On choice in a complex environment

Murali Agastya *
m.agastya@econ.usyd.edu.au

Arkadii Slinko †
slinko@math.auckland.ac.nz

April 20, 2009

Abstract

A Decision Maker (DM) must choose at discrete moments from a finite set of actions that result in random rewards. The environment is complex in that she finds it impossible to describe the states and is thus prevented from application of standard Bayesian methods. This paper presents an axiomatic theory of choice in such environments.

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 observed rewards. 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 are necessary and sufficient for the DM's preferences to admit a utility representation. The DM's behavior in this case will be akin to fictitious play. We then show that, if rewards are drawn by a stationary stochastic process, the observed behavior of such a DM almost surely cannot be distinguished from anyone who is fully cognizant of the environment.

Keywords: Ex-post utility maximization, Choice under ignorance, Multisets, Fictitious play.

*Murali Agastya, Economics Discipline, H04 Merewether Building University of Sydney NSW 2006 AUSTRALIA

†A.M.Slinko, Department of Mathematics, University of Auckland, Private Bag 92019, Auckland NEW ZEALAND

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 Savage (1954) or Anscombe and Aumann (1963), would then be inapplicable.[1] Yet, decision makers often find themselves in these situations and do somehow make choices, the complexity of the environment notwithstanding. This paper offers a theory of choice in such environments.

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 $\mathcal{D} := (\succeq_{h_t})_{h_t \in H}$, where $\succeq_{h_t}$ 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. We impose axioms on $\mathcal{D}$.

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 using a rule of thumb. To use a term coined by Reinhard Selten, the DM indulges in *ex-post rational* behavior.[2] In this literature, the "stimulus" is almost always modeled as a real number which is interpreted as a monetary payoff or an exogenously specified cardinal utility assignment to a reward. The rule that maps these past payoffs to actions is either fully specified or is assumed to have some desirable properties. The focus is the 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 Börgers, Morales, and Sarin (2004), Schlag (1998), Gigerenzer and Selten

[1] Knight (1921) and Ellsberg (1961) concern the existence of a prior. More recent and a more direct questioning of the assumption that a DM may have a well defined state space (let alone a known prior) have lead to Gilboa and Schmeidler (1995), Easley and Rustichini (1999), Dekel, Lipman, and Rustichini (2001), Gilboa and Schmeidler (2003) and Karni (2006) among others. Gilboa and Schmeidler (1995), in particular, forcefully argues how in many environments there is no naturally given state space and how the language of expected utility theory precludes its application in these cases.

[2] See Selten (2001) and the informal discussion available at http://www.strategy-business.com/press/16635507/05209.

(2001), Fudenberg and Levine (1998) and the references therein.

An exception to the above is Easley and Rustichini (1999) (hereafter ER). Instead of directly assuming rules that map payoffs to actions, ER, like us, consider an abstract individual decision problem modeled as a family of preference relations, not on actions but on lotteries over them, and impose axioms on it. The use of the axiomatic approach to dynamic choice under ignorance makes ER the closest relative of this paper. We defer a complete discussion of the relation of this work to ER (and other literature) to Section 7. We do note here however that there are significant differences both in the formal modeling details and in the conceptual basis for the axioms. For instance, our formulation allows for considerable *path dependence* of the DM's preferences over actions across time which is ruled out in ER. Our results will also show that the DM can be initially ambiguous on how to value the rewards but becomes increasingly precise over time. This feature too is absent in ER (since they also assume that rewards/payoffs are monetary). In fact, the class of adaptive learning procedures that are axiomatized here resemble fictitious play in contrast to the replicator dynamics characterized in their work.[3]

What we do share with ER and many of the works cited above is that the DM operates in a social environment in which there are other decision makers. For, we assume that at each date the DM is able to observe rewards that occur to each of the actions, including those that she herself did not choose. Such an assumption on observability of rewards seems particularly natural for situations such as betting on horses or investing on a share-market. For, in these cases there is a sufficient diversity of preferences so that all the actions are chosen in each period by various individuals and outcomes are publicly observable.

There are three results in the main — Theorem 1 is a "utility representation" result for $\mathcal{D}$. Theorem 2 uses the previous result to show that the observed behavior of a DM who obeys our axioms is virtually indistinguishable from a fully rational DM provided the rewards are generated by a stationary stochastic process. Proposition 3 is a simple empirical test for refuting the axioms. The novelty in the proof of Theorem 1 is the identification of a certain isomorphism between preferences over actions across time and a binary relation over *multisets* of rewards. We then prove Proposition 2, a representation result for orderings of *multisets* — a technical result which we expect to be of independent interest with applications elsewhere in Decision Theory and Social Choice. Theorem 1 is then deduced from Proposition 2. We shall now elaborate on the axioms and these results.

[3]Fictitious play was introduced by Brown (1951). See Fudenberg and Levine (1998) for variations of fictitious play. See Hopkins (2002) for a nice comparison of fictitious play and replicator dynamics.

There are three axioms. The first axiom requires that a comparison of a pair of actions at any date depends on the historical sequence of observed rewards corresponding to only that pair. The second axiom captures the bounded rationality of the DM. It insists that for any sequence of rewards attributed to an action in any history, the DM is able to track only the *number of times* various rewards have accrued. The final axiom concerns the updating of preferences in response to the rewards and is loosely based on the principle of insufficient reason: if a pair of actions receive the same reward in the current period following a history h_t, then their current relative ranking is carried forward to the next period.

One way of ranking actions that would satisfy the above axioms would be for the DM to assign utility weights to the underlying set of rewards and, just as in fictitious play, the utility of an action at any date is the *average* utility of the rewards that have occured to the action until then. Our first result, Theorem 1 in Section 2, shows the above axioms are equivalent to precisely this procedure with the following caveat. The set of endogenously determined utility weights for rewards that are available to the DM at any date are not necessarily unique (even after applying positive affine transformations). Rather, the DM can choose the utility weights for the rewards from a certain convex polytope U_t in $\mathbb{R}^n$ for each date t such that $U_{t+1} \subseteq U_t$. It is worth noting that the non-uniqueness in the valuation of rewards coexists with the DM's preferences over actions being complete and transitive at every date.

We refer to the above axiomatized procedure as *ex-post rationality*. Thus, in a nutshell, Theorem 1 shows that our axioms are equivalent to the agent choosing between the empirical distributions of the rewards to different actions as if she is an expected utility maximiser. The fact that $U_{t+1} \subseteq U_t$ means an ex-post rational DM learns more about her imputed utilities for the rewards over time.

The *simultaneous* determination of both the value of rewards and the ranking of actions over time given by Theorem 1 sets our work apart from the existing literature on dynamic learning procedures cited previously. Moreover the evolution of utility weights permitted by our result shows that our framework allows for classes of behavior that are not usually captured in the above literature. For instance,[4] in evaluating a pair of treatments (actions), suppose a doctor finds the first action has resulted in much better outcomes in the

[4]This example is related to one given in Gilboa and Schmeidler (2003) who attribute it to Peyton Young when discussing scenarios where their Combination Axiom *fails*. Their Combination Axiom is related to Axiom 3 here.

past but most recently has resulted in a fatality. The second action has no such record. Then, in our framework, even with a single such outcome, it is consistent for the doctor to strictly prefer the second action. That is, in our framework, it is possible that some rewards are implicitly considered to be "infinitely" more relevant than others.

The intersection of all polytopes of the sequence $(\mathbf{U}_t)_{t\geq 1}$ is always a singleton. However, it does not necessarily constitute a valid assignment of utility weights to the rewards. In the event it does, just this one vector of utility weights may be used to describe the DM's behavior in all time periods. In this case, the DM is said to admit a "global utility representation". An ex-post rational DM with a global utility representation would simply be engaging in fictitious play. In Theorem 2 (see Section 6.2), we show that in stationary stochastic environments, an external observer will typically be unable to distinguish between an ex-post DM and a rational DM that is fully cognizant of the environment and maximizes expected utility.

