Axioms for a Class of Algorithms of Sequential Decision Making

Murali Agastya[1], Arkadii Slinko[2]

[1] Economics Discipline, University of Sydney NSW 2006 Australia
m.agastya@econ.usyd.edu.au
[2] Department of Mathematics, The University of Auckland, Private Bag 92019,
Auckland, New Zealand
a.slinko@auckland.ac.nz

Abstract. We axiomatically characterise a class of algorithms for making sequential decisions in situations of complete ignorance. These algorithms assume that a decision maker (DM) (human or or a software agent) has exogenously defined utilities for prizes and she uses the empirical distribution of prizes to calculate the "expected utility" of each action maximising this expected utility at each stage of the decision making process. We show that this class of algorithms is defined by three simple axioms that highlight the independence of the given actions, the bounded rationality of the agent, and the principle of insufficient reason at margin.

Key words: sequential decision making, ex-post rationality, fictitious play, multiset

1 Introduction

Consider a Decision Maker (DM) who has to repeatedly choose from a finite set of actions. Each action results in a random reward, also drawn from a finite set. The environment is complex in the sense that the DM is either unable to offer a complete description of the states of the world or is unable to construct a meaningful prior probability distribution. Naturally, the well established Bayesian methods of say [12] or [1] would then be inapplicable.

Our approach is to postulate that the DM has a preference relation defined directly over the set of actions which is updated over time in response to the sequences of observed rewards. Thus, if $\mathcal{A}$ denotes the set of all actions and H the set of all histories, the DM is completely described by the family $D := (\succeq_{h_t})_{h_t \in H}$, where $\succeq_{h_t} \subseteq \mathcal{A} \times \mathcal{A}$ is a well defined preference relation on the actions following a history h_t at date t. A history consists of the sequences of rewards, drawn from a finite set $\mathcal{R}$, that are obtained over time to each of the actions. Later we will impose axioms on D of procedural rationality type.

There is a considerable literature in economics and psychology on a variety of "stimulus-response" models of individual choice behavior. In these models, the DM does not attempt to learn the environment, instead she looks at the past experiences and takes her decisions on the basis of her observations. Most of this literature prescribes some boundedly rational rule(s) for updating and the focus is on analysis of implied adaptive dynamics. These imputed rules of updating vary widely. They range from modifications of fictitious play and reinforcement learning to imitation of peers etc. See for example [2], [13], [7] and the references therein.

Our approach outlined above is different. We do not consider any particular updating rules but impose axioms on the updating procedure. These axioms impose some structural restrictions and postulate certain independence and we derive an ex-post utility representation for such a DM. This approach may be found in [4] where they axiomatically characterised replicator dynamics which makes [4] the closest relative of this paper. We note that the Case Based Decision Theory of Gilboa and Schmeidler [8], [9] is not applicable due to the assumption of infinitude of cases and the Archimedean axiom that they impose.

Chapter 2 introduces the model, Chapter 3 defines the ex-post utility representation, Chapter 4 introduces the axioms, Chapter 5 formulates the main theorem and outlines its proof, and finally Section 6 fills all the gaps and completes the proof of the main theorem.

2 The Model

A Decision Maker must choose from a finite set of m actions $\mathcal{A} = \{a_1, \ldots, a_m\}$, at each moment $t = 0, 1, 2, \ldots$. Every action results in a reward, drawn from a finite set $\mathcal{R} = \{1, \ldots, n\}$. The rewards are governed by a stochastic process unknown to the DM. Following her choice at date t, the vector of realised rewards, $\mathbf{r}_t = (r_1^{(t)}, \ldots, r_m^{(t)})$, where $r_i^{(t)}$ is the reward to action a_i at moment t, is revealed to the DM. Thus the DM observes the rewards for *all* actions and not only for the one she has chosen. A *history* at date t is a sequence of vectors of rewards $h_t = (\mathbf{r}_0, \ldots, \mathbf{r}_{t-1})$.

The sequential decisions of the DM are guided by the following principle. Following any history h_t, the DM works out a preference relation[3] $\succeq_{h_t}$ on the set of actions $\mathcal{A}$. At date t she chooses one of the maximal actions with respect to $\succeq_{h_t}$, observes the set of outcomes $\mathbf{r}_t$ and calculates a new preference relation $\succeq_{h_{t+1}}$ where $h_{t+1} = (h_t, \mathbf{r}_t)$. At the outset the DM is indifferent between all the actions so she chooses a random one.

Let H_t denote the set of all histories at date t and $H = \bigcup_{t \geq 1} H_t$. Thus, the family of preference relations $D := (\succeq_h)_{h \in H}$ completely describes the DM. Our objective is to discuss the behavior of this learning agent through the imposition of certain axioms that encapsulate the DM's procedural rationality. For a DM satisfying these axioms we will derive an ex-post utility representation theorem that is based on the empirical distribution of rewards in any history.

Before proceeding any further with the analysis, it is important to point out two salient features of the above formulation of the DM.

