', A) = \int_A E\tilde{v}(b', x)\, d\lambda = \int_A P\{\tilde{v}(x) = v'\}\, d\lambda$$

while

$$EU(b', A|S) = \int_A E[\tilde{v}(b', x)|S]\, d\lambda = \int_A P\{\tilde{v}(x) = v'|S\}\, d\lambda.$$

But $(*)$ implies that

$$\int_A [P\{\tilde{v}(x) = v' \mid \tilde{v}(\bar{x}) = \bar{v}\} - P\{\tilde{v}(x) = v'\}\, d\lambda > 0$$

so $EU(b', A|S) > EU(b', A)$, a contradiction with B.4.

Turning to the inductive step, suppose that the claim holds for all $n' < n$. We note that Lemma A.2 below implies that: there is a set $A^n \in \overline{\mathscr{A}}^n$ with $\lambda^n(A^n) = 1$ such that for every $\{x_1, \ldots, x_n\} \in A^n$, $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n)\}$ are independent. We now mimic our earlier proof for $n = 1$.

Suppose that $(P, \mathscr{M})$ satisfies B.3 and B.4, but B.4* fails for n. Then there must λ^n-positive measure set in $\mathscr{A}^n$ such that for any element of that subset, $(\bar{x}_1, \ldots, \bar{x}_n)$, there is a set $A_{\bar{x}_1,\ldots,\bar{x}_n} \subset X$ of positive λ-measure such that for every $x \in A_{\bar{x}_1,\ldots,\bar{x}_n}$, $\{\tilde{v}(\bar{x}_1), \ldots, \tilde{v}(\bar{x}_n), \tilde{v}(x)\}$ are not independent. By the observation of the last paragraph, we may choose $\{\tilde{v}(\bar{x}_1), \ldots, \tilde{v}(\bar{x}_n)\}$ to be independent.

The above implies that $z(x, \bar{x}_1, \ldots, \bar{x}_n) > 0$ for every $x \in A_{\bar{x}}$. The definition of z and the fact that V is finite imply that there must be $(\bar{v}_1, \ldots, \bar{v}_n) \in V^n$ and $v' \in V$, and a set $A \subset A_{\bar{x}_1,\ldots,\bar{x}_n}$ with $\lambda(A) > 0$, such that the expression

$$z(x, \bar{x}_1, \ldots, \bar{x}_n; v', \bar{v}_1, \ldots, \bar{v}_n)$$

is either always strictly positive, or strictly negative for all $x \in A$. Note that in either case, this implies $P\{\tilde{v}(\bar{x}_i) = \bar{v}_i,\ i = 1, \ldots, n\} > 0$ (from the definition of z and the fact that $\{\tilde{v}(\bar{x}_1), \ldots, \tilde{v}(\bar{x}_n)\}$ are independent). We also note that $P\{\tilde{v}(\bar{x}_i) = \bar{v}_i,\ i = 1, \ldots, n\} < 1$, otherwise $\{\tilde{v}(\bar{x}_1), \ldots, \tilde{v}(\bar{x}_n), \tilde{v}(x)\}$ would be trivially independent.

The remainder of the argument continues as in the $n = 1$ case, with the event S defined as $\{\tilde{v}(\bar{x}_1) = \bar{v}_1, \ldots, \tilde{v}(\bar{x}_n) = \bar{v}_n\}$. $\square$

It is clear from the proof that the claim also holds "n-by-n" in the sense that, for any $n \geqslant 1$, B.5 holds for all $n' \leqslant n$ if and only if B.4* holds for $n+1$.

Lemma A.3. *If P satisfies B.4*, then for any collection of random variables $\{\tilde{f}(x),\ x \in X\}$ such that $x \mapsto E\tilde{f}(x)$ is measurable and any $A \in \mathscr{A}$, the Pettis integral $\int_A \tilde{f}(x)\, d\lambda$ is constant almost surely, and equals $\int_A E\tilde{f}(x)\, d\lambda$.*

Proof. By definition, the Pettis integral $\int_A \tilde{f}(x)\, d\lambda$ satisfies, for every $z \in L^2$, the equation:

$$\left(z \middle| \int_A \tilde{f}(x)\, d\lambda\right) = cov\left(z, \int_A \tilde{f} d\lambda\right) + Ez\ E \int_A \tilde{f} d\lambda$$
$$= \int_A cov(z, \tilde{f}(x))\, d\lambda + Ez \int_A E\tilde{f}(x)\, d\lambda.$$