Despite Theorem 2, it is important to realize that all our axioms are imposed on behavior following observed data and are hence refutable. In Proposition 3 (see Section 6.3), we present a simple condition for checking whether a DM is consistent with our axioms. The condition involves simply checking whether a certain finite system of linear inequalities admits a solution.

The rest of the paper is organised as follows. Section 2 introduces the basic setup. The axioms are formally presented and discussed in Section 3. In Section 4 we formally introduce "ex-post utility representation" and "ex-post rational behavior" and study their properties. The representation results are in Section 5. Proposition 2, the representation result for multisets that may be of independent interest occurs in Section 5.1. Section 6 discusses various aspects of the paper.

Relation of our work to the literature is in Section 7. Besides the connections to the literature on learning procedures, the nature of the representation result for $\mathcal{D}$ is bound to invite a comparison with *Case Based Decision Theory* developed by Itzhak Gilboa and David Schmeidler. We refer the reader to their book Gilboa and Schmeidler (2001) for an overview of their various contributions to this theory. The relation of our model to their theory is also given in Section 7. Section 8 concludes.

2 The Model

A Decision Maker must choose from a finite set $\mathcal{A} = \{a_1, \ldots, a_m\}$ of m actions 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. Let $r_i^{(t)}$ denote the reward to an action a_i at moment t and $\mathbf{r}_t = (r_1^{(t)}, \ldots, r_m^{(t)})$ the vector of rewards to the various actions. A *history* at date t is a sequence of vectors of rewards $h_t = (\mathbf{r}_0, \ldots, \mathbf{r}_{t-1})$.

The DM makes no attempt to learn the characteristics of the underlying data generating process. Rather, she relies on precedent to determine her preferences over actions. That is, upon observing a h_t at date t, the DM works out a preference relation[5] $\succeq_{h_t}$ 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 $h_{t+1} = (h_t, \mathbf{r}_t)$. We will soon impose a set of axioms that govern these preferences.

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 $\mathcal{D} := (\succeq_h)_{h \in H}$ completely describes the DM. Our objective is to discuss the behavior of this learning agent through the imposition of certain axioms that encapsulate her procedural rationality. For a DM satisfying these axioms we will derive a utility representation theorem that is based on the empirical distribution of rewards in the history.

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

First, as in Easley and Rustichini (1999), 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 we require a preference on actions to be specified after every conceivable history. 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. The assumption is

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

also much in the spirit of the theoretical developments in virtually all decision theory. For instance, in Savage (1954), a ranking of all conceivable acts is required. (See Aumann and Dreze (2008) or Blume, Easley, and Halpern (2006) however.) The presumption underlying such a requirement 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.

Non-trivial $\mathcal{D}$. We maintain a non-triviality assumption on $\mathcal{D}$ for the rest of this paper. That is, we assume that there exists some one-period history $h_1 \in H_1$ and a pair of actions a, b such that $a \succ_{h_1} b$. It is worth emphasizing that this does not entail any loss in generality. Indeed, should this not be the case, the implication in conjunction with our axioms will be that the DM is indifferent between all actions following all histories making any analysis redundant.

2.1 Multisets

For the axioms of dynamic choice and the thumb rule for choice that will ultimately be characterized, the number of times different rewards accrue to a given action during a history is important. To describe this, it is convenient to use *multisets*. We remind the reader that 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, a typical multiset is 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 $\sum_{i=1}^{n} \mu(i)$. Let $\mathcal{P}_t$ denote the subset of all such multisets of cardinality t whereupon

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

denotes the set of all non-empty multisets over $\mathcal{R}$. We will write $\mathcal{P}_t[n]$ or $\mathcal{P}[n]$ when we need to emphasize the number of available rewards. 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 $i \in \mathcal{R}$. In other words, $\mu \cup \nu = \mu + \nu$, the usual sum of two vectors (of integers). 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 corresponding to action a and $\mu(a, h) = (\mu_1(a, h), \ldots, \mu_n(a, h))$. We will refer to $\mu(a, h)$ as the *multiset of prizes* corresponding to the pair (a, h).

Example 1. Suppose that for $t = 9$ and $n = 5$ the history of rewards for action a is

$$h(a) = (1, 1, 3, 5, 2, 5, 2, 2, 2), \quad \text{then} \quad \mu(a, h) = (2, 4, 1, 0, 2).$$

An alternative self-explanatory notation for this multiset that is often used in mathematics is $\mu(a, h) = \{1^2, 2^4, 3, 5^2\}$.

At various places we will also need to consider preference relations on $\mathcal{P}_t$. We will use $\succeq_t$ to denote a relation on $\mathcal{P}_t$. We alert the reader that this should not be confused with $\succeq_h$ which is a preference relation on $\mathcal{A}$ the set of actions.

3 Axioms

We impose three axioms on the DM's behavior. The first axiom says that in comparing a pair of actions, the information regarding the other actions is irrelevant. Formally, given a history $h_t \in H_t$ and an action $a \in \mathcal{A}$, let $h_t(a)$ be the sequence of rewards corresponding to this action.

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 $a \succeq_{h'_t} b$.*

The next axiom aims to capture the bounded rationality of the agent. Although the agent has the entire history at her disposal, we postulate that for any action, she can only track the number of times different rewards were realised. Thus, if the empirical distribution of rewards corresponding to the two actions a and b is the same in a history h_t, then the DM is indifferent between them.

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

Our final axiom describes how the DM revises 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, an implication of Axiom 3 is 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,

then DM continues to prefer a to b. We view Axiom 3 as loosely capturing the "principle of insufficient reason at the margin". In principle it allows for a wide range of behavior. For instance it allows for the fact that some rewards are infinitely more "important" than others. For instance, after any history, ranking actions by lexicographically ordering their corresponding multisets of prizes is entirely consistent with this axiom.

We view the axioms as mostly plausible hypotheses of behavior under ignorance. However, it is worth noting that Axiom 1 is reminiscent of the Independence of Irrelevant Alternatives and could be subjected to a similar sort of criticism. Axiom 3 also rules out certain kinds of behavior that may be considered intuitive on some grounds. For instance, consider a situation where an action a has resulted in a "high" or a "low" reward an even number of times while b has resulted in a "medium" reward in every period over a long horizon t. It is conceivable that the DM prefers b to a for the security it offers at date t. Now suppose that at $t+1$ both a and b yield the low reward. One might argue that the DM's belief that b never delivers a low reward is shaken whereupon she revises her preference away from b.

It is worth pausing to compare the above axioms with those in ER. In their work, much of the focus is on the transition of preferences over actions from date t to date $t+1$, i.e. the more serious axiomatic treatment in their work concerns assumptions in the spirit of Axiom 3 above. It is therefore not possible to find direct counterparts of Axiom 1 and Axiom 2 in their work. Nonetheless, their Assumption 5.4 (PC-Pairwise Comparisons), namely that the "new measure of relative preference between action a and b is independent of the payoffs to the other actions" is precisely in the spirit of Axiom 1. Likewise, their Assumption 6.2 (E-Exchangeability) which "requires that the time order in which states are observed is unimportant" corresponds to Axiom 2.

We do not assume that rewards are monetary but if one does so, Axiom 3 would then be weaker than their Monotonicity assumption on the transition of preferences. But we emphasize that the key difference is that here Axiom 3 allows for considerable *path dependence* in the revision of preferences. In other words, it is entirely possible in our framework that there can be a pair of t period histories h_t, h'_t such that $\succeq_{h_t} = \succeq_{h'_t}$ and yet when followed by the same reward vector at $h_{t+1} = (h_t, \mathbf{r}_t)$ and $h'_{t+1} = (h'_t, \mathbf{r}_t)$ we have $\succeq_{h_{t+1}} \neq \succeq_{h'_{t+1}}$. In their setting, $\succeq_{h_t} = \succeq_{h'_t}$ implies $\succeq_{h_{t+1}} = \succeq_{h'_{t+1}}$ for all $\mathbf{r}_t$.