[3] Throughout, by a *preference relation* on any set, we mean a binary relation that is a complete, transitive and reflexive ordering of the elements.

First, as in [4], a history describes the rewards to all the actions in each period, including those that the DM did not choose. This implicitly assumes that decisions are taken in a social context where other people are taking other actions and the rewards for each action are publicly announced. Examples of such situations are numerous and include investing in a share market and betting on horses. Relaxing this assumption of learning in a social context is a topic of future research.

Second, note that the description requires a preference on actions to be specified after every conceivable history. This is much in the spirit of the theoretical developments in virtually all decision theory. The presumption underlying such an abstraction is that any subset of these acts may be presented to the DM and that a necessary aspect of a theory is that it is applicable with sufficient generality. Given the temporal nature of the problem at hand this assumption may be quite natural. For, all conceivable histories may appear by assuming that the underlying random process generates every $\mathbf{r} \in \mathcal{R}^m$ with a positive probability.

We make a non-triviality assumption on D for the rest of this paper. We assume that the DM is not indifferent between all actions following all histories.

3 Multisets & Ex-Post Utility Maximisation

Here we will introduce the rule (a class of algorithms) that we will eventually axiomatise. For this rule, the number of times different rewards accrue to given action during a history is important. To progress further, we will need to introduce the idea of a *multiset*. A multiset over an underlying set may contain several copies of any given element of the latter. The number of copies of an element is called its *multiplicity*. Our interest is in multisets over $\mathcal{R}$. Therefore, multiset μ is identified with a vector $\mu = (\mu(1), \ldots, \mu(n)) \in \mathbb{Z}_+^n$, where $\mu(i)$ is the multiplicity of the ith prize and the *cardinality* of this multiset is $\sum\limits_{i=1}^{n} \mu(i)$. Let $\mathcal{P}_t[n]$ denote the subset of all such multisets of cardinality t whereupon

$$\mathcal{P}[n] = \bigcup_{t=1}^{\infty} \mathcal{P}_t[n] \tag{1}$$

denotes the set of all non-empty multisets over $\mathcal{R}$. Mostly, we will write $\mathcal{P}_t$ instead of $\mathcal{P}_t[n]$ when the number of prizes is clear. The *union* of $\mu, \nu \in \mathcal{P}$ is defined as the multiset $\mu \cup \nu$ for which $(\mu \cup \nu)(i) = \mu(i) + \nu(i)$ for any $i \in \mathcal{R}$. Observe that whenever $\mu \in \mathcal{P}_t$ and $\nu \in \mathcal{P}_s$, then $\mu \cup \nu \in \mathcal{P}_{t+s}$.

Given any history $h \in H_t$, let $\mu_i(a, h)$ denote the number of times the reward i has occured in the history of rewards $h(a)$ corresponding to action a and $\mu(a, h) = (\mu_1(a, h), \ldots, \mu_n(a, h))$.

For any two vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ of $\mathbb{R}^n$, we let $x \cdot y$ denote their dot product, i.e. $x \cdot y = \sum_{i=1}^{n} x_i y_i$.

Here comes the rule. A DM applying this rule must have exogenously defined utilities of the prizes. Let $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ be the vector of her utilities, where

u_i is the utility of the ith prize. At any moment t the DM calculates the total utility of the prices for each given action in the past and chooses the action which performed best in the past and for which the total utility of prizes is at least as high as for any other action. In other words she chooses any action belonging to $\text{argmax}_i(\mu(a_i, h) \cdot \mathbf{u})$.

The problem of the DM is that she does not know the probabilities. In the absence of any knowledge about the environment the most reasonable thing to do is to assume that the process of generating rewards is stationary and to replace the probabilities of the rewards with their empirical frequencies. Due to the assumed stationarity of the process she expects that these frequencies approximate probabilities well (at least in the limit), so in a way the DM acts as an expected utility maximiser relative to the empirical distribution of rewards. This rule is very much in the spirit of the so-called fictitious play[4].

There is a good reason to allow the DM to use different vectors of utilities at different moments. This will allow the DM, at each moment, to refine her utilities from the previous period to reflect her preferences on larger multisets and longer histories. An obvious consistency condition must however be imposed: we require that the vector of utilities the DM uses at time t must be also suitable to evaluate actions in all previous moments.

Definition 1 (Ex-Post Utility Representation). *A sequence $(\mathbf{u}_t)_{t\geq 1}$ of vectors of $\mathbb{R}^n_+$ is said to be an* ex-post utility representation *of $D = (\succeq_h)_{h\in H}$ if, for all $t \geq 1$,*

$$a \succeq_h b \;\Leftrightarrow\; \mu(a, h) \cdot \mathbf{u}_t \geq \mu(b, h) \cdot \mathbf{u}_t \quad \forall\, a, b \in \mathcal{A}, \;\; \forall h \in H_s, \tag{2}$$

for all $s \leq t$. The representation is said to be global *if $\mathbf{u}_t \equiv \mathbf{u}$ for some $\mathbf{u} \in \mathbb{R}^n_+$.*

In what follows, we shall say that the DM is *ex-post rational* if she admits an ex-post utility representation.

