On choice in a complex environment

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

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

[Draft: please do not quote – 14 March, 2009]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 of say Savage (1954) or Anscombe and Aumann (1963). This paper presents an axiomatic foundation and 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 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 sufficient to show that the DM has an ex-post utility representation and behaves as if she maximises expected utility in a certain sense. This also enables us to show that, if rewards are drawn by a stationary stochastic process, the observed behavior of such a DM almost surely cannot be distinguished from one that is fully cognizant of the environment.

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

*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] 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. The introduction to 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.

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 $D := (\succeq_{h_t})_{h_t \in H}$, 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. We impose axioms on 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 on the basis of her observations. To use a term coined by Reinhard Selten, the DM indulges in some kind of *ex-post rational* behavior[2] when one looks at what might have been better last time and adjusts the next decision in this direction. 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 Börgers, Morales, and Sarin (2004), Schlag (1998), Gigerenzer and Selten (2002) 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 Easley and Rustichini (1999) (hereafter ER) which makes it the closest relative of this paper.

We defer a complete discussion of the relation of this work to ER (and other literature) to Section 5. We do note here however that there are significant differences both in the formal modeling details and in the conceptual basis for the axioms. The axiomatised class of adaptive learning procedures in their paper is very different and includes, for example, the replicator dynamics. These, for instance, in our formulation allow for considerable *path dependence* of the DM's preferences over actions across time which are ruled out by ER. Furthermore, as we explain below, our results will show that the DM may be initially ambiguous on how to value the rewards but becomes increasingly precise over time. This feature too is absent in ER.

What we do share with ER and many of 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 sharemarket. For, in these cases there is a enough diversity of preferences so that all the actions are chosen in each period by various individuals and outcomes are publicly

[2]See the Chapter "What is Bounded Rationality" in Gigerenzer and Selten (2002) and the informal discussion available at http://www.strategy-business.com/press/16635507/05209.

observable.

We establish three results — Theorem 1, a representation result for D, Theorem 2, a characterization of DM's observed behavior when the DM has a utility representation and rewards are generated by a stationary stochastic process and Theorem 3, a simple empirical test for refuting the axioms. We shall describe them presently after describing the axioms.

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.

Theorem 1 in Section 2 is an "ex-post utility representation" of D for whom the three axioms described above are satisfied. Essentially, it shows that the above axioms are equivalent to assigning utilities to each of the underlying rewards following any history h_t, calculating the implied utility of any action as the *average* utility of all the rewards realised from that action, and choosing the action with the maximal utility for the next round. These vectors of utilities are not unique and the set of possible utility assignments at time t form a convex polytope U_t in $\mathbb{R}^n$. We show that $U_{t+1} \subseteq U_t$. Thus, in a nutshell, at any moment, the agent chooses between the empirical distributions of the rewards to different actions as if she is an expected utility maximiser. In other words, our three axioms axiomatise an *ex-post rational* DM. Since the polytopes of utility assignments shrink with time, at any moment, the DM learns a bit more about her imputed utilities for the rewards.

There are two noteworthy aspects of Theorem 1. First, it is proved although there are only finitely many rewards — there are no topological assumptions nor do we rely on the possibility of mixed strategies. The proof involves a novel transformation of preferences over actions to a binary relation on the space of *multisets* of rewards. We then prove a representation result, Proposition 1 (see Section 3.1), for orders on multisets to conclude Theorem 1. Proposition 1 is an important technical result that we expect to be of independent interest with applications to other areas of Decision Theory and Social Choice.

Second, the representation given in Theorem 1 is local in the sense that an element of U_t can only be used to rank histories until date t. As t increases to infinity, the intersection of the respective polytopes either consists of a single point or is empty. To ensure a global utility representation, namely a non-empty intersection, requires an axiom of the Archimedean type that is not assumed here.

Next, we ask how a DM that is consistent with our axioms fares relative to one that is fully aware of the environment. Assuming that the rewards are drawn by a stationary stochastic process, we show via Theorem 2 that she behaves almost surely as a DM that

knows the enviroment and maximises expected utility.

Among other issues, we also address the empirical verifiability of the model. Note that all the axioms concern behavior conditional on observed data. One might in principle directly verify the consistency of DM's behavior with regard to the axioms in the observed period of time. This would however be cumbersome. Our second result, Theorem 3, in Section 5.3 shows that consistency with the axioms involves simply checking whether a certain finite system of linear inequalities admit a solution. As such this theorem constitutes a simple test for the empirical refutability of the model.