For any z such that $Ez = 0$,

$$cov\left(z, \int_A \tilde{f} d\lambda\right) = \int_A cov(z, \tilde{f}(x))\, d\lambda.$$

B.4* implies that any z of the form $z = \tilde{f}(x) - E\tilde{f}(x),\ x \in X$ is orthogonal to almost every $\tilde{f}(x)$, so the integral on the RHS is zero. This is also clearly true for every z that is a linear combination of random variables of the form $\tilde{f}(x) - E\tilde{f}(x),\ x \in X$ and, by continuity, to limits of such combinations. We conclude that

$$cov\left(z, \int_A \tilde{f} d\lambda\right) = 0$$

for every mean-zero z, so $\int_A \tilde{f} d\lambda$ is constant a.s. The last assertion, $E \int_A \tilde{f}(x)\, d\lambda = \int_A E\tilde{f}(x)\, d\lambda$, follows from Lemma A.1. □

Proof of Proposition A.4. Under B.3 and B.4, Theorem 4 implies that B.4* holds. The result then follows by applying Lemma A.3. □

A.4. Proof of Theorem 2

We begin with the proof of Lemma 1.

Proof of Lemma 1. Let $\mathscr{A}_n$ be the (finite) partition of X determined by conditioning on the first n features, and let $\overline{\mathscr{A}_n}$ be the σ-algebra it generates. By Proposition A.1,

$\overline{\mathscr{A}}$ is the σ-algebra generated by $\bigcup_{n=1}^{\infty} \overline{\mathscr{A}}_n$. Define $l^n : X \to \Delta^V$ by

$$l^n(x) = \begin{pmatrix} \sum_{i=1}^{2^n} EU(b_1, A_i)\chi_{A_i}(x) \\ \vdots \\ \sum_{i=1}^{2^n} EU(b_K, A_i)\chi_{A_i}(x) \end{pmatrix},$$

where χ_{A_i} is the characteristic function of A_i. Note that l^n is $\mathscr{A}_n$-measurable.

We now show that l^n is a martingale with respect to the filtration $\{\overline{\mathscr{A}}_n\}$. That $E|l^n| < \infty$ is obvious. We need only show that $E[l^{n+1}|\overline{\mathscr{A}}_n](x) = l^n(x)$, λ-a.s. Let A be an element of the partition $\mathscr{A}_n$, and let $\{B_1, B_2\}$ be the partition of A in $\mathscr{A}_{n+1}$ (recall our assumption that each feature can take only two possible values). Then,

$$\begin{aligned} E[l^{n+1} \mid x \in A] &= \lambda(B_1 \mid A) \begin{pmatrix} EU(b_1, B_1) \\ \vdots \\ EU(b_K, B_1) \end{pmatrix} + \lambda(B_2 \mid A) \begin{pmatrix} EU(b_1, B_2) \\ \vdots \\ EU(b_{K-1}, B_2) \end{pmatrix} \\ &= \begin{pmatrix} EU(b_1, A) \\ \vdots \\ EU(b_K, A) \end{pmatrix} = l^n(x) \end{aligned}$$

(the last equality follows from A.2*). Thus, $(l^n, \overline{\mathscr{A}}_n)$ forms a martingale. By the martingale convergence theorem, there is an essentially unique $\overline{\mathscr{A}}$-measurable function, $l : X \to \Delta^V$ such that $\lim_{n\to\infty} l^n(x) = l(x)$ for λ-a.e. [17, p. 476].

To complete the proof, we show that for any $A \in \mathscr{A}$, and $b_k \in B$:

$$\begin{aligned} \int_A l_k(x)\, d\lambda &= \int_A \lim_{n\to\infty} l_k^n(x)\, d\lambda \\ &= \lim_{n\to\infty} \int_A l_k^n(x)\, d\lambda \quad (a) \\ &= \lim_{n\to\infty} \int_A \sum_{i=1}^{2^n-1} EU(b_k, A_i)\chi_{A_i}(x)\, d\lambda \\ &= \lim_{n\to\infty} \sum_{i=1}^{2^n-1} EU(b_k, A_i) \int_A \chi_{A_i}(x)\, d\lambda \\ &= \lim_{n\to\infty} \sum_{i=1}^{2^n-1} EU(b_k, A_i)\lambda(A \cap A_i) \\ &= \lambda(A) EU(b_k, A). \end{aligned}$$