4 Ex-post Utility

Our axioms will ultimately characterize a thumb rule for dynamic choice that entails utility maximization in a certain ex-post sense. Our aim in this section is to offer an independent motivation for this procedure and study some of its properties.

The rule that underlies what we will presently define to be "ex-post rational" behavior is to closely related to fictitious play, a widely studied learning procedure in games. (See the references given in Footnote 3.) As under fictitious play, at any moment the DM looks at the empirical distribution of rewards obtained to a given action in the past. She then ranks these empirical distributions by assigning utility weights $\mathbf{u} = (u_1, \ldots, u_n)$ to the underlying rewards and taking the expected values of the empirical distributions. Unlike in the usual fictitious play, there is a *set* of these weights which may be revised at each point in time. Ex-post utility maximization places some restrictions on how these weights are revised. The definition below makes this precise.

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

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* $\mathcal{D} = (\succeq_h)_{h \in H}$ *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* $\mathbf{u} \in \mathbb{R}^n_+$*.*

A plausible rationale for the DM to engage in the above behavior is as follows. Recall that the DM she is ignorant of the probabilities. In the absence of any knowledge about the environment, a 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. There is a good reason to allow the DM to use different vectors of utilities at different moments. This will allow her to refine her utility weights, at each moment, from the previous period to reflect her preferences over longer histories.[6] Therefore, in an ex-post representation, not

[6] Allowing for the utility weights to vary over time also has the advantage of accommodating $\mathcal{D}$ that have lexicographic properties. (See Section 6.1.)

only the vector $\mathbf{u}_t$ but any $\mathbf{u}_{t+k}$ with $k \geq 1$ may also be used to represent $\succeq_{h_t}$.

Definition 2 (Ex-post rational)**.** *The DM is said to be ex-post rational if $\mathcal{D}$ admits an ex-post utility representation.*

We emphasise that the object that is of ultimate interest is the ranking of the actions following histories, namely $\mathcal{D}$. Clearly, the same $\mathcal{D}$ can admit several ex-post utility representations. Indeed, should $(\mathbf{u}_t)_{t\geq 1}$ be an ex-post utility representation of some $\mathcal{D}$, then any sequence $(\mathbf{u}'_t)_{t\geq 1}$ obtained by applying some positive affine transformations $\mathbf{u}'_t \mapsto \alpha_t \mathbf{u}_t + \beta_t$ (with $\alpha_t > 0$) is also an ex-post utility representation. The next step is therefore to offer a succinct characterization of all the ex-post utility representations of a given $\mathcal{D}$.

It is clear from above that, with no loss in generality, we may begin by assuming that every $\mathbf{u}_t$ in an ex-post utility representation $(\mathbf{u}_t)_{t\geq 1}$ lies in $\Delta \subseteq \mathbb{R}^n$, the $n-1$ 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 coordinates are equal. Hence we may assume that at any $\mathbf{u}_t = (u_1, \ldots, u_n)$ in a representation, $\min\{u_i\} = 0$ (and $\max\{u_i\} > 0$). Hence, $\mathbf{u}_t$ may in fact be assumed to lie in the following subset of the unit simplex:

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

which is one of the facets[7] of Δ.

Next, note that by arbitrarily choosing $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ as the utility weights, we obtain an order $\succeq_{\mathbf{u}}$ on $\mathcal{P}_t$,[8] whereby for any two multisets $\mu, \nu \in \mathcal{P}_t$,

$$\mu \succeq_{\mathbf{u}} \nu \iff \sum_{i=1}^{n} \mu(i) u_i \geq \sum_{i=1}^{n} \nu(i) u_i. \tag{4}$$

Definition 3 (Representable ordering of $\mathcal{P}_t$)**.** *A preference relation $\succeq_t$ on $\mathcal{P}_t$ is said to be* representable *if there exists some $\mathbf{u} \in \mathbb{R}^n$ such that $\succeq_t = \succeq_{\mathbf{u}}$.*

The interest in representable orders over $\mathcal{P}_t$ for any t should be clear since any ex-post utility representation of $\mathcal{D}$ induces a representable order, namely $\succeq_{\mathbf{u}_t}$, on $\mathcal{P}_t$. The following Lemma is a key step for obtaining all equivalent ex-post utility representations of $\mathcal{D}$.

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

[8] There is a slight abuse of notation here – $\succeq_{\mathbf{u}}$ being an ordering on $\mathcal{P}_t$ must depend on t. The value of t will be clear from the context.

Let $ri(C)$ denote the relative interior of a convex set C.

Lemma 1. *The set of distinct utility representations of a representable order on $\mathcal{P}_t$ are positive affine transformations of some element $\mathbf{u} \in ri(U_t)$ for a unique convex polytope $U_t \subseteq \Delta^i$ for some i.*

Using this Lemma, we can now give a complete description of all distinct ex-post utility representations of an ex-post rational DM.

Proposition 1. *Suppose the DM with preferences $\mathcal{D}$ is ex-post rational. There is a unique sequence of non-empty convex polytopes $(U_t)_{t\geq 0}$ such that*

1. *$U_t \subseteq \Delta^i$ for all $t \geq 0$, for some i,*
2. *$U_{t+1} \subseteq U_t$ for all $t \geq 1$.*
3. *a sequence $(\mathbf{u}_t)_{t\geq 1}$ of vectors of $\mathbb{R}^+_n$ is an ex-post utility representation of $\mathcal{D}$ if and only if $\mathbf{u}_t$ is a positive affine transformation of some $\mathbf{u}'_t \in ri(U_t)$.*
4. *$\bigcap_{t=1}^{\infty} U_t$ consists of a single vector which is a global utility representation if $\bigcap_{t=1}^{\infty} U_t$ is in the interior of every U_t.*

Remark 1. It is worth drawing attention in particular to the fact that $\bigcap_{t=1}^{\infty} U_t$ is a singleton which is to say that a global utility representation, if it exists, must be unique. We refer the reader to Section 6.1 for a further discussion of this issue. Some readers may also find that Example 2 given there is a useful illustration of the above proposition.

Proof of Proposition 1. Every ex-post utility representation $(\mathbf{u}_t)_{t\geq 1}$ describes a representable order $\succeq_{\mathbf{u}_t}$ on $\mathcal{P}_t$. Lemma 1 then gives us Part 3. Moreover, observe from Definition 1 that in an ex-post representation, $\mathbf{u}_t$ and $\mathbf{u}_{t+1}$ induce the same representable order on $\mathcal{P}_t$. This gives $U_{t+1} \subseteq U_t$, i.e. Part 2. In our earlier discussion we have already argued that one may normalize so that $\mathbf{u}_t \in \Delta^i$ for some i. The fact that $U_{t+1} \subseteq U_t$ for all t also gives us Part 1.

To prove Part 4 suppose, by way of contradiction, that $\bigcap_{t=1}^{\infty} U_t$ has more than one element and without loss of generality, set $i = n$ in Part 1. Then there exist $\mathbf{u}, \mathbf{v} \in ri(U_t)$ for all t such that $\mathbf{u} \neq \mathbf{v}$. Since $\mathbf{u} \neq \mathbf{v}$, 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 $\mathbf{x} \cdot \mathbf{v} < 0$. These being strict inequalities, we may assume that $\mathbf{x}$ has rational coordinates and multiplying by their common denominator we may assume that

the coordinates are in fact integers. After that we may change the ith coordinate x_i of $\mathbf{x}$ to x'_n so as 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, by the above inequalities, we have $\mu \succ_{\mathbf{u}} \nu$ but $\mu \prec_{\mathbf{v}} \nu$, which is to say $\mathbf{u}$ and $\mathbf{v}$ describe different representable orders on $\mathcal{P}_t$, in contradiction of the fact that $\mathbf{u}$ and $\mathbf{v}$, being in $ri(U_t)$, must describe the same representable order on $\mathcal{P}_t$. Since only points in the relative interior of any U_t are valid utility representations, $\cap_{t \geq 1} U_t$ cannot be a global utility representation unless it lies in $ri(U_t)$ for all t.