The rest of the paper is organised as follows. Section 2 introduces the basic setup followed by a definition of ex-post utility representation and a formal description of the axioms in Section 2.2 followed by a discussion. The statement of Theorem 1 is in Section 2.3. Some elements of the proof of this theorem follow the statement of the theorem. However, completion of its proof requires us to consider properties of orderings over multisets. Section 3 is devoted entirely to this. Some readers may find this material to be of independent interest as the representation result proved there can be expected to be applicable in other contexts. Others may want to skip to Section 4 which contains a discussion of various aspects of our results including those mentioned above and future directions of research and then return to Section 3. Theorem 2 and Theorem 3 also appear in this section. Section 5 further reviews the literature and Section 6 concludes.

2 The Model

A Decision Maker must choose from a finite set of $m \geq 3$ actions[3], $\mathcal{A} = \{a_1, \ldots, a_m\}$, 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[4] $\succeq_{h_t}$ 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 $h_{t+1} = (h_t, \mathbf{r}_t)$.

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 her 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.

[3]Our proof depends on the availability of at least three actions. We do not know if this restriction can be removed. However, it looks that, like in voting, the case $m = 2$ is quite special.

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

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 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. 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 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 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 the listed axioms is that the DM is indifferent between all actions following all histories making any analysis redundant.

2.1 Multisets & Ex-Post Utility Maximisation

Here we will introduce the choice rule that we are going to 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 $\sum_{i=1}^{n} \mu(i)$. 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 corresponding to action a and $\mu(a,h) = (\mu_1(a,h), \ldots, \mu_n(a,h))$.

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 $\mu(a,h) = \{1^2, 2^4, 3, 5^2\}$.

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. $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$.

Here comes the rule. A DM applying this rule must know the utilities of the prizes. Let $\mathbf{u} = (\mathbf{n}_1, \ldots, \mathbf{n}_)$ be his vector of 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 $\operatorname{argmax}_i(\mu(a_i,h) \cdot \mathbf{u})$.

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.

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 $\mathbf{u}'_t \mapsto \alpha_t \mathbf{u}_t + \beta_t$ (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 $\Delta \subseteq \mathbb{R}^m$ 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 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[5] of Δ.

2.2 Axioms

Next, we turn to the axioms that are necessary and sufficient for D to admit an ex-post utility representation. 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. 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 $a \succeq_{h'_t} b$.*

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. This also 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] Facet of a polytope is a face of the maximal dimension.

We emphasise that the DM in this model does not try to predict the future outcomes but evaluates the actions based on their past performance. Yet, even if DM were to engage such prediction, intuitively, Axiom 1 would be consistent with a belief that the random environment she faces is one in which the consequences of actions are statistically uncorrelated. Likewise Axiom 2 would be consistent with a belief that the rewards are being generated by a stationary stochastic process.

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 it is worth reiterating 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 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 $\mathbf{r}_t$.

2.3 The Main Theorem

In this section we will formulate and give an outline of the proof of the main theorem which fully characterises a DM satisfying axioms 1–3 as the one which has an ex-post utility representation. Recall that $ri(C)$ denotes the relative interior of a convex set C.

Theorem 1 (Representation Theorem)**.** *The following are equivalent:*

1. *$D = (\succeq_h)_{h\in H}$ satisfies Axioms 1– 3.*
2. *D has an ex-post utility representation. Moreover, there exist a unique sequence of non-empty convex polytopes $(U_t)_{t\geq 0}$ such that $U_t \subseteq \Delta^i$ for some i and*
   (a) *$U_{t+1} \subseteq U_t$ for all $t \geq 1$.*
   (b) *$\bigcap_{t=1}^{\infty} U_t$ consists of a single utility 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.*

   Moreover, 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.

Remark 2. Below, we prove that $\mathcal{D}$, under Axioms 1-3 is essentially equivalent to a partial order over $\mathcal{P}$ that satisfies certain properties. The rest of the proof of this theorem heavily relies on orders on multisets which is taken up in Section 3. The proof is completed by appealing to Lemma 1 that appears toward the end of Section 3.

Proof of Theorem 1. It is easy to show that any DM with an ex-post utility representation satisfies the axioms. For example, let us prove Axiom 3. 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 $a \succeq_{h_{t+1}} b$.