We emphasise that the object that is of ultimate interest is the ranking of the actions following a history. The utility representation of a DM involves assigning non-negative weights to the rewards. However this assignment is not unique. A sequence $(\mathbf{u}'_t)_{t\geq 1}$ obtained by applying some positive affine transformations $\mathbf{u}'_t = \alpha_t \mathbf{u}_t + \beta_t$ (with $\alpha_t > 0$) to a given utility representation $(\mathbf{u}_t)_{t\geq 1}$ is also a utility representation.

Therefore, we should adopt a certain normalisation. By $\varDelta \subseteq \mathbb{R}^m$ we denote the $m - 1$ dimensional unit simplex consisting of all non-negative vectors $\mathbf{x} = (x_1, \ldots, x_n)$ such that $x_1 + \ldots + x_n = 1$. Due to the non-triviality assumption, for any $\mathbf{u}_t$, not all utilities are equal. Hence we may assume that at any $\mathbf{u}_t = (u_1, \ldots, u_n)$ in a representation, $\min\{u_i\} = 0$. We may further normalise the coordinates to sum to one so that every $\mathbf{u}_t$ may be assumed to lie in one of the following subsets of the unit simplex:

$$\varDelta^i = \{\mathbf{u} = (u_1, \ldots, u_n) \in \varDelta \mid u_i = 0\}, \tag{3}$$

[4] Ficitious play was introduced by [3]. See [5] for variations of fictitious play.

which is one of the facets[5] of $\varDelta$.

4 Axioms

Next, we turn to the axioms that are necessary and sufficient for D to admit an ex-post utility representation. The first axiom says that in comparing a pair of actions, the information regarding the other actions is irrelevant. Intuitively, this amounts to asserting that the agent believes that she is facing an environment in which consequences of actions are statistically uncorrelated.

Axiom 1 *Consider h_t, h'_t and actions $a, b \in \mathcal{A}$ such that $h_t(a) = h'_t(a)$ and $h_t(b) = h'_t(b)$. Then $a \succeq_{h_t} b$ if and only if $a \succeq_{h'_t} b$.*

Although the agent has the entire history at her disposal, we postulate that for any action, the algorithm only tracks the number of times different rewards were realised. This means that the agent believes that she is facing an environment generated by a stationary stochastic process.

Axiom 2 *Consider a history h_t at which for two actions a and b the multisets of prizes are the same, i.e. $\mu(a, h_t) = \mu(b, h_t)$. Then $a \sim_{h_t} b$.*

The next axiom describes how the DM learns to revise her preferences in response to new information.

Axiom 3 *For any history h_t and any $r \in \mathcal{R}$, if $h_{t+1} = (h_t, \mathbf{r}_t)$ where $\mathbf{r}_t = (r, \ldots, r)$, then $\succeq_{h_{t+1}} = \succeq_{h_t}$.*

Due to Axiom 1, it implies that if at some history h_t the DM (weakly) prefers an action a to b and in the current period both these actions yield the same reward, according to the next axiom, the DM continues to prefer a to b. We view Axiom 3 as loosely capturing the "principle of insufficient reason at the margin".

5 The Main Theorem

In this section we will formulate and give an outline of the proof of the main theorem. Recall that $ri(C)$ denotes the relative interior of a convex set C.

Theorem 1 (Representation Theorem). *Suppose $m \geq 3$. The following are equivalent:*

1. *$D = (\succeq_h)_{h \in H}$ satisfies Axioms 1– 3.*
2. *D has an ex-post utility representation. There exists a unique sequence of non-empty convex polytopes $(U_t)_{t \geq 0}$ such that $U_t \subseteq \varDelta^i$ for some i and*
   (a) $U_{t+1} \subseteq U_t$ for all $t \geq 1$.

[5] Facet of a polytope is a face of the maximal dimension.

(b) $\bigcap_{t=1}^{\infty} U_t$ *consists of a single vector.*

(c) A sequence $(\mathbf{u}_t)_{t\geq 1}$ *of vectors of* $\mathbb{R}_n^+$ *is a utility representation of* D *if and only if* $\mathbf{u}_t$ *is a positive affine transformation of some* $\mathbf{u}'_t \in ri(U_t)$*. In particular, any sequence* $(\mathbf{u}_t)_{t\geq 1}$ *such that* $\mathbf{u}_t \in ri(U_t)$ *is a utility representation of* D*.*

(d) If $\bigcap_{t=1}^{\infty} U_t$ *is in the interior of every* U_t*, then the representation is global.*

Remark 1. We note that despite an expected-utility-like calculation that is implicitly involved in Theorem 1, it is important to note that there is no connection with the expected utility hypothesis. Our DM is only ex-post rational.