□

5 Representation Results

In the previous sections, we have given a set of axioms that describe the behavior of the DM and discussed a class of preferences of the DM that we termed ex-post rational. We will now show the following:

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

1. *$\mathcal{D} = (\succeq_h)_{h \in H}$ satisfies Axioms 1– 3.*
2. *$\mathcal{D}$ is ex-post rational.*

Remark 2. Taken together with Proposition 1, the above Theorem shows that a $\mathcal{D}$ that satisfies Axioms 1-3 is uniquely identified by a non-increasing sequence of convex polytopes whose relative interiors determine the ex-post utility representations.

Remark 3. It is worth noting that Theorem 1 obtains despite the fact that at each date there are only finitely many rewards — there are no topological assumptions nor do we rely on the possibility of mixed strategies. The strategy of proof for showing the non-trivial part of Theorem 1, namely that $1 \Rightarrow 2$, is as follows. We first show that under Axioms 1-3, $\mathcal{D}$ is equivalent to a partial order over all multisets $\mathcal{P}$ that satisfies certain properties. Ex-post representability of $\mathcal{D}$ is then easily seen to be equivalent to that ordering in $\mathcal{P}$ admitting a certain utility representation. We shall therefore prove this latter representation in Section 5.1 and then we will give the proof of Theorem 1 in

Section 5.2. Besides being important for the proof of Theorem 1, we expect the material presented in Section 5.1 to be of independent interest with applications elsewhere.

Example 2. 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 $m = 2$. Pick any utility vector $\mathbf{u} = (u_1, \ldots, u_n)$ for the rewards and define $\mathcal{D}$ 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.

5.1 A representation result for orders on multisets

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 different cardinalities (it would if we had missing observations). Due to this, our main focus 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 naturally induced partial (but reflexive and transitive) binary relation on $\mathcal{P}$ whereby for any $\mu, \nu \in \mathcal{P}$, $\mu \succeq \nu$ if both μ and ν are of the same cardinality, say t, and $\mu \succeq_t \nu$ and $\succeq$ is undefined otherwise.[9]

A typical $\succeq$ involves a comparison of only multisets of equal cardinality. Let us consider the following condition that relates orders of different cardinalities.

Definition 4 (Consistency)**.** *An order $\succeq = (\succeq_t)_{t\geq 1}$ on $\mathcal{P}$ is said to be* consistent *if for any $\mu, \nu \in \mathcal{P}_t$ and any $\xi \in \mathcal{P}_s$,*

[9]Mathematically speaking $\mathcal{P}$ here is considered as an object *graded* by positive integers. In a graded object all operations and relations are defined on its homogeneous components only.

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

One simple example of a non-trivial consistent order is to fix a vector of (not all equal) utility weights $\mathbf{u} = (u_1, \ldots, u_n)$ and take $\succeq_t$ to be the representable order $\succeq_{\mathbf{u}}$ on $\mathcal{P}_t$. A larger class of consistent orders are those that satisfy the following condition.

Definition 5 (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*

$$\mu \succeq_s \nu \iff \mu \cdot \mathbf{u}_t \ \geq \ \nu \cdot \mathbf{u}_t \qquad \forall\, \mu, \nu \in \mathcal{P}_s, \quad \forall s \leq t. \tag{6}$$

A sequence $(\mathbf{u}_t)_{t\geq 1}$ is said to locally represent $\succeq$ if (6) holds. The order $\succeq$ is said to be globally representable if there exist $\mathbf{u} \in \mathbb{R}^n$ such that (6) is satisfied for $\mathbf{u}_t = \mathbf{u}$ for all t.

The lexicographic ordering of all multisets is locally representable but not globally. It is easy to check that any locally representable linear order on $\mathcal{P}$ is consistent. More interestingly, we have the following:

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

Remark 4. By the above Proposition, every $\succeq_t$ in a consistent order $\succeq$ on $\mathcal{P}$ is representable. Applying Lemma 1 and repeating the proof of Proposition 1 virtually *ad verbatim*, we note that any consistent order $(\succeq_t)_{t\geq 1}$ is uniquely identified by a sequence of polytopes $(U_t)_{t\geq 0}$ that satisfies the properties listed in Proposition 1.

Proof of Proposition 2. The "if" part is straightforward to verify. 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 $\mu_2, \nu_2 \in \mathcal{P}_s$,

$$\mu_1 \succeq_t \nu_1 \text{ and } \mu_2 \succeq_s \nu_2 \implies \mu_1 \cup \mu_2 \ \succeq_{t+s} \ \nu_1 \cup \mu_2 \ \succeq_{t+s} \ \nu_1 \cup \nu_2, \tag{7}$$

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 $\mu_2 \succ_s \nu_2$.

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

$$\mu_1 \succeq_t \mu_2 \succeq_t \cdots \succeq_t \mu_{N-1} \succeq_t \mu_N. \tag{8}$$

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 (8) are equalities, then it is representable and can be obtained by assigning 1 to all of the utilities. Hence at least one ranking in (8) 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

$$\sum_{i=1}^{N-1} c_i(\mu_i - \mu_{i+1}) = 0 \tag{9}$$

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 Gale (1960), 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 (9) 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 (9) can be rewritten as

$$\sum_{i=1}^{N-1} c_i\mu_i = \sum_{i=1}^{N-1} c_i\mu_{i+1}, \tag{10}$$

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

$$\bigcup_{i=1}^{N-1} \underbrace{\mu_i \cup \ldots \cup \mu_i}_{c_i} = \bigcup_{i=1}^{N-1} \underbrace{\mu_{i+1} \cup \ldots \cup \mu_{i+1}}_{c_i} \tag{11}$$

which contradicts to $c_j > 0$, $\mu_j \succ \mu_{j+1}$ and (7). This contradiction proves the proposition. □

5.2 Proof of Theorem 1

Proof of Theorem 1. 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 $\mathcal{P}_t$ as follows: for any $\mu, \nu \in \mathcal{P}_t$,

$$\begin{aligned} \mu \succeq_t^* \nu \quad \text{or } \mu \succ_t^* \nu \quad \Longleftrightarrow \quad & \text{there exists } a, b \in \mathcal{A} \text{ and a history } h_t \in H_t \\ & \text{such that } \mu = \mu(a, h_t) \text{ and } \nu = \mu(b, h_t) \text{ and} \\ & a \succeq_{h_t} b \text{ or } a \succ_{h_t} b \text{ respectively.} \end{aligned} \tag{12}$$

We need to show that $\succ_t^*$ is antisymmetric. Otherwise, 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. By a repeated application of Axiom 3, we see at once that

Claim 3. *The sequence $\succeq^* = (\succeq_t^*)_{t \geq 1}$ is a consistent order on $\mathcal{P}$ (in the sense of Definition 4).*

Applying Proposition 2 we note that $(\succeq_t^*)_{t \geq 1}$ is locally representable. Any $(\mathbf{u}_t)_{t \geq 1}$ representation of $(\succeq_t^*)_{t \geq 1}$ then, by construction of the latter, will constitute an ex-post utility representation of $\mathcal{D}$.