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 $\mu, \nu \in \mathcal{P}_t$,

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

We define also a strict version of it by

$$\begin{aligned} \mu \succ_t^* \nu \iff & \text{ there exists } a, b \in \mathcal{A} \text{ and a history } h_t \in H_t \\ & \text{such that } \mu = \mu(a, h_t) \text{ and } \nu = \mu(b, h_t) \text{ and} \\ & a \succ_{h_t} b \end{aligned} \tag{5}$$

This 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 $\xi \in \mathcal{P}_s$,*

$$\mu \succeq_t^* \nu \iff \mu \cup \xi \succeq_{t+s}^* \nu \cup \xi \tag{6}$$

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

The remainder of the proof will follow from Lemma 1 where orders with property (6) will be studied and their representability proved. □

3 Orders on Multisets and Their Geometric Representation

3.1 Consistent 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 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.[6] To make the orders $\succeq_t$, for various t, related to each other we need to impose some kind of consistency condition on them.

To complete the proof of the main theorem we must study orders on $\mathcal{P}$ with the property (6). 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 (6) from Claim 3, that is, for any $\mu, \nu \in \mathcal{P}_t$ and any $\xi \in \mathcal{P}_s$,*

$$\mu \succeq_t \nu \iff \mu \cup \xi \succeq_{t+s} \nu \cup \xi. \tag{7}$$

[6] 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.

We note that, due to the twosidedness of the arrow in (7), we have also

$$\mu \succ_t \nu \Longleftrightarrow \mu \cup \xi \succ_{t+s} \nu \cup \xi. \tag{8}$$

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*

$$\mu \succeq_s \nu \iff \mu \cdot \mathbf{u}_t \geq \nu \cdot \mathbf{u}_t \qquad \forall\, \mu, \nu \in \mathcal{P}_s, \quad \forall s \leq t. \tag{9}$$

A sequence $(\mathbf{u}_t)_{t\geq 1}$ is said to locally represent $\succeq$ if (9) holds. The order $\succeq$ is said to be globally representable if there exist $\mathbf{u} \in \mathbb{R}^n$ such that (9) is satisfied for $\mathbf{u}_t = \mathbf{u}$ for all t.

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

Proof. See Appendix. □

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 to characterize all such representations. To do this, we need to describe different consistent orders geometrically. This is taken up in the next section.

3.2 Geometric Representation of Consistent Orders

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 arrangment 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

$$\mu \succeq_{\mathbf{u}} \nu \iff \sum_{i=1}^{n} \mu(i)u_i \geq \sum_{i=1}^{n} \nu(i)u_i. \tag{10}$$

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 arrangment.

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

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

and consider the hyperplane arrangement

$$A(t, n) = \big\{ L(\mu, \nu) \;\mid\; \mu, \nu \in \mathcal{P}_t[n] \big\}. \tag{11}$$

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 (10) 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:

$$\begin{array}{rcll} (\mu - \nu) \cdot \mathbf{u} & = & 0 & \text{whenever } \mu \sim_{\mathbf{u}} \nu, \\ (\mu - \nu) \cdot \mathbf{u} & > & 0 & \text{whenever } \mu \succ_{\mathbf{u}} \nu, \end{array} \qquad \forall\, \mu, \nu \in \mathcal{P}_t. \tag{12}$$

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

Proposition 2. *Let $\mathbf{u} \neq \mathbf{v}$ 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. See Appendix. □

To enable easy reference later in the paper, we collect these observations in the form of a Lemma below.

Lemma 1. *Any consistent order $\succeq = (\succeq_t)_{t\geq 1}$ on $\mathcal{P}$ corresponds to a sequence of convex polytopes $(U_t)_{t\geq 0}$ such that $U_t \subseteq \Delta^i$ for some i and*

1. *$U_{t+1} \subseteq U_t$ for all $t \geq 1$ and $\bigcap_{t=1}^{\infty} U_t$ consists of a single utility vector.*
2. *a sequence $(\mathbf{u}_t)_{t\geq 1}$ of vectors of $\mathbb{R}^+_n$ is a utility representation of $\succeq$ if and only if $\mathbf{u}_t$ is an affine transformation of some $\mathbf{u}'_t \in ri(U_t)$.*

Moreover, $\succeq$ is globally representable iff $\bigcap_{t=1}^{\infty} U_t$ is contained in the relative interior $ri(U_t)$ for all $t \geq 1$.