Proof. It is easy to show that any DM with an ex-post utility representation satisfies the axioms. Let us show the non-trivial part of the theorem, which is, $1 \Rightarrow 2$. We begin by defining, for each $t \geq 1$, a binary relation $\succeq_t^*$ on $\mathcal{P}_t = \mathcal{P}_t[n]$ as follows: for any $\mu, \nu \in \mathcal{P}_t$,

$$\begin{aligned} \mu \succeq_t^* \nu \iff & \text{ there exists } a, b \in \mathcal{A} \text{ and a history } h_t \in H_t \\ & \text{such that } \mu = \mu(a, h_t) \text{ and } \nu = \mu(b, h_t) \text{ and} \\ & a \succeq_{h_t} b \end{aligned} \tag{4}$$

Analogously we define also a strict version $\succ_t^*$ of $\succeq_t^*$. The latter needs to be proved to be antisymmetric. For, for a certain pair of multisets $\mu, \nu \in \mathcal{P}_t$, different choices of histories and actions can result in both $\mu \succeq_t^* \nu$ and $\nu \succ_t^* \mu$ at once. However, we claim that:

Claim 1 *For any* $a, b, c, d \in \mathcal{A}$ *and any two histories* $h_t, h'_t \in H_t$ *such that* $\mu(a, h_t) = \mu(c, h'_t)$ *and* $\mu(b, h_t) = \mu(d, h'_t)$*,*

$$a \succeq_{h_t} b \iff c \succeq_{h_{t'}} d.$$

The above claim ensures that $\succ_t^*$ is antisymmetric since $\succ_h$ is antisymmetric. It is now also clear that the sequence $\succeq^* = (\succeq_t^*)_{t\geq 1}$ inherits the non-triviality assumption in the sense that for some t the relation $\succeq_t^*$ is not a complete indifference. Next we claim that

Claim 2 $\succeq_t^*$ *is a preference ordering on* $\mathcal{P}_t$*.*

Both of the above claims only rely on Axiom 1 and Axiom 2. The proofs of Claim 1 and Claim 2 are straightforward but nevertheless relegated to the Appendix. By a repeated application of Axiom 3, we see at once that

Claim 3 *The sequence* $\succeq^* = (\succeq_t^*)_{t\geq 1}$ *satisfies the following property: for any* $\mu, \nu \in \mathcal{P}_t$ *and any* $\xi \in \mathcal{P}_s$*,*

$$\mu \succeq_t^* \nu \iff \mu \cup \xi \succeq_{t+s}^* \nu \cup \xi \tag{5}$$

for all $t, s \in \mathbb{Z}_+$*.*

The remainder of the proof will follow from Theorem 2 proved in the next section and further considerations.

The requirement in Theorem 1 that there are at least three actions for the agent to choose from cannot be dropped. To see this we have the following counter-example with $m = 2$.

Example 1. Pick any utility vector $\mathbf{u} = (u_1, \ldots, u_n)$ for the rewards and define D as follows:

Following a history $h_t \in H_t$,

1. If $\mu(a_i, h_t) \cdot \mathbf{u} > \mu(a_j, h_t) \cdot \mathbf{u}$, the DM strictly prefers a_i to a_j, where $i \neq j$ and $i, j = 1, 2$.
2. If $\mu(a_1, h_t) \cdot \mathbf{u} = \mu(a_2, h_t) \cdot \mathbf{u}$, then
   (a) If the corresponding multisets of rewards are the same, i.e. $\mu(a_1, h_t) = \mu(a_2, h_t)$, then the actions are indifferent.
   (b) Otherwise a_1 is strictly preferred.

It may be readily verified that $\mathcal{D}$ described above satisfies Axioms 1-3 but does not admit an ex-post utility representation.

6 Orders on Multisets and Their Utility Representation

This section completes the proof of the main theorem.

As we know from Section 2, multisets of cardinality t are important for a DM as they are closely related to histories at date t. The DM has to be able to compare them for all t. At the same time in the context of this paper it does not make much sense to compare multisets of cardinalities of different sizes (it would if we had missing observations). Due to this, our main object in this subsection is a family of orders $(\succeq_t)_{t \geq 1}$, where $\succeq_t$ is an order on $\mathcal{P}_t$. In this case we denote by $\succeq$ the partial (but reflexive and transitive) binary relation on $\mathcal{P}$ whereby for any $\mu, \nu \in \mathcal{P}$, where $\mu \succeq \nu$ if both μ and ν are of the same cardinality, say t, and $\mu \succeq_t \nu$ and $\mu \succeq \nu$ is undefined otherwise.

To complete the proof of the main theorem we must study orders on $\mathcal{P}$ with the property (5). Due to their importance we will give them a special name.