The proof of Theorem 1 is therefore complete upon giving the proofs of Claim 1 and Claim 2 and verifying that fact that $2 \Rightarrow 1$. All of these are relatively straightforward but nevertheless relegated to the Appendix. □

6 Discussion

In this section, we shall discuss several aspects our model including the empirical implications and robustness of Theorem 1. Let $L(\mathbf{n})$ in $\mathbb{R}^n$ denote the linear hyperplane whose normal is $\mathbf{n}$, i.e.

$$L(\mathbf{n}) \;=\; \{\mathbf{x} \in \mathbb{R}^n \;:\; \mathbf{n} \cdot \mathbf{x} = 0\}. \tag{13}$$

6.1 On the set of all utility representations

Theorem 1 shows that a utility representation obtains under fairly weak assumptions. The set of feasible utility assignments were given in Proposition 1. Note that since a utility assignment $\mathbf{u} \in U_t$ is already normalised, no two elements of $ri(U_t)$ are affine transformations of each other (see the proof of Lemma 1). In this sense, the DM may be ambiguous about the actual value she assigns to individual rewards although the relative ranking of the rewards remains unchanged over time. The following example illustrates how the possible utility assignments to the rewards, i.e. the polytopes in Proposition 1, evolve.

Example 3. Assume there are three rewards, i.e. $\mathcal{R} = \{1, 2, 3\}$. Recall from the proof of Theorem 1 that a $\mathcal{D}$ that satisfies Axioms 1-3 is equivalent to a consistent ordering over $\mathcal{P}$ as given in Definition 4 and an ex-post utility representation of $\mathcal{D}$ is a local utility representation of $\succeq$ as given in Definition 5. Let $\succeq = (\succeq_t)_{t \geq 1}$ be that ordering over $\mathcal{P}$.

Since $\mathcal{P}_1 = \mathcal{R}$, the order $\succeq_1$ is simply a ranking of the three rewards. Let us assume that $1 \succ_1 2 \succ_1 3$. Then any choice of utilities for the rewards $u_1 > u_2 > u_3$ would represent $\succeq_1$ on $\mathcal{P}_1$. One can normalise these by setting the least utility to zero and scaling them to add to one so that vectors from the relative interior of

$$U_1 = \{(u_1, 1 - u_1, 0) \mid u_1 \in [1/2, 1]\}$$

effectively give us all representations of $\succeq_1$.

Next, we consider $\mathcal{P}_2$. The multisets in $\mathcal{P}_2$ are listed in the table below with multiplicities for each multiset appearing in the first three columns. In the rightmost column we give the notation for each multiset.

	1	2	3	*Notation*
μ_1	2	0	0	1^2
μ_2	1	1	0	12
μ_3	1	0	1	13

	1	2	3	*Notation*
μ_4	0	2	0	2^2
μ_5	0	1	1	23
μ_6	0	0	2	3^2

Table 1: $\mathcal{P}_2 = \{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6\}$.

Consistency requires that $\succeq_2$ must necessarily rank $1^2 \succeq_2 12$ as the top two multisets and $23 \succeq_2 3^2$ as two bottom ones, Furthermore, 13 and 2^2 must be placed in between 12 and 23 although we have freedom to choose the relation between them. Thus, we have three possible orderings of $\mathcal{P}_2$ that would be consistent with the given $\succeq_1$ depending on how this ambiguity is resolved. If $13 \sim_2 2^2$, representability gives $u_1 = 2u_2$, which immediately pins down $U_2 = \{(2/3, 1/3, 0)\}$. Moreover, for all $t > 2$ we will also have $U_t = U_2 = \{(2/3, 1/3, 0)\}$.

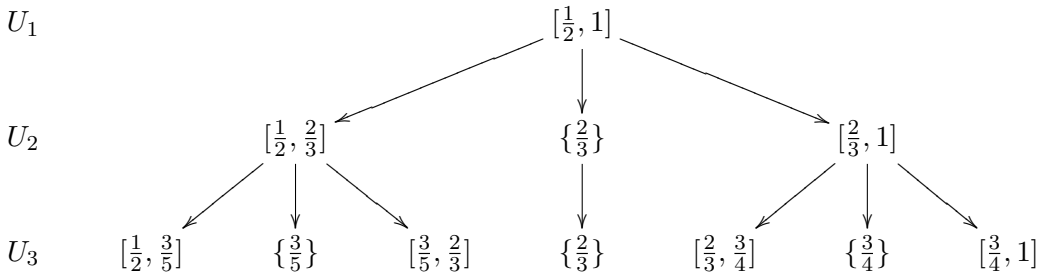

Figure 1: Schematic description of consistent orders on $\mathcal{P}_t[3]$, $t \leq 3$, when $1 \succ_1 2 \succ_1 3$.

If, on the other hand, $13 \succ_2 2^2$, we have $U_2 = \{(u_1, 1 - u_1, 0) \,|\, u_1 \in [2/3, 1]\}$ and in the residual case of $2^2 \succ_2 13$, we have $U_2 = \{(u_1, 1 - u_1, 0) \,|\, u_1 \in [1/2, 2/3]\}$. Going further to $\mathcal{P}_3 = \mathcal{P}_3[3]$, the possibilities are listed in Figure 3. In the figure, the set U_t is encoded by the interval of values that u_1 is allowed to take. For a u_1 that lies in the different sets listed in the terminal nodes of the graph, we obtain a distinct preference relation on $\mathcal{P}_3$ that is consistent with $1 \succ_1 2 \succ_1 3$. The above process can be continued for $t > 3$ along similar lines.

As illustrated in the above example, the DM becomes increasingly precise over the values assigned to the rewards. This is also true in general as $U_{t+1} \subseteq U_t$. However, the limiting set $\cap_{t\geq 1} U_t$ may not constitute a representation as shown in the following example.

Example 4. Consider the case where there are three rewards, i.e. $\mathcal{R} = \{1, 2, 3\}$ and $\mathcal{D} = (\succeq_h)_{h\in H}$ is the lexicographic ordering, where

$$a \succ_h b \quad \Leftrightarrow \quad \begin{cases} \text{if } \mu_1(a,h) > \mu_1(b,h) \\ \text{if } \mu_1(a,h) = \mu_1(b,h) \text{ and } \mu_2(a,h) > \mu_2(b,h). \end{cases} \tag{14}$$

This ordering is represented by choosing U_t whose elements are of the form $(u_1, u_2, u_3) = (u_1, 1-u_1, 0)$ where $u_1 \in (t/(t+1), 1)$. And yet, there cannot be a global representation of this lexicographic ordering since the intersection $\bigcap_{t=1}^{\infty} U_t = (1, 0, 00)$ is a boundary point.

Recall that although a global utility representation may not exist but if one does, it must be unique. (See Proposition 2.)

To ensure the existence of a global utility representation, one requires some form of the Archimedean axiom on the DM's behavior. We do not pursue this here since the role of such axioms is well understood in Decision Theory.

6.2 Random Rewards and Observed Behavior

For the rest of this section, suppose that there is a stationary stochastic process X_t that generates the rewards. From the probability measure that governs this process, one can compute the probability that an action a_i receives the reward j at any given date. Denote this probability by q_{ij}. To each action $a_1, \ldots, a_m$, we then have a corresponding lottery $\mathbf{q}_i = (q_{i1}, \ldots, q_{in})$ over the set of rewards.

Consider, for the moment, a DM that is fully aware of the environment and satisfies the expected utility hypothesis. Given vNM utility vector for the rewards $\mathbf{u} = (u_1, \ldots, u_n)$, naturally we shall say that an action a_{i^*} is a *best action* for the DM if

$$\mathbf{u} \cdot \mathbf{q}_{i^*} \quad \geq \quad \mathbf{u} \cdot \mathbf{q}_i \quad \text{for all } 1 \leq i \leq m. \tag{15}$$

Our interest here is in the observed behavior in the above environment of a DM who does not know the environment but satisfies Axioms 1-3 vis-a-vis a DM that knows the environment. We will show the following.

