Bounded Rationality: Static versus Dynamic Approaches

Suren Basov[1]

February 13, 2003

[1]The University of Melbourne, Melbourne, Victoria 3010, Australia. JEL classification: C0, D7.

Abstract

Two kinds of theories of boundedly rational behavior are possible. Static theories focus on stationary behavior and do not include any explicit mechanism for temporal change. Dynamic theories, on the other hand, explicitly model the fine-grain adjustments made by the subjects in response to their recent experiences. The main contribution of this paper is to argue that the restrictions usually imposed on the distribution of choices in the static approach are generically not supported by a dynamic adjustment mechanism. The genericity here is understood both in the measure theoretic and in the topological sense.

1 INTRODUCTION

There is a growing empirical evidence that calls into question the utility maximization paradigm. For a description of systematic errors made by experimental subjects, see Arkes and Hammond (1986), Hogarth (1980), Kahneman, Slovic, and Tversky (1982), Nisbett and Ross (1980), and the survey papers by Payne, Bettman, and Johnson (1992) and by Pitz and Sachs (1984). On the basis of this and similar evidence, Conlisk (1996) convincingly argued for the incorporation of bounded rationality in economic models.

Some early attempts to incorporate boundedly rational decision making in economics were made by Alchian (1950), Simon (1957), and Nelson and Winter (1982) among others. But a universal model of boundedly rational behavior still does not exist. The existing models can be divided into two classes: *static* and *dynamic*.

In *static* models individuals choose the better alternatives more frequently than the inferior ones. They were introduced in economics by Luce (1959). It is typical in this type of models to impose some intuitive restrictions on the choice probabilities and study the probability distributions that satisfy these restrictions. Such probabilistic choice models have already found their application in economics. See, for example, McKelvey, Palfrey (1995, 1998), Chen, Friedman, Thisse (1997), Anderson, Goeree, and Halt (1998), Offerman, Schram, Sonnemans (1998), and Anderson, Goeree, and Halt (2001).

In *dynamic* models individuals are assumed to adjust their choices over time in the directions that appear beneficial. The dynamic approach originated in the work of Bush and Mosteller (1955), was introduced in economics by Arrow and Hurwicz (1960), and is represented, for example, by papers of Foster and Young (1990), Fudenberg and Harris (1992), Kandori, Mailath, Rob (1993), Young (1993), Friedman and Yellin (1997), Anderson, Goeree, and Holt (1999), and Friedman (2000).

The distinctive feature of this type of models is an attempt to capture the fine-grain adjustments made by the individuals on the basis of their current experiences. On a very general level, such adjustments produce a stochastic process on the choice set. The probability distribution of choices of a static model can be naturally viewed as the steady state distribution of the stochastic process arising from a dynamic model. For a study of a broad class of dynamic adjustment processes, see Basov (2001).

This paper studies the connections between the properties of the static and the dynamic models. Many dynamic models assume that the process of choice adjustment leads to better choices on average. For the purposes of this paper, I will formalize this idea using the notion of *a locally improving* adjustment process.

Let us call an adjustment process *locally improving* (LI) if the vector of the expected adjustment points into a direction of the increase of the utility. In particular, I will consider a broad class of locally improving Markov processes, I call them *PDS* processes, for which the deterministic part of the generator is linked to the gradient of the utility by a constant symmetric positively definite linear transformation. A restrictive assumption here is that the coefficients of the linear transformation a constant, i. e. they do not vary over the choice space. This assumption is, however, not too restrictive provided the choice space is sufficiently small. Since the results of the paper do not depend on the size of the choice space, this assumption does not drive the results.

A question I will address is: Can the restrictions usually imposed on the probability distribution of choices by the static approach be supported by a generic locally improving adjustment process of this kind? In other words, does the steady state density of a generic locally improving process satisfy the usual axioms of the static approach? To address this question let us start by introducing two important concepts: *payoff monotonicity* (PM) and *Independence of Irrelevant Alternatives* (IIA).

An adjustment process is PM if for any admissible choice set the density of the steady state distribution at x_1 is greater than the density of the steady state distribution at x_2 if and only if x_1 is preferred to x_2. An adjustment process satisfies IIA if the ratio of the steady state probability densities at any two feasible points do not depend on what other choices are available. As we will see below, under some mild regularity assumptions the steady state of each dynamic adjustment process is unique. This, together with the requirement that the restrictions on the steady state density should hold for *any* admissible choice set, implies that the payoff monotonicity and the IIA characterize the *process* rather then a particular distribution.