Here, (a) follows from the dominated convergence theorem and the fact that l^n converges to l for almost every x. $\square$

Proof of the theorem. We first define the state space (Ω, Σ) using standard Kolmogorov construction [17]. The set of states Ω is the set of all functions $\omega : X \to V$. Viewing Ω as a product space, projection on the x coordinate, denoted $\pi_x : \Omega \to V$, is $\pi_x(\omega) = \omega(x)$. We define Σ to be the σ-algebra generated by projections; that is, Σ is the smallest σ-algebra containing all sets of the form $\{\omega : \pi_x^{-1}(v)\}$, for $x \in X$ and $v \in V$. For each x, define $\tilde{v}_x : \Omega \to V$ by $\tilde{v}(x, \omega) = \omega(x)$ (i.e., the projection of the state ω on x). Clearly, $\tilde{v}(x)$ is measurable with respect to Σ.

Let $l(x)$ denote the (essentially unique) local weights derived from behavior in Lemma 1. We use $l(x)$ to define a probability distribution P on (Ω, Σ). For every finite set of contingencies $\{x_1, \ldots, x_S\}$, define the probability distribution $P_{\{x_1,\ldots,x_S\}}$ on the finite set V^S so that the random variables $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_S)\}$ are independently distributed with $E\tilde{v}(x_s) = l(x_s)$. Clearly, the $P_{\{x_1,\ldots,x_S\}}$'s satisfy the consistency condition in Kolmogorov's Extension Theorem [17, Theorem 4, p. 165], so there is a unique probability measure P that agrees with every finite dimensional distribution $P_{\{x_1,\ldots,x_S\}}$.[44]

Set $\mathscr{M} = \Sigma$. Then B.1 and B.3–4 are satisfied by construction. B.2 follows from the fact that the function l is measurable, so any expression of the form $\prod_{i=1}^{n} l(x_i)$ is also measurable. It only remains to show that P rationalizes U. Since the $\tilde{v}$'s are independent by construction, Lemma A.3 implies that for every A, $\int_A \tilde{v}(x)\, d\lambda$ is degenerate random variable, equal a.s. to $\int_A E\tilde{v}(x)\, d\lambda = \int_A l(x)\, d\lambda$. $\square$

A.5. Proof of Theorem 3

Suppose that a problem is complex over A. Then $\sup_{f \in \mathscr{F}} E\tilde{U}(f, A) < 1$. We want to show that there is $A' \subset A$, with $\lambda(A') > 0$ such that $l(x) \notin V$ for almost every $x \in A'$. Suppose to the contrary that for almost all $x \in A$ we have $l(x) := E\tilde{v}(x) \in V$. Now define $f(x)$ as follows. For $x \in A$, if $l(x) = v_j$, $j \in \{1, \ldots, K\}$, let $f(x) = b_j$. For $x \notin A$, choose $f(x)$ arbitrarily. Note that $E[\tilde{v}(f(x), x)] = 1$. Thus $E\tilde{U}(f, A) = \frac{1}{\lambda(A)} \int_A E[\tilde{v}(f(x), x)]\, d\lambda(x) = \frac{1}{\lambda(A)} \int_A d\lambda(x) = 1$. Contradiction.

In the other direction suppose there exists $A' \subset A$, with $\lambda(A') > 0$ such that $l(x) \notin V$ for almost every $x \in A'$. Let $l^{\max}(x) = \max\{l_1(x), \ldots, l_K(x)\}$. Note that $l^{\max}(x)$ is measurable and $l^{\max}(x) < 1$ for almost all $x \in A'$. Thus $\sup_{f \in \mathscr{F}} E\tilde{U}(f, A') \leqslant \frac{1}{\lambda(A')} \int_{A'} l^{\max}(x)\, d\lambda(x) < 1$. Since $\lambda(A') > 0$, this implies $\sup_{f \in \mathscr{F}} E\tilde{U}(f, A) < 1$.