Theorem 2. *Consider a DM that is consistent with Axioms 1-3 and admits a global utility representation* $\mathbf{u}$*. Suppose the stationary stochastic process* X_t *is such that there is a unique best action. Then, with probability one, the DM chooses the best action corresponding to* $\mathbf{u}$ *at all but finitely many dates.*

Remark 5. The best action is determined by a finite set of linear inequalities (15). For a generic choice of probabilities and global utility vectors, the existence of a unique best action is therefore assured. Thus, the existence of a unique best action in Theorem 2 is a weak requirement.

To see how Theorem 2 obtains, pick any two actions, say a_1 and a_2. Suppose that our stationary stochastic process produces reward r_i for a_1 and reward r_j for a_2 with probability p_{ij}. We model this event by the vector $\mathbf{f}_{ij} = \mathbf{e}_i - \mathbf{e}_j$. So without loss of generality we may assume that the stochastic process X_t actually produces not prizes but these vectors and let $Y_t = X_1 + \cdots + X_t$. To illustrate, suppose $\mathcal{R} = \{1, 2, 3\}$ and the following sequences of prizes are realized

$$\begin{array}{lcccccccccccc} a_1\text{:} & 1 & 1 & 2 & 3 & 2 & 2 & 1 & 3 & 3 & 3 & 1 & 2 & \ldots \\ a_2\text{:} & 2 & 3 & 1 & 1 & 3 & 1 & 2 & 2 & 1 & 2 & 3 & 3 & \ldots \end{array}$$

The initial five realizations of our stochastic process X_1, X_2, X_3, X_4 and X_5 are respectively

$$\mathbf{f}_{12} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},\ \mathbf{f}_{13} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},\ \mathbf{f}_{21} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},\ \mathbf{f}_{31} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},\ \mathbf{f}_{23} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

and correspondingly

$$Y_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},\ Y_2 = \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix},\ Y_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},\ Y_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},\ Y_5 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

We are interested in the behavior of $Y_t = X_1 + X_2 + \ldots + X_t$. For, by Theorem 1, a DM with a global utility representation $\mathbf{u}$ chooses the first action at moment t if $Y_t \cdot \mathbf{u} > 0$, chooses the second action at moment t if $Y_t \cdot \mathbf{u} < 0$ and chooses any action when $Y_t \cdot \mathbf{u} = 0$.

Observe that the coordinates of Y_t will necessarily sum to zero. Therefore, Y_t lies on the hyperplane $L(\mathbf{1})$ where $\mathbf{1} = (1, \ldots, 1)$. In fact, Y_t is a random walk on the integer grid in $L(\mathbf{1})$ generated by the vectors $\mathbf{f}_{ij}$. These vectors are not linearly independent. For instance, in the above example, we have $\mathbf{f}_{12} + \mathbf{f}_{23} = \mathbf{f}_{13}$. Thus if we take $\mathbf{f}_{12}$ and $\mathbf{f}_{23}$ as a basis for this grid, then $\mathbf{f}_{13}$ will represent a diagonal move. In general, the $m-1$ vectors $\{\mathbf{f}_{12}, \mathbf{f}_{23}, \ldots, \mathbf{f}_{m-1\,m}\}$ form a basis, so that having m prizes we have a walk on an $m-1$ dimensional grid with a drift

$$\mathbf{d} = \sum_{i \neq j} p_{ij} \mathbf{f}_{ij}.$$

We are now ready to prove the theorem:

Proof of Theorem 2. Consider the hyperplane $L = L(\mathbf{u})$. With no loss in generality, label the unique best action as a_1 and pick any other action and label it a_2. It suffices to show that with probability one, the DM chooses a_1 in all but finitely many periods. Axiom 1 will then complete the proof.

First, note that[10]

$$\mathbf{q}_1 - \mathbf{q}_2 \quad = \quad \mathbf{d}. \tag{16}$$

By hypothesis then, $\mathbf{u} \cdot \mathbf{d} > 0$ which is to say that $\mathbf{d}$ lies above the hyperplane L. By the Strong Law of Large numbers, $\frac{1}{t} Y_t$ converges almost surely to $\mathbf{d}$. Hence, with probability one, Y_t also lies above L for all but finitely many t. Recalling that the DM may choose a_2 only when $\mathbf{Y}_t \cdot \mathbf{u} \leq 0$, the claim follows readily upon appealing to Axiom 1. □

6.3 Empirical Test of the Axioms

In this section, our interest is in what an external observer can infer about a DM, who is consistent with Axiom 1-3, simply by observing her sequential choices and the sequence of rewards.

To first illustrate and simplify exposition, assume that there are only two actions and $\mathcal{R} = \{1, 2, 3\}$. Suppose the following sequence of rewards are realised:

[10] To see (16), note that $q_{1i} = \sum_{j=1}^{n} p_{ij}$ and $q_{2i} = \sum_{j=1}^{n} p_{ji}$. Next, observe that the ℓth coordinate of any $\mathbf{f}_{ij}$ is non-zero only if ℓ is either i or j. Therefore, $d_i \mathbf{e}_i = \sum_{j=1}^{n} p_{ij} \mathbf{f}_{ij} + \sum_{j=1}^{n} p_{ji} \mathbf{f}_{ji}$ or that $d_i \mathbf{e}_i = \big(\sum_{j=1}^{n} (p_{ij} - p_{ji})\big) \mathbf{e}_i$.

a_1:	1	1	2	3	2	2	1	3	3	3	1	2
a_2:	2	3	1	1	3	1	2	2	1	2	3	3

By observing the choices of the DM along this sequence, the DM's preferences over the actions following *all* two period histories (i.e. $\succeq_{h_2}$ for $h_2 \in H_2$) will be revealed. Indeed, to discover this relation, all we need to do is figure out how she ranks the six multisets in $\mathcal{P}_2$ listed in Table 3. The comparisons $1^2 \,?\, 2^2$, $2^2 \,?\, 3^2$ and $1^2 \,?\, 3^2$ will be encountered at moments 1,5 and 9. The comparisons $1^2 \,?\, 23$, $13 \,?\, 2^2$ and $12 \,?\, 3^2$ will be encountered at moments 4,8 and 12, respectively. When the DM resolves these comparisons by choosing one action or another the whole preference order on $\mathcal{P}_2$ will be revealed. On the other hand, if 2 is the least valued prize, the sequences

a_1:	1	3	1	3	1	3	...
a_2:	2	2	2	2	2	2	...

never reveals agent's preferences between rewards 1 and 3.

More generally, one can design particular sequences of rewards and by observing those rewards, one can figure out what $\succeq_{h_t}$ is for all $h_t \in H_t$. This amounts to constructing a sequence of rewards that reveals the implied preferences on $\mathcal{P}_t[n]$. The idea is, at every step, to undo all the previous comparisons and then to present the agent with the new one. Also note that for such revelation to occur the DM must switch from one action to another. Such sequences and switching can be engineered via experiments in a laboratory. However, if rewards are instead drawn at random, we know from Theorem 2 and Remark 5, the DM rarely switches.[11]

The point is, that while it is feasible to discover a DM's characteristics using experimental data from the laboratory, typically only very limited conclusions can be drawn of a DM using the empirical data on her choices out in the field (where the rewards are drawn at random). We emphasise however, that the inability to deduce the preference relation does undermine the refutability of our Axioms.

Proposition 3. *Suppose we observe that a DM, who is known to have non-trivial preferences, has chosen actions $a_1, \ldots, a_t, \ldots, a_T$ in successive periods and the history h_T. Suppose that any $\mathbf{u} = (u_1, \ldots, u_n)$ that satisfies the system of inequalities*

$$\mu(a_t, h_t) \cdot \mathbf{u} \;\geq\; \mu(a, h_t) \cdot \mathbf{u} \quad a \in \mathcal{A},\ t = 1, \ldots, T, \tag{17}$$

[11] Should the non-generic possibility of a driftless $\{Y_t\}$ occur (the random process described Section 6.2) with $n = 3$ rewards, the walk will be recurrent and the utilities will still be revealed. Not so for $n > 3$.

where h_t is the sub-history of h_T until period t, is necessarily of the form $u_i = u_j$ for all $i, j = 1, \ldots, n$. Then the DM violates one of the Axioms 1-3.