The first main finding of the paper is the following:

(T1) *any PM process is LI.*

This result is rather intuitive. It claims that for the long-run choice probabilities to be increasing in the payoffs for any choice set the expected adjustment vector should point into a direction of the increase of the utility function. The second main finding of the paper is less intuitive. It states that:

(T2) *a generic PDS process,(i.e. a process for which the deterministic part of the generator is linked to the gradient of the utility by a symmetric positively definite linear transformation), is neither PM nor IIA. Moreover, a generic PDS process that satisfies IIA is not PM.*

The lack of the payoff monotonicity means that given a generic adjustment process one can find a choice set, a pair of choices in it, and a pair of equimeasurable neighborhoods of these choice, such that choices in a neighborhood of the one with a higher payoff are chosen less often in the steady state. The violation of IIA means that given a generic adjustment process one can find a pair of choice sets and a pair of choices that belong to the intersection of these choice sets such that the ratio of the steady state probability densities of the choices depends on the choice set.

Genericity in the above statement can be interpreted both in the measure theoretic and the topological sense. Genericity in the measure theoretic sense means that a property in question does not hold only on a set of dynamic adjustment processes of measure zero. Genericity in the topological sense means that the property in question is violated for a nowhere dense set of dynamic adjustment processes. Note, that the measure theoretic and the topological genericity are not implied by each other. For a discussion see, for example, Oxtoby (1980). In fact, the statement I prove is even stronger. I prove that the set of the PM (IIA) processes can be embedded as a submanifold of a lower dimension into an appropriate subset of LI processes. The same comments about the meaning of genericity as above apply to the second part of the claim. I also show that any PM process is LI, while there is no connection between the LI property and the IIA.

These findings suggest that the usual restrictions on the probability density in the static approach are too strong. They are not supported by a generic dynamic adjustment process. If interpreted from an evolutionary perspective they imply that the adjustment rule the human beings evolved to use with probability one is neither PM nor IIA. Moreover, it is not close to any PM or IIA adjustment rule. Therefore, an explicit modelling of the dynamic adjustment process is important when describing boundedly rational behavior.

This paper is organized as follows. Section 2 describes a broad class of stochastic adjustment processes. In Section 3 I define the main concepts of the paper. Section 4 contains the main results. It states the connections between the concepts defined in Section 3. Section 5 contains the proofs. Section 6 concludes.

2 A MODEL OF INDIVIDUAL BEHAVIOR

Let us assume that an individual repeatedly faces with a problem of choosing an alternative from an open, bounded set $\Omega \subset R^n$ with a smooth boundary. I will refer to such sets as admissible. She adjusts her choices gradually in response to her recent experiences. The adjustment rule produces a stochastic process on the choice set. The expected adjustment vector can be interpreted as an attempt to increase the individual's utility, while the difference between actual and expected adjustment be interpreted as experimentation. I will assume that the stochastic process is Markov and that it possesses a generator. The first assumption is essentially a finite memory assumption, while the last one is purely technical in nature and is made to allow us to employ the continuous time technique.

To build a formal model of adjustment, assume that Σ is a sigma-algebra on Ω, and for any $\Gamma \in \Sigma$, define $P(x(t), \Gamma, \tau)$ to be the transition probability, that is the probability that the individual who at time t made a choice $x(t)$, will make a choice $w \in \Gamma$ at time $t + \tau$. Note that P does not depend on t explicitly, since the process is assumed to be Markov.

Assumption The following limit exists

$$\pounds = \lim_{\tau \to 0} \frac{P(x, \Gamma, \tau) - I}{\tau}, \quad (1)$$

where I is the identity operator.

Operator $\pounds$ is known as the generator of the Markov process (Rogers and Williams, 1994). It can be shown that

$$\pounds = \mu(x)\nabla + \Gamma(x)D^2, \quad (2)$$

where

$$\mu(x) = \lim_{\tau \to 0} \frac{1}{\tau} E(x(t+\tau) - x(t)) \quad (3)$$

$$\Gamma(x) = \lim_{\tau \to 0} \frac{1}{\tau} Var(x(t+\tau) - x(t)) \quad (4)$$

(see, for example, Kanan, 1979). Vector μ captures the deterministic trend in the adjustment rule, while matrix Γ is the covariance matrix of the experimentation errors. The Markov process is completely characterized by vector μ and matrix Γ. I sometimes refer to it as process (μ, Γ).