[44] The consistency condition here says that for any two subsets of contingencies $\{y_1, \ldots, y_T\} \subset \{x_1, \ldots, x_S\}$, and any $(v_1, \ldots, v_T) \in V^T$, $P_{\{y_1,\ldots,y_T\}}(v_1, \ldots, v_T) = P_{\{x_1,\ldots,x_S\}}(v_1, \ldots, v_T)$. This is true by our construction of the P's as independent distributions based on the same expectation function $l(x)$.

A.6. Proof of Theorem 5

From our earlier results we may assume that B.5 is satisfied. Fix an action b; we simplify notation by writing $f(x) = E\tilde{v}(b,x)$. We prove the claim in the case $A = X$; the argument does not depend on this assumption, however.

We show first that for λ^∞-a.e. sequence of instances $\{x_1, \ldots\}$,

$$EU(b,X) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n). \tag{+}$$

Suppose first that f is a step function, i.e., takes finitely many values $\alpha_1, \ldots, \alpha_K$, then from B.5, the law of large numbers, and the definition of the integral, it follows that there is a set of λ^∞-measure 1 of sequences $\{x_1, \ldots\}$ such that for any such sequence,

$$\int_X f\, d\lambda = \sum_{k=1}^{K} \alpha_k\, \lambda(f^{-1}(\alpha_k)) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n).$$

Next, suppose that f is the *uniform* limit of a sequence of step functions, $\{f_m\}$. Then (using Eq. (2)), for any $\varepsilon > 0$,

$$\begin{aligned} EU(b,X) &= \int_X f\, d\lambda \\ &= \lim_{m\to\infty} \int_X f_m\, d\lambda \\ &= \lim_{m\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f_m(x_n) \\ &\leqslant \lim_{m\to\infty} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} [f(x_n) + \varepsilon] \\ &= \varepsilon + \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n). \end{aligned}$$

We may similarly conclude that $EU(b,X) \geqslant -\varepsilon + \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n)$. Since ε is arbitrarily, we have $EU(b,X) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n)$. Finally, for any measurable function f and any $\eta > 0$ there is $B \subset X$, $\lambda(B) > 1 - \eta$ and a sequence of step functions $\{f_m\}$ that converges uniformly on B [16, Proposition 23, p. 71]. Taking η small enough shows (+) for any measurable function f.

Since the variances of the random variables $\tilde{v}(x)$ are uniformly bounded, by the law of large numbers [17, Theorem 2, p. 364], there is P-probability 1 that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{v}(b, x_n, \omega) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n).$$

From this and $EU(b,X) = \int_X f\, d\lambda$, we conclude that, with P-probability 1,

$$EU(b,X) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \tilde{v}(b, x_n, \omega).$$

A.7. Proof of Theorem 6

The proof of Theorem 4 already establishes that B.4* must hold: there we have shown that the failure of B.4* implies the existence of an event S such that conditioning on this event changes expected utility for some option on some subset. Given this, the utility distribution cannot be degenerate, contradicting A.3.

It only remains to show:

Proposition A.5. *Any probabilistic model rationalizing U that satisfies* A.1–2 *must satisfy* B.3*.

Proof. We first note that if B.3* fails at x, then the event $\{\omega : E[\tilde{v}(x)|\mathscr{M}](\omega) \in V\}$ has probability less than 1 (i.e., the conditional expectation $E[\tilde{v}(x)|\mathscr{M}](\omega)$ is in the interior of Δ^V with positive probability). To see this, note first that the conditional expectations $E[\tilde{v}(x)|\mathscr{M}]$ must, by definition, be $\mathscr{M}$-measurable, i.e., $E[\tilde{v}(x)|\mathscr{M}]^{-1}(v) \in \mathscr{M}$ for any $v \in V$. Note also that $E[\tilde{v}(x)|\mathscr{M}]^{-1}(v) \subset \{\omega : \tilde{v}(x) = v\}$. If the claim failed (i.e., $E[\tilde{v}(x)|\mathscr{M}](\omega) \in V$ with probability 1), then $P\{\bigcup_{v\in V} E[\tilde{v}(x)|\mathscr{M}]^{-1}(v)\} = 1$. But since $P\{\bigcup_{v\in V} \{\omega : \tilde{v}(x) = v\}\} = 1$, the difference $\{\omega : \tilde{v}(x) = v\} - E[\tilde{v}(x)|\mathscr{M}]^{-1}(v)$ must have probability 0 for every v. Augmenting one of the sets $E[\tilde{v}(x)|\mathscr{M}]^{-1}(v)$ by the probability zero set $\Omega - \bigcup_{v\in V} E[\tilde{v}(x)|\mathscr{M}]^{-1}(v)$ if necessary, we obtain a partition $\mathscr{M}_x$ of Ω that satisfies the requirements in B.3*, contradicting that this condition fails at x.