Proof. Suppose that, by way of contradiction, the DM obeys Axioms 1-3. By Theorem 1, the DM is then ex-post rational. Applying Proposition 1 and choosing any ex-post utility representation $(\mathbf{u}_t)_{t\geq 1}$ of such a DM, we note that $\mathbf{u}_T \in U_T$ must be a solution to the above system of inequalities. Furthermore, since U_T lies on a facet of the unit simplex $\Delta \subset \mathbb{R}^n$, $\mathbf{u}_T$ is in fact a solution of (17) such that not all of its coordinates are equal. This contradiction establishes the Proposition. □

6.4 Ex-post Rationality with Bounded Recall

Throughout, we had assumed that the DM can track the entire history. An alternative hypothesis is that she can only track the last k observations. In fact such a hypothesis may be more plausible if the underlying process X_t which produces rewards for actions is not stationary. Indeed, if X_t becomes uncorrelated after time k, then, even if the DM remembers old observations, they become of no use. A DM who understands this aspect of the environment (but still possibly ignorant about other aspects) may use only the last k observations.

With bounded recall then, the DM is only required to rank in a consistent fashion multisets of cardinality not greater than k. But then, Proposition 2 breaks down.

The following is a consistent linear order on $\mathcal{P}_3[4]$ (taken from Sertel and Slinko (2005)) but is not representable.

$$1^3 \succ 1^2 2 \succ 1^2 3 \succ 1^2 4 \succ 12^2 \succ 123 \succ 124 \succ 13^2 \succ 134 \succ 2^3 \succ$$
$$2^2 3 \succ 14^2 \succ 2^2 4 \succ 23^2 \succ 234 \succ 24^2 \succ 3^3 \succ 3^2 4 \succ 34^2 \succ 4^3.$$

Indeed we have:

$$2^2 3 \succ 14^2, \quad 24^2 \succ 3^3, \quad 134 \succ 2^3. \tag{18}$$

If this ranking were representable then the respective system of inequalities

$$\begin{array}{rcl} 2u_2 + u_3 & \geq & u_1 \\ u_2 & \geq & 3u_3 \\ u_1 + u_3 & \geq & 3u_2 \end{array}$$

would have a non-zero non-negative solution, but it has not. These inequalities imply $u_1 = u_2 = u_3 = u_4 = 0$.

Whether some weaker form of representability of the DM can be achieved remains a topic for future research.

7 Related Literature

There is a large body of literature that *begins* with the assumption that the DM is a long run expected utility maximiser. Certain simple thumb rules are posited and the question is if these simple rules yield the optimizing behavior of a fully rational player. See Lettau and Uhlig (1995), Schlag (1998) and Robson (2001) among others. Given the Axioms and the representation, the analysis presented in Section 6.2 is in this spirit.

The main focus of this paper is however on the axiomatic development of the DM's behavior that attempts to capture from first principles how a DM learns. From this standpoint, Börgers, Morales, and Sarin (2004) and ER are two works that share this concern. The former considers behavioral rules that take the action/payoff pair that was realised in the previous period and map it to a mixed strategy on $\mathcal{A}$. The desirable properties that are imposed on a behavioral rule (monotonicity, expediency, unbiasedness etc.) involve comparing the payoffs realised in the previous periods. Thus, no distinction is being made between payoffs and rewards.

ER is the closest relative of this work as it explicitly considers axioms on sequences of preferences in a dynamic context. Like us, ER study a family of preference relations $\{\succeq_{h_t}\}_{t\geq 1}$ on the set of actions $\mathcal{A}$ indexed by histories. There are however both formal and conceptual differences. Unlike us, they find it necessary to extend $\succeq_{h_t}$ to a preference relation over $\Delta(\mathcal{A})$, the set of all lotteries over $\mathcal{A}$ while in our paper we do not need lotteries. They too, just as in Börgers, Morales, and Sarin (2004), assume that the rewards are monetary payoffs. In our setting the outcome of an action is an arbitrary reward. This distinction is important since, as we have seen, at each stage, there is in fact a convex polytope of endogenously determined utilities for the rewards that determines the DM's behavior. Interestingly, our representation result Theorem 1 shows that our three axioms enough to at once *jointly* determine the updating method and the payoffs to underlying rewards.

Conceptually, ER's focus is on the transition from the preference relation $\succeq_{h_t}$ to $\succeq_{h_{t+1}}$ in response to the most recently observed rewards. A driving assumption in their work is

to treat history as being important only to the extent of determining the current preference relation on $\Delta(\mathcal{A})$. On the other hand, only Axiom 3 here relates preferences of one date to another but it is too weak to allow to determine $\succeq_{h_{t+1}}$ given $\succeq_{h_t}$ and the current vector of rewards. Under our set of axioms, it is entirely possible that DM's ordering of the actions at a given date coincide after two different histories but subjected to the same vector of rewards in the current period this ordering can be updated to two different rankings. In other words, one can have $\succeq_{h_t}=\succeq_{h'_t}$ but $\succeq_{h_{t+1}}\neq\succeq_{h'_{t+1}}$ for $h_{t+1}=(h_t,\mathbf{r})$ and $h'_{t+1}=(h'_t,\mathbf{r})$. In other words, our formulation allows a level of *path dependence* that is absent in their model.

It may also be mentioned that the axioms of ER are in the spirit of reinforcement learning – upon observing the rewards to various actions, the relative probability of choosing an action is revised with an eye on the size of the reward. Axiom 3 here on the other hand, places a restriction on the updating behavior only upon the realization of a reward vector that is constant across actions. This allows the analysis here to be (trivially) in the spirit of the learning direction theory presented in Selten and Buchta (1999) and Selten and Stoecker (1986). Not surprisingly our results on the expected-utility-like maximization behavior of the DM is in sharp contrast to the replicator dynamic (or its generalizations) characterised in ER.

Next, we address the relation of our work to *Case Based Decision Theory* of Gilboa and Schmeidler. We shall restrict the comparison of this work with Gilboa and Schmeidler (2003) that is most characteristic of their work. Their framework consists of two primitives. First, in their framework there is a set of objects denoted by X and interpreted varyingly as eventualities or actions, that need to be ranked. Second, there is a set of all conceivable "cases", which they denote by $\mathbb{C}$ and which is assumed to be infinite. A case should be interpreted as a "distinct view" or an occurrence that offers credence to the choice of one act over another or a relative increase in the likelihood of one eventuality over another. Their decision maker is thus a family of binary relations $(\succeq_M)$ on X, where $M\subseteq\mathbb{C}$ is the set of actual cases that are available in the agent's database at the time of making a choice. (See also Gilboa and Schmeidler (1995).) M is assumed to be finite but $\mathbb{C}$ is necessarily infinite. Translated to our framework, $X=\mathcal{A}$ and the set of all conceivable "cases" would be the set of all vectors of rewards $\mathbf{r}=(r_1,\ldots,r_m)\in\mathcal{R}^m=\mathbb{C}$. As $\mathbb{C}$ is then finite, formally it is not possible to embed our model in theirs.

There is also a conceptual difference. They consider each case to be kind of a "distinct view" that gives additional credence to the choice of an act. In our analysis, it is not

just the set of "distinct views" but also "how many" times any of those given views are expressed is important. To elaborate further, Gilboa and Schmeidler (2003) work with a family of relations $\succeq_M \subseteq X \times X$ with M a finite set of $\mathbb{C}$ being the parameter. $\mathbb{C}$ is necessarily infinite. We, on the other hand, work with a family of relations $\succeq_\mu \subseteq X \times X$ where the parameter μ is a *multiset* of $\mathbb{C}$ (a finite set).