Definition 2 (Consistency). *An order $\succeq = (\succeq_t)_{t \geq 1}$ on $\mathcal{P}$ is said to be* consistent *if it satisfies the condition (5) from Claim 3, that is, for any $\mu, \nu \in \mathcal{P}_t$ and any $\xi \in \mathcal{P}_s$,*

$$\mu \succeq_t \nu \Longleftrightarrow \mu \cup \xi \succeq_{t+s} \nu \cup \xi. \tag{6}$$

We note that, due to the twosidedness of the arrow in (6), we have also

$$\mu \succ_t \nu \Longleftrightarrow \mu \cup \xi \succ_{t+s} \nu \cup \xi. \tag{7}$$

One consistent linear order that immediately comes to our mind is the lexicographic order which is an extension of a linear order on $\mathcal{R}$. But, of course, this is not the only consistent order. Now we will define a large class of consistent orders on $\mathcal{P}$ to which the lexicographic order belongs.

Definition 3 (Local Representability). *An order $\succeq := (\succeq_t)_{t\geq 1}$ on $\mathcal{P}$ is locally representable if, for every $t \geq 1$, there exist $\mathbf{u}_t \in \mathbb{R}^n$ such that*

$$\mu \succeq_s \nu \Longleftrightarrow \mu \cdot \mathbf{u}_t \;\geq\; \nu \cdot \mathbf{u}_t \qquad \forall\, \mu, \nu \in \mathcal{P}_s, \quad \forall s \leq t. \tag{8}$$

A sequence $(\mathbf{u}_t)_{t\geq 1}$ is said to locally represent $\succeq$ if (8) holds. The order $\succeq$ is said to be globally representable if there exist $\mathbf{u} \in \mathbb{R}^n$ such that (8) is satisfied for $\mathbf{u}_t = \mathbf{u}$ for all t.

The lexicographic order is locally representable but not globally.

Theorem 2. *An order $\succeq = (\succeq_t)_{t\geq 1}$ on $\mathcal{P}$ is consistent if and only if it is locally representable.*

Proof. If the order is locally representable it is straightforward to verify that it is consistent. Suppose the sequence of vectors $(\mathbf{u}_t)_{t\geq 1}$ represents $\succeq = (\succeq_t)_{t\geq 1}$. Let $\mu, \nu \in \mathcal{P}_s$ with $\mu \succeq_s \nu$ and $\eta \in \mathcal{P}_t$. Then $\mu \cdot \mathbf{u}_{s+t} \geq \nu \cdot \mathbf{u}_{s+t}$ since $\mathbf{u}_{s+t}$ can be used to compare multisets of cardinality t as $t < t + s$. But now

$$(\mu + \eta) \cdot \mathbf{u}_{s+t} - (\nu + \eta) \cdot \mathbf{u}_{s+t} = \mu \cdot \mathbf{u}_{s+t} - \nu \cdot \mathbf{u}_{s+t} \geq 0$$

which means $\mu + \eta \succeq_{s+t} \nu + \eta$.

To see the converse, let $\succeq = (\succeq_t)_{t\geq 1}$ be consistent. An immediate implication of consistency is that for any $\mu_1, \nu_1 \in \mathcal{P}_t$ and $\mu_2, \nu_2 \in \mathcal{P}_s$,

$$\mu_1 \succeq_t \nu_1 \text{ and } \mu_2 \succeq_s \nu_2 \Longrightarrow \;\mu_1 \cup \mu_2 \;\succeq_{t+s}\; \nu_1 \cup \nu_2, \tag{9}$$

where we have $\mu_1 \cup \mu_2 \;\succ_{t+s}\; \nu_1 \cup \nu_2$ if and only if either $\mu_1 \succ_t \nu_1$ or $\mu_2 \succ_s \nu_2$. Indeed by consistency, we have

$$\mu_1 \cup \mu_2 \;\succeq_{t+s}\; \nu_1 \cup \mu_2 \;\succeq_{t+s}\; \nu_1 \cup \nu_2.$$

Now suppose, by way of contradiction, that local representability fails at some t which means that $\mathbf{u}_t$ is the first vector that cannot be found. Note that there are $N = \binom{n+t-1}{t}$ multisets of cardinality t in total. Let us enumerate all the multisets in $\mathcal{P}_t$ so that

$$\mu_1 \succeq_t \mu_2 \succeq_t \cdots \succeq_t \mu_{N-1} \succeq_t \mu_N. \tag{10}$$

Some of these relations may be equivalencies, the others will be strict inequalities. Let $I = \{i \mid \mu_i \sim_t \mu_{i+1}\}$ and $J = \{j \mid \mu_j \succ_t \mu_{j+1}\}$. If $\succeq_t$ is complete indifference, i.e. all inequalities in (10) are equalities, then it is representable and can be obtained by assigning 1 to all of the utilities. Hence at least one ranking in (10) is strict or $J \neq \emptyset$.

The non-representability of $\succeq_t$ is equivalent to the assertion that the system of linear equalities $(\mu_i - \mu_{i+1}) \cdot \mathbf{x} = 0$, $i \in I$, and linear inequalities $(\mu_j - \mu_{j+1}) \cdot \mathbf{x} > 0$, $j \in J$, has no semi-positive solution.