To prove the proposition, suppose by way of contradiction that there is $A \in \mathscr{A}$ on which B.3* fails. By the previous argument, for any $x \in A$, $E \max_{b\in B} E[\tilde{v}(b,x)|\mathscr{M}](\omega) < 1$. Let g denote the "grand option" $g(x) = B$ for all x. Then,

$$\begin{aligned}\lambda(A) E\tilde{U}(g,A) &= E \int_A \max_{b\in B} E[\tilde{v}(b,x)|\mathscr{M}](\omega)\, d\lambda(x) \\ &= \int_A E \max_{b\in B} E[\tilde{v}(b,x)|\mathscr{M}](\omega)\, d\lambda(x) < 1,\end{aligned}$$

where the second equality follows from Lemma A.1 (with $S = \Omega$) and the integral in the first equation is the Pettis integral. But, Condition A.1 requires that the agent puts value 1 on g, contradicting the assumption that P rationalizes U. □

A.8. Proof of Theorem 7

We first note the following:

Lemma A.4. *If $(P, \mathcal{M})$ and $(P', \mathcal{M}')$ are two probabilistic models rationalizing behavior U satisfying* A.1–2. *Let $l(x)$ denote the local weights given by Lemma* 1. *Then for almost every $x \in X$ and for every v_k, $P\{\tilde{v}(x) = v_k\} = P'\{\tilde{v}(x) = v_k\} = l_k(x)$.*

Proof. Fix k, and consider the constant rule b_k. By Lemma 1, $\lambda(A)EU(b_k, A) = \int_A l_k(x)\, d\lambda$. Since P rationalizes U, we must also have $\lambda(A)EU(b_k, A) = E\int_A \tilde{v}(b_k, x)\, d\lambda = \int_A E\tilde{v}(b_k, x)\, d\lambda = \int_A P\{\tilde{v}(x) = v_k\}\, d\lambda$. Thus, we have that for every A:

$$\int_A l_k(x)\, d\lambda = \int_A P\{\tilde{v}(x) = v_k\}\, d\lambda.$$

But this implies that $l_k(x) = P\{\tilde{v}(x) = v_k\}$ for almost every x. The claim of the lemma follows by noting that a similar argument holds also for P'. □

Proof of Part (i). By Proposition A.5 both $(P, \mathcal{M})$ and $(P', \mathcal{M}')$ must satisfy B.3*. The claim now follows by applying the previous lemma. □

Proof of Part (ii). Clearly, P and P' must agree on the local weights $l(x)$ on a set of measure 1 of contingencies. Given $n \geqslant 2$, from the previous proposition, there is a set A^n of probability 1 such that both P and P' satisfy B.5 on A^n. But this implies that P and P' have the same joint distribution on $\{\tilde{v}(x_1), \ldots, \tilde{v}(x_n)\}$ for every set of contingencies $\{x_1, \ldots, x_n\} \in A^n$. □

A.9. Proof of Theorem 8

To prove part (i), define first the correspondence

$$G(x) = \underset{C \in 2^B \setminus \emptyset}{\operatorname{argmax}} \sum_{v_1, v_2} l(x)(v_1, v_2) v_1(b(C, v_2)).$$

Our problem is to show that we can select an option g, such that $g(x) \in G(x)$ for all x. To prove this, we show that G is a measurable correspondence, in which case a measurable selection exists by Klein and Thompson [10, p. 163]. That is, we must show that for every closed $F \subset 2^B \setminus \emptyset$,

$$\{x : G(x) \cap F \neq \emptyset\} \in \overline{\mathcal{A}}.$$

Since $2^B \setminus \emptyset$ is finite, we may restrict attention to singleton sets $F = \{C\}$ (since $\overline{\mathcal{A}}$ is preserved under finite unions and intersections). Thus, for $C \in 2^B \setminus \emptyset$, we have

$$\begin{aligned}\{x : G(x) \cap \{C\} \neq \emptyset\} &= \{x : C \in G(x)\} \\ &= \{x : E_{l(x)} v_1(b(C, v_2)) \geqslant E_{l(x)} v_1(b(C', v_2)), \forall C' \in 2^B \backslash \emptyset\} \\ &= \bigcap_{C' \in 2^B \backslash \emptyset} \{x : E_{l(x)} v_1(b(C, v_2)) \geqslant E_{l(x)} v_1(b(C', v_2))\} \\ &= \bigcap_{C' \in 2^B \backslash \emptyset} \{x : E_{l(x)} [v_1(b(C, v_2)) - v_1(b(C', v_2))] \geqslant 0\}. \quad (**)\end{aligned}$$