8 Conclusion and Future Work

In this paper, we have presented a theory of choice in a complex environment, a theory that does not rely on the action/state/consequence approach. Three simple axioms secure that the DM has an ex-post utility representation and behaves as an expected utility maximiser with regard to the empirical distribution of rewards.

In future work we expect to relax the following assumptions:

(a) that the agent is learning in a social setting. A history in this case would contain missing observations,

(b) allow the DM to have bounded recall,

(c) allow for the possibility that the DM faces a possibly different problem in each period (thus making the analysis comparable to case based decision theory of Gilboa and Schmeidler (1995)).

Appendix

Proof of Lemma 1. We recall a few basic facts about hyperplane arrangements in $\mathbb{R}^n$ (see Orlik and Terao (1992) for more information about them). A *hyperplane arrangement* A is any finite set of hyperplanes. Given a hyperplane arrangement A and a hyperplane J, both in $\mathbb{R}^n$, the set

$$A^J = \{L \cap J \mid L \in A\}$$

is called the *induced arrangement of hyperplanes* in J.

A *region* of an arrangement A is a connected component of the complement U of the union of the hyperplanes of A, i.e., of the set

$$U = \mathbb{R}^n \setminus \bigcup_{L \in A} L.$$

Any region of an arrangement is an open set.

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

$$\mu \succeq_{\mathbf{u}} \nu \Longleftrightarrow \sum_{i=1}^{n} \mu(i)u_i \geq \sum_{i=1}^{n} \nu(i)u_i. \tag{19}$$

Any order on $\mathcal{P}_t$ that can be expressed as above 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 arrangement.

Consider the hyperplane arrangement

$$A(t, n) = \big\{ L(\mu - \nu) \;\mid\; \mu, \nu \in \mathcal{P}_t[n] \big\}. \tag{20}$$

where $L(\mu - \nu)$ is as given by Eq. (13).

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.

Let us note that in (19) 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.

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:

$$\begin{array}{rcll} (\mu-\nu)\cdot\mathbf{u} & = & 0 & \text{whenever } \mu \sim_{\mathbf{u}} \nu, \\ (\mu-\nu)\cdot\mathbf{u} & > & 0 & \text{whenever } \mu \succ_{\mathbf{u}} \nu, \end{array} \qquad \forall\, \mu,\nu \in \mathcal{P}_t. \tag{21}$$

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 $\succeq = (\succeq_t)_{t\geq 1}$ be a consistent order on $\mathcal{P}$. By Proposition 1 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 also note that the intersection $\bigcap_{t=1}^{\infty} U_t$ has exactly one element. This is immediately implied by the following $\square$

Proof of Theorem 1. We give the proofs of Claim 1, Claim 2 and the fact that $2 \Rightarrow 1$ in sequence.

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 $\mu \succeq_t^* \xi$.

Finally, we turn to the implication $2 \Rightarrow 1$. That Axioms 1 and 2 are met by an ex-post rational $\mathcal{D}$ is easy to see. To prove that Axiom 3 holds, suppose that the sequence of utility vectors $(\mathbf{u}_t)_{t\geq 1}$ represents the DM and suppose $a \succeq_{h_t} b$ and at the moment t both actions a and b yield a reward i. Then we have $\mu(a, h_{t+1}) = \mu(a, h_t) + \mathbf{e}_i$ and $\mu(b, h_{t+1} = \mu(b, h_t) + \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith vector of the standard basis of $\mathbb{R}^n$. Due to consistency condition, the utility vector $\mathbf{u}_{t+1}$ can also be used for comparisons of histories shorter than $t + 1$, so we have

$$\mu(a, h_t) \cdot \mathbf{u}_{t+1} \geq \mu(a, h_t) \cdot \mathbf{u}_{t+1}$$

From here we obtain:

$$\mu(a, h_{t+1}) \cdot \mathbf{u}_{t+1} = (\mu(a, h_t) + \mathbf{e}_i) \cdot \mathbf{u}_{t+1} \geq (\mu(b, h_t) + \mathbf{e}_i) \cdot \mathbf{u}_{t+1} = \mu(b, h_{t+1}) \cdot \mathbf{u}_{t+1}.$$

Hence $a \succeq_{h_{t+1}} b$.

□

References

Anscombe, F., and R. Aumann (1963): "A definition of subjective probabiliy," *Annals of Mathematical Statistics*, 34, 199–205.

Aumann, R. J., and J. H. Dreze (2008): "Rational Expectations in Games," *American Economic Review*, 98(1), 72–86.

Blume, L. E., D. A. Easley, and J. Y. Halpern (2006): "Redoing the Foundations of Decision Theory," in *Tenth International Conference on Principles of Knowledge Representation and Reasoning KR(2006)*, pp. 14–24.

Börgers, T., A. J. Morales, and R. Sarin (2004): "Expedient and Monotone Learning Rules," *Econometrica*, 72(2), 383–405.

Brown, G. W. (1951): "Iterative solutions of games by fictitious play," in *In Activity Analysis of Production and Allocation*, ed. by T. C. Koopmans. New York: Wiley.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69(4), 891–934.

Easley, D., and A. Rustichini (1999): "Choice without Beliefs," *Econometrica*, 67(5), 1157–1184.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 74(4).

Fudenberg, D., and D. K. Levine (1998): *The Theory of Learning in Games*. MIT Press.

Gale, D. (1960): *The Theory of Linear Economic Models*. McGraw-Hill, New-York.

Gigerenzer, G., and R. Selten (eds.) (2001): *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

Gilboa, I., and D. Schmeidler (1995): "Case-Based Decision Theory," *The Quarterly Journal of Economics*, 110(3), 605–39.

——— (2001): *A Theory of Case-Based Decisions*. Cambridge University Press.

——— (2003): "Inductive Inference: An Axiomatic Approach," *Econometrica*, 71(1), 1–26.

Hopkins, E. (2002): "Two Competing Models of How People Learn in Games," *Econometrica*, 70(6), 2141–2166.

Karni, E. (2006): "Subjective expected utility theory without states of the world," *Journal of Mathematical Economics*, 42(3), 325–342.

Knight, F. H. (1921): *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co.

Lettau, M., and H. Uhlig (1995): "Rules of Thumb and Dynamic Programming," *Tilburg University Discussion Paper*.

Orlik, P., and H. Terao (1992): *Arrangements of Hyperplanes*. Springer-Verlag, Berlin.

Robson, A. J. (2001): "Why Would Nature Give Individuals Utility Functions?," *Journal of Political Economy*, 109, 900–914.

Savage, L. J. (1954): *The Foundations of Statistics*. Harvard University Press, Cambridge, Mass.

Schlag, K. H. (1998): "Why Imitate, and If So, How?, : A Boundedly Rational Approach to Multi-armed Bandits," *Journal of Economic Theory*, 78(1), 130–156.

Selten, R. (2001): "What is bounded rationality?," in *Bounded Rationality: The Adaptive Toolbox*, ed. by G. Gigerenzer, and R. Selten. MIT Press.

Selten, R., and J. Buchta (1999): "Experimental Sealed-Bid First Price Auctions with Directly Observed Bid Functions.," *In D. Budescu, I. Erev, and R. Zwick (eds.),* Games and Human Behavior: Essays in the Honor of Amnon Rapoport *NJ: Lawrenz Associates Mahwah.*

Selten, R., and R. Stoecker (1986): "End Behavior in Sequences of Finite Prisoners' Dilemma Supergames: A Learning Theory Approach," *Journal of Economic Behavior and Organization.*, 7, 47–70.

Sertel, M. R., and A. Slinko (2005): "Ranking Committees, Income Streams or Multisets," *Economic Theory*.