A standard linear-algebraic argument tells us that inconsistency of the system above is equivalent to the existence of a nontrivial linear combination

$$\sum_{i=1}^{N-1} c_i(\mu_i - \mu_{i+1}) = 0 \tag{11}$$

with non-negative coefficients c_j for $j \in J$ of which at least one is non-zero (see, for example, Theorem 2.9 of [6], page 48). Coefficients c_i, for $i \in I$, can be replaced by their negatives since the equation $(\mu_i - \mu_{i+1}) \cdot \mathbf{x} = 0$ can be replaced with $(\mu_{i+1} - \mu_i) \cdot \mathbf{x} = 0$. Thus we may assume that all coefficients of (11) are non-negative with at least one positive coefficient c_j for $j \in J$. Since the coefficients of vectors $\mu_i - \mu_{i+1}$ are integers, we may choose $c_1, \ldots, c_n$ to be non-negative rational numbers and ultimately non-negative integers.

The equation (11) can be rewritten as

$$\sum_{i=1}^{N-1} c_i\mu_i = \sum_{i=1}^{N-1} c_i\mu_{i+1}, \tag{12}$$

which can be rewritten as the equality of two unions of multisets:

$$\bigcup_{i=1}^{N-1} \underbrace{\mu_i \cup \ldots \cup \mu_i}_{c_i} = \bigcup_{i=1}^{N-1} \underbrace{\mu_{i+1} \cup \ldots \cup \mu_{i+1}}_{c_i} \tag{13}$$

which contradicts to $c_j > 0$, $\mu_j \succ \mu_{j+1}$ and (9). This contradiction proves the theorem.

The above equivalence lies at the heart of proof Theorem 1. Indeed, it already implies, via Claims 1-3 given in the previous section, that Axioms 1-3 imply the existence of an ex-post representation for $\mathcal{D}$. What remains to be shown is the characterization of all such representations.

Consistent orders on $\mathcal{P}_t$ can be represented geometrically [14]. Every point $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ defines an order $\succeq_{\mathbf{u}}$ on $\mathcal{P}_t$, which obtains when we allocate utilities $u_1, \ldots, u_n$ to prizes $i = 1, 2, \ldots, n$, that is

$$\mu \succeq_{\mathbf{u}} \nu \Longleftrightarrow \sum_{i=1}^{n} \mu(i)u_i \geq \sum_{i=1}^{n} \nu(i)u_i. \tag{14}$$

Any order on $\mathcal{P}_t$ that can be expressed as $\succeq_{\mathbf{u}}$ for some $\mathbf{u} \in \mathbb{R}^n$ is said to be *representable*. We will now argue that the representable linear orders on $\mathcal{P}_t$ are in one-to-one correspondence with the regions of the following hyperplane arrangment.

For any pair of multisets $\mu, \nu \in \mathcal{P}_t[n]$, we define the hyperplane

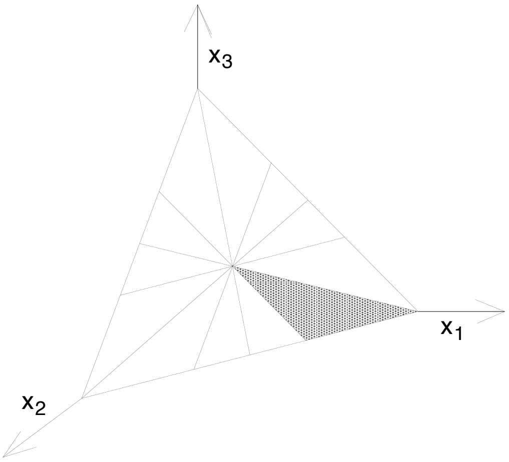
$$L(\mu, \nu) \;=\; \left\{ \mathbf{x} \in \mathbb{R}^n \mid \sum_{i=1}^{n} \mu(i)x_i - \sum_{i=1}^{n} \nu(i)x_i = 0 \right\}$$

and consider the hyperplane arrangement

$$A(t,n) = \big\{L(\mu,\nu) \ \mid \ \mu,\nu \in \mathcal{P}_t[n]\big\}. \tag{15}$$

The set of representable linear orders on $\mathcal{P}_t[n]$ is in one-to-one correspondence with the regions of $A = A(t,n)$. In fact, then the linear orders $\succeq_{\mathbf{u}}$ and $\succeq_{\mathbf{v}}$ on $\mathcal{P}_t$ will coincide if and only if $\mathbf{u}$ and $\mathbf{v}$ are in the same region of the hyperplane arrangement A. This immediately follows from the fact that the order $\mu \succ_{\mathbf{x}} \nu$ changes to $\mu \prec_{\mathbf{x}} \nu$ (or the other way around) when $\mathbf{x}$ crosses the hyperplane $L(\mu,\nu)$. The closure of every such region is a convex polytope.

Example 2. The 12 regions on the figure below represent all 12 representable orders on $\mathcal{P}_2[3]$.



with the shaded region corresponding to the lexicographic order $1^2 \succ 12 \succ 13 \succ 2^2 \succ 23 \succ 3^2$.