4 Discussion

4.1 On the set of all utility representations

Theorem 1 shows that a utility representation obtains under fairly weak assumptions. Also 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 (remember, α and β in the previous section were found uniquely). 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 Theorem 1 evolve.

Example 2. 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 2 and an ex-post utility representation of $\mathcal{D}$ is a local utility representation of $\succeq$ as given in Definition 3. 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 $\succeq_1$. This set can be encoded by the interval $[1/2, 1]$ for u_1 and we will use this abbreviation below on Figure 2.

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 inbetween 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 $U_t = U_2 = \{(2/3, 1/3, 0)\}$.

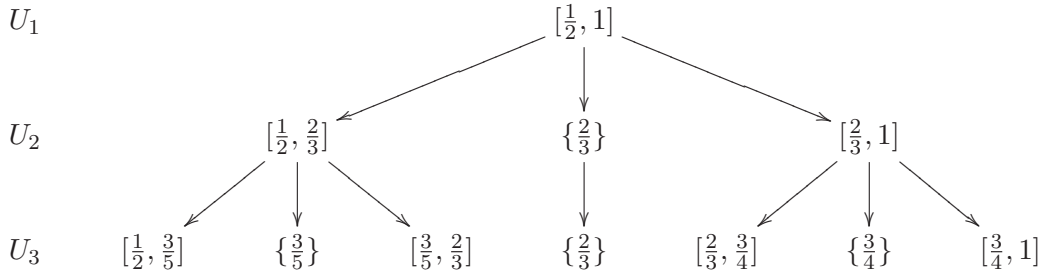

Figure 1: Schematic description of consistent orders on $\mathcal{P}_t[3]$, $t \leq 3$, when $1 \succ_1 2 \succ_1 3$.

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 the figure below. 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 $\mathbb{U}_{t+1} \subseteq \mathbb{U}_t$. In the limit, if a global utility representation exists, the ranking becomes cardinal. However, without further assumptions, it is in general not possible to obtain a global representation. The typical example involving lexicographic preferences illustrates.

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

$$a \succ_h b \quad \Leftrightarrow \quad \begin{cases} \text{if } \mu_1(a,h) > \mu_1(b,h) \\ \text{if } \mu_1(a,h) = \mu_1(b,h) \text{ and } \mu_2(a,h) > \mu_2(b,h). \end{cases} \tag{13}$$

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 $\bigcap_{t=1}^{\infty} U_t = \{1\}$ is a boundary point.

We note that while a global utility representation may not exist, if one exists, it must be unique. Indeed, the set $\bigcap_{t=1}^{\infty} U_t$ has only one element. (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.

4.2 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 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

$$\mathbf{u} \cdot \mathbf{q}_{i^*} \quad \geq \quad \mathbf{u} \cdot \mathbf{q}_i \quad \text{for all } 1 \leq i \leq m. \tag{14}$$

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 at all but finitely many dates.*

Remark 3. Since the best action is determined by a finite set of linear inequalities, for a generic choice of probabilities and global utility vectors, the existence of a unique best action is assured. Thus, the existence a unique best action in Theorem 2 is a weak assumption.

To see how why 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

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

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 H for which $(1, \dots, 1)$ is the normal. In fact, Y_t is a random walk on the integer grid in H 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 $\{\mathbf{f}_{12}, \mathbf{f}_{23}, \dots, \mathbf{f}_{m-1\,m}\}$ 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. Let $L = u^{\perp}$ be the hyperplane of which $\mathbf{u}$ is the normal vector. 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[7]

$$\mathbf{q}_1 - \mathbf{q}_2 \quad = \quad \mathbf{d}. \tag{15}$$

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. □

4.3 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 following sequence of rewards are realised:

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 2. 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 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 $\succeq_{h_t}$ 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. Such sequences can of course be done via experiments in a laboratory. Also note that for such revelation to occur the DM must switch from one action to another. However,

[7] To see (15), 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 $d_i \mathbf{e}_i = \big(\sum_{j=1}^n (p_{ij} - p_{ji}) \big) \mathbf{e}_i$.

if rewards are instead drawn at random, we know from Theorem 2 and Remark 3, the DM rarely switches.[8]

The point is, that while it is feasible to discover a DM's characteristics using expermental 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.