The expression $E_{l(x)}[v_1(b(C, v_2)) - v_1(b(C', v_2))]$ depends only on the probability weights $l(x)(v_1, v_2)$. Since the joint distribution $l(x)$ is measurable with respect to $\overline{\mathscr{A}}$, every set in the finite intersection in (**) is in $\overline{\mathscr{A}}$. Thus, G is indeed a measurable correspondence, so a measurable selection exists.

To prove part (ii), suppose that g is a locally optimal option that we know exists from part (i). First we show that $U(g) \geqslant U(g')$ for all $g' \in \mathscr{G}$ and thus g is a globally optimal option.

We know that

$$U(g) = \int_X E_{l(x)} v_1(b(g(x), v_2)) \, d\lambda$$

almost every x. Take any other option $g' \in \mathscr{G}$. For almost every x

$$E_{l(x)} v_1(b(g(x), v_2)) \geqslant E_{l(x)} v_1(b(g'(x), v_2))$$

optimal option is also globally optimal.

Now to prove the other direction suppose a globally optimal option $g^+ \in \mathscr{G}$ is not locally optimal. Let $g \in \mathscr{G}$ be a locally optimal option. Then for almost every x

$$E_{l(x)} v_1(b(g(x), v_2)) \geqslant E_{l(x)} v_1(b(g^+(x), v_2)),$$

$U(g) > U(g^+)$, which is a contradiction.

References

[1] L. Anderlini, L. Felli, Incomplete written contracts: undescribable states of nature, Quart. J. Econom. 109 (1994) 1085–1124.

[2] L. Anderlini, H. Sabourian, Cooperation and effective computability, Econometrica 63 (1995) 1337–1370.

[3] K. Binmore, Modeling rational players, Part I, Econom. Philos. 3 (1987) 179–214.

[4] E. Dekel, B. Lipman, A. Rustichini, Representing preferences with a unique subjective state space, Econometrica 69 (2001) 894–934.

[5] L. Devroye, L. Gyorfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, Berlin, 1996.

[6] J. Diestel, J. Uhl, Vector Measures, Mathematical Surveys No. 15, American Mathematical Society, Providence, RI, 1977.

[7] I. Gilboa, D. Schmeidler, Infinite histories and steady orbits in repeated games, Games Econom. Behav. 6 (1994) 370–399.

[8] J. Holland, K. Holyoak, R. Nisbett, P. Thagard, Induction, MIT Press, Cambridge, MA, 1989.

[9] J. Jordan, Classification Dynamics, Mimeo, University of Minnesota, 1995.
[10] E. Klein, A.C. Thompson, Theory of Correspondences, Wiley–Interscience Publication, New York, 1984.
[11] D. Kreps, A representation theorem for "preference for flexibility", Econometrica 47 (1979) 565–577.
[12] D. Kreps, Static choice in the presence of unforeseen contingencies, in: O. Hart, P. Dasgupta, D. Gale, E. Maskin (Eds.), Economic Analysis of Markets and Games, MIT Press, Cambridge, MA, 1992, pp. 258–281.
[13] B. Lipman, Information processing and bounded rationality: a survey, Canad. J. Econom. 28 (1995) 42–67.
[14] K. Nehring, Preference for flexibility in a savage framework, Econometrica 67 (1999) 101–120.
[15] E. Ozdenoren, Completing the state space with subjective states, J. Econom. Theory 105 (2002) 531–539.
[16] H.L. Royden, Real Analysis, MacMillan Publishing Co., New York, 1968.
[17] A.N. Shiryayev, Probability, Springer, New-York, 1984.
[18] H. Uhlig, A law of large numbers for large economies, Econom. Theory 8 (1996) 41–50.