Let us note that in (14) we can divide all utilities by $u_1 + \ldots + u_n$ and the inequality will still hold. Hence we could from the very beginning consider that all vectors of utilities are in the hyperplane J given by $x_1 + \ldots + x_n = 1$ and even in the simplex Δ given by $x_i \geq 0$ for $i = 1,2,\ldots,n$.

Thus, every representable linear order on $\mathcal{P}_t$ is associated with one of the regions of the induced hyperplane arrangement $A^J = \{L \cap J \mid L \in A\}$.

Let us note that due to our non-triviality assumption the vector $\left(\frac{1}{n},\ldots,\frac{1}{n}\right)$ does not correspond to any order. Consider a utility vector $\mathbf{u} \in \Delta$ different from $\left(\frac{1}{n},\ldots,\frac{1}{n}\right)$ lying in one of the regions of A^J whose closure is V. We then can normalise $\mathbf{u}$ applying a positive affine linear transformation which makes its lowest utility zero. Indeed, suppose that without loss of generality $u_1 \geq u_2 \geq \ldots \geq u_n \neq \frac{1}{n}$. Then we can solve for α and β the system of linear equations $\alpha + n\beta = 1$ and $\alpha u_n + \beta = 0$ and since the determinant of this system is $1 - nu_n \neq 0$ its solution is unique. Then the vector of utilities $\mathbf{u}' = \alpha\mathbf{u} + \beta \cdot \mathbf{1}$ will lie on the facet Δ^n of Δ and we will have $\succeq_{\mathbf{u}'} = \succeq_{\mathbf{u}}$. Hence the polytope V

has one face on the boundary of Δ. We denote it U. So if the order $\succeq$ on $\mathcal{P}_t$ is linear the dimension of U will be $n-2$.

In general, when the order on $\mathcal{P}_t$ is not linear, the utility vector $\mathbf{u}$ that represents this order must be a solution to the finite system of equations and strict inequalities:

$$\begin{array}{ll}(\mu-\nu)\cdot\mathbf{u}=0 & \text{whenever } \mu\sim_{\mathbf{u}}\nu,\\ (\mu-\nu)\cdot\mathbf{u}>0 & \text{whenever } \mu\succ_{\mathbf{u}}\nu,\end{array}\qquad \forall\,\mu,\nu\in\mathcal{P}_t. \tag{16}$$

Then $\mathbf{u}$ will lie in one (or several) of the hyperplanes of $A(k,n)$. In that hyperplane an arrangement of hyperplanes of smaller dimension will be induced by $A(k,n)$ and $\mathbf{u}$ will belong to a relative interior of a polytope U of dimension smaller than $n-2$.

Let now $\succeq=(\succeq_t)_{t\ge 1}$ be a consistent order on $\mathcal{P}$. By Theorem 2 it is locally representable. We have just seen that in such case, for any t, there is a convex polytope U_t such that any vector $\mathbf{u}_t\in ri(U_t)$ represents $\succeq_t$. Due to consistency any vector $\mathbf{u}_s\in ri(U_s)$, for $s>t$ will also represent $\succeq_t$ so $U_t\supseteq U_s$. Thus we see that our polytopes are nested. Note that only points in the relative interior of U_t are suitable points of utilities to rationalise $\succeq_t$. We have almost proved our main theorem. The only thing which is left to note is that the intersection $\bigcap_{t=1}^{\infty}U_t$ has exactly one element. This is immediately implied by the following

Proposition 1. *Let $\mathbf{u}\ne\mathbf{v}$ be two distinct vectors of normalised non-negative utilities. Then there exist a positive integer t and two multisets $\mu,\nu\in\mathcal{P}_t$ such that $(\mu-\nu)\cdot\mathbf{u}>0$ but $(\mu-\nu)\cdot\mathbf{v}<0$.*

Proof. Since $\mathbf{u}$ and $\mathbf{v}$ are normalised we have, in particular, $u_n=v_n=0$. Since $\mathbf{u}\ne\mathbf{v}$, there will be a point $\mathbf{x}=(x_1,\ldots,x_n)\in\mathbb{R}^n$ such that $\mathbf{x}\cdot\mathbf{u}>0$ but $\mathbf{x}\cdot\mathbf{v}<0$. As rational points are everywhere dense in $\mathbb{R}^n$ we may assume that $\mathbf{x}$ has rational coordinates. Then multiplying by their common denominator we may assume all coefficients are integers. After that we may change the last coordinate x_n of $\mathbf{x}$ to x'_n so that to achieve $x_1+x_2+\ldots+x'_n=0$. Now since $u_n=v_n=0$, we will still have $\mathbf{x}'\cdot\mathbf{u}>0$ and $\mathbf{x}'\cdot\mathbf{v}<0$ for $\mathbf{x}'=(x_1,x_2,\ldots,x'_n)$. Now $\mathbf{x}'$ is uniquely represented as $\mathbf{x}'=\mu-\nu$ for two multisets μ and ν. Since the sum of coefficients of $\mathbf{x}'$ was zero, the cardinality of μ will be equal to the cardinality of ν. Let this common cardinality be t. Then $\mu,\nu\in\mathcal{P}_t$ and they are separated by a hyperplane from $A(t,n)$. The proposition is proved.