Indeed, suppose that the DM has a utility representation $(\mathbf{u}_t)_{t\geq 1}$ and we can observe one particular history of payoffs and her corresponding choices. Observing k first choices we discover how her vector of utilities $\mathbf{u}_k$ is positioned relative to k hyperplanes in the hyperplane arrangement given in Section 3.2. The DM will have to compare the multisets in the following pairs:

$$\mu(a_1, h_i) \,?\, \mu(a_2, h_i) \qquad (i = 1, 2, \ldots, k)$$

and the ith comparison will give us the information on which side of the hyperplane $L(\mu(a_1, h_i), \mu(a_2, h_i))$ her vector of utilities $\mathbf{u}_i$ (and the same can be said about $\mathbf{u}_k$ too) lies. More exactly, from the choice of DM's action we will learn that $\mu(a_1, h_i) \succeq \mu(a_2, h_i)$ or $\mu(a_2, h_i) \succeq \mu(a_1, h_i)$. We note that neither of them precludes $\mu(a_2, h_i) \sim \mu(a_1, h_i)$. (Indeed, if the DM is indifferent between these two multisets she still has to choose one of the actions.) This will give us k closed half-spaces which may or may not have a non-zero intersection. This gives the key for the following theorem.

Theorem 3. *Suppose that we observe the actions of DM $a_{i_1}, \ldots, a_{i_t}$ during a history $h = h_t$. The DM is consistent with Axioms 1–3 if and only if there exists $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ such that*

$$\mu(a_{i_\ell}, h_\ell) \cdot \mathbf{u} \geq \mu(a, h_\ell) \cdot \mathbf{u} \quad \forall a \in A,\ \ell = 1, \ldots, t, \tag{16}$$

where h_ℓ is the sub-history of h up to date ℓ, which is equivalent to the respective half-spaces having a non-zero intersection.

4.4 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. This is closely related to allowing the random process X_t, which produces rewards for actions, to be not stationary Indeed, if the random process 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 the enviroment (but still possibly ignorant about other aspects) will use only the last k ones.

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 1 breaks down. The

[8] Should the non-generic possibility of a driftless $\{Y_t\}$ occur (the random process described Section 4.2) with $n = 3$ rewards, the walk will be recurrent and the utilities will still be revealed. Not so for $n > 3$.

following consistent 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:

$$2^2 3 \succ 14^2, \quad 24^2 \succ 3^3, \quad 134 \succ 2^3. \tag{17}$$

If this ranking were representable then the respective system of inequalities

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

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.

5 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 4.2 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 consider 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.

Our framework and in particular 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 various contributions to the theory. We shall restrict the comparison of this work with Gilboa and Schmeidler (2003) that is most characteristic of their contributions. 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. 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 $\mathbb{C}$ while $\mathbb{C}$ itself is taken to be finite. Unlike them, we do not need any kind of Archimedean axiom to prove our main theorem.

6 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 Proposition 1. The implication $1 \Rightarrow 2$ 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 $\mu_2, \nu_2 \in \mathcal{P}_s$,

$$\mu_1 \succeq_t \nu_1 \text{ and } \mu_2 \succeq_s \nu_2 \implies \mu_1 \cup \mu_2 \succeq_{t+s} \nu_1 \cup \nu_2, \tag{18}$$

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

$$\mu_1 \succeq_t \mu_2 \succeq_t \cdots \succeq_t \mu_{N-1} \succeq_t \mu_N. \tag{19}$$

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 (19) are equalities, then it is representable and can be obtained by assigning 1 to all of the utilities. Hence at least one ranking in (19) 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

$$\sum_{i=1}^{N-1} c_i(\mu_i - \mu_{i+1}) = 0 \tag{20}$$

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 (20) 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 (20) can be rewritten as

$$\sum_{i=1}^{N-1} c_i \mu_i = \sum_{i=1}^{N-1} c_i \mu_{i+1}, \tag{21}$$

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

$$\bigcup_{i=1}^{N-1} \underbrace{\mu_i \cup \ldots \cup \mu_i}_{c_i} = \bigcup_{i=1}^{N-1} \underbrace{\mu_{i+1} \cup \ldots \cup \mu_{i+1}}_{c_i} \tag{22}$$

which contradicts to $c_j > 0$, $\mu_j \succ \mu_{j+1}$ and (18). This contradiction proves the proposition. □

Proof of Proposition 2. Since $\mathbf{u}$ and $\mathbf{v}$ are normalised we have, in particular, $u_n = v_n = 0$. 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 $\mathbf{x} \cdot \mathbf{v} < 0$. 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. □

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

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.

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), 643–669.

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

Gigerenzer, G., and R. Selten (2002): *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.

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., 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*.