Appendix

Proof of Claim 1. Take the hypothesis as given. If the actions $a,b,c,d\in\mathcal{A}$ are distinct, consider a history $g_t\in H_t$ such that $g_t(a)=h_t(a)$, $g_t(b)=h_t(b)$, $g_t(c)=h'_t(a)$ and $g_t(d)=h'_t(b)$. Applying Axiom 2, $a\sim_{g_t}c$ and $b\sim_{g_t}d$ and therefore, $a\succeq_{g_t}b\Leftrightarrow c\succeq_{g_t}d$. Apply Axiom 1 to complete the claim.

Suppose now that a,b,c,d are not all distinct. We will prove that if $\mu(a,h)=\mu(c,h')$ and $\mu(b,h)=\mu(b,h')$, then

$$a\succeq_{h_t}b\Longleftrightarrow c\succeq_{h'_t}b,$$

which is the main case. Let us consider five histories presented in the following table:

	h	h^1	h^2	h^3	h'
a	$h(a)$	$h(a)$	$h'(b)$	$h'(b)$	$h'(a)$
b	$h(b)$	$h(b)$	$h(b)$	$h'(b)$	$h'(b)$
c	$h(c)$	$h'(c)$	$h'(c)$	$h'(c)$	$h'(c)$

In what follows we repeatedly use Axiom 1 and Axiom 2 and transitivity of $\succeq_{h^i}$, $i = 1, 2, 3$. Comparing the first two histories, we deduce that $c \sim_{h^1} a \succeq_{h^1} b$ and $c \succeq_{h^1} b$. Now comparing h^1 and h^2 we have $c \succeq_{h^2} b \sim_{h^2} a$ and $c \succeq_{h^2} a$. Next, we compare h^2 and h^3 and it follows that $c \succeq_{h^3} a \sim_{h^2} b$, whence $c \succeq_{h^3} b$. Now comparing the last two histories we obtain $c \succeq_{h'} b$, as required.

Proof of Claim 2. Given the fact that actions must be ranked for all conceivable histories, $\succeq_t^*$ is a complete ordering of $\mathcal{P}_t$. From its construction, $\succeq_t^*$ is also is reflexive. Again, through appealing to Axiom 1 and Axiom 2 repeatedly, it may be verified that it is also transitive. Indeed, choose $\mu, \nu, \xi \in \mathcal{P}_t$ such that $\mu \succeq_t^* \nu$ and $\nu \succeq_t^* \xi$. Pick three distinct actions $a, b, c \in \mathcal{A}$ and consider a history $h_t \in H_t$ such that $\mu(a, h_t) = \mu$, $\mu(b, h_t) = \nu$ and $\mu(c, h_t) = \xi$. By definition, $a \succeq_{h_t} b$ and $b \succeq_{h_t} c$ while transitivity of $\succeq_{h_t}$ shows that $a \succeq_{h_t} c$. Hence $\mu \succeq_t^* \xi$.

References

1. Anscombe, F., Aumann, R.: A definition of subjective probabiliy. Annals of Mathematical Statistics 34:199–205, (1963)
2. Börgers, T., Morales, A.J., Sarin, R. Expedient and monotone learning rules. Econometrica 72(2):383–405 (2004)
3. Brown, G.W.: Iterative solutions of games by fictitious play. In: Koopmans, T.C. (ed.) Activity Analysis of Production and Allocation, 374-376. New York: Wiley (1951).
4. Easley, D., Rustichini. A.: Choice without beliefs. Econometrica 67(5):1157–1184 (1999)
5. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games, MIT Press (1998)
6. Gale, D.: The Theory of Linear Economic Models. McGraw-Hill, New-York (1960)
7. Gigerenzer, G., Selten, R.: Bounded Rationality: The Adaptive Toolbox. MIT Press (2002)
8. Gilboa, I., Schmeidler, D.: A theory of case-based decisions. Cambridge University Press (2001)
9. Gilboa, I., Schmeidler, D.: Inductive Inference: An Axiomatic Approach. Econometrica 71(1):1 – 26 (2003)
10. Lettau, M., Uhlig, H.: Rules of thumb and dynamic programming. Tilburg University Discussion Paper (1995)
11. Milnor, J.H.: Games against nature. In: Davis, C.H, Coombs, R.M., R.L. Thrall, R.L (eds), Decision Processes, 49–60. John Wiley & Sons, Inc., (1954)
12. Savage, L.J.: The Foundations of Statistics. Harvard University Press, Cambridge, Mass. (1954)
13. Schlag, K.H.: Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. Journal of Economic Theory 78(1):130–156 (1998)
14. Sertel, M. R., Slinko, A.: Ranking Committees, Words or Multisets, Economic Theory, 30(2): 265–287 (2007)