Assume that Σ is Borel sigma algebra. The definition of Γ implies that it is positively semi-definite. I will assume that it is positively definite. From an economic perspective, it means that the experimentation has a full range. Then, if the initial distribution of choices can be characterized by a density function it can also be characterized by a density function at any $t > 0$ and the evolution of the density is determined by the following system:

$$\frac{\partial f}{\partial t} + div(\mu(x)f) = \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2(\Gamma_{ij}(x)f)}{\partial x_i \partial x_j}, \tag{5}$$

$$\sum_{i=1}^{n}(\frac{1}{2}\sum_{j=1}^{n}\frac{\partial(\Gamma_{ij}(x)f)}{\partial x_j} - \mu_i(x)f)n_i(x) = 0 \text{ on } \partial\Omega, \tag{6}$$

where $n(x)$ is the unit vector normal to the boundary of the choice set $\partial\Omega$ (Ito, 1992). In the rest of the paper I will assume that matrix Γ does not depend on x. The assumption is made for the sake of simplicity only and does not seriously affect the results. The preferences of the individual are given by a twice continuously differentiable utility function $U(\cdot)$.

3 DEFINITION OF SOME CLASSES OF ADJUSTMENT PROCESSES

In this Section I am going to define the main classes of Markov adjustment processes studied in the paper. Let us start with defining the concept of a locally improving process.

Definition 1 *A Markov adjustment process is called locally improving (LI) if*

$$\langle \mu(x), \nabla U(x) \rangle \geq 0 \text{ for } \forall x \in \Omega. \tag{7}$$

Here and throughout the paper $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors. In words, a process is LI if the vector of the expected adjustment of the choice, μ, points into a direction of an increase of the utility. The space of all LI processes is a functional space of an infinite dimension. Next, I define two finite-dimensional subsets of LI.

Definition 2 *Markov adjustment process* (μ, Γ) *is called PD (PDS) if*

$$\mu(x) = B\nabla U,$$

where B is a (symmetric) positively definite matrix with constant coefficients.

Note that $PDS \subset PD \subset LI$ and sets PDS and PD are finite dimensional. The dimension of PDS *is* $n(n+1)$, while that of PD is $n^2+n(n+1)/2$. Hence, both of them can be embedded into space R^k with an appropriate k endowed with the Lebesque measure and be considered as measure spaces.

The concept of a locally improving process is a dynamic concept. Next I am going to define two concepts: payoff monotonicity and independence of irrelevant alternatives. One can naturally think of them as static concepts, since they put restrictions on the steady state density function. However, demanding that these restrictions should hold for any choices space, they can be made the properties of the process. To ensure the soundness of this procedure we need a result from the theory of the stochastic processes. To formulate the result let us assume that $f_{s,\Omega}(x)$ is the stationary solution of the system (5)-(6) normalized by

$$\int_{\Omega} f_{s,\Omega}(x)dx = 1,$$

when the choice set is Ω. It is also known as the steady state density of the Markov process.

Lemma 1 *Assume matrix* Γ *is positively definite. There exists a unique twice continuously differentiable stationary normalized solution of system (5)-(6). Moreover, it is positive everywhere on* Ω *and asymptotically stable.*

For a proof, see Ito (1992). The result states that the steady state density is well defined and is determined by the process rather than by the initial conditions. This allows us to give the following definition.

Definition 3 *A Markov adjustment process is called payoff monotone (PM) if for any choice set* $\Omega \subset R^n$ *and any* x_1, $x_2 \in \Omega$

$$(f_{s,\Omega}(x_1) \geq f_{s,\Omega}(x_2)) \Leftrightarrow (U(x_1) \geq U(x_2)). \tag{8}$$

In words, a Markov adjustment process is PM if for any sufficiently small $\varepsilon > 0$ the steady state probability that the choice is in the $\varepsilon-$ ball centered at the point x_1 is higher then the probability that it is in the $\varepsilon-$ ball centered at point x_2 if and only if alternative x_1 is preferred to alternative x_2. Note, that the payoff monotonicity refers to the *process* rather than to a particular steady state distribution because the latter depends on Ω, while the payoff monotonicity requires (8) to hold for any Ω.

Another important restriction often imposed in the static approach is IIA.

Definition 4 *A Markov adjustment process satisfies independence of irrelevant alternatives (IIA) if for any two choice sets* Ω_1 *and* Ω_2 *and any* $x \in \Omega_1 \cap \Omega_2$

$$f_{s,\Omega_1}(x) = f_{s,\Omega_2}(x). \tag{9}$$

Again, Lemma 1 allows us to talk about IIA *processes* rather then the distributions satisfying IIA. In words, IIA states that the ratio of the steady state probability densities of two choices does not depend on what other choices are available. My next task is to investigate the connections between the PM, the IIA, and the LI (PD, PDS).

4 THE MAIN RESULTS

In this Section I study the connections between the LI, the PM and the IIA and formulate the main results of the paper. I formulate two Theorems. Theorem 1 states that any PM process is LI. Formally, the following result is true.

Theorem 1 *Assume that*

1.

$$\forall x \in \Omega \ (\mu(x) = 0) \Leftrightarrow (\nabla U(x) = 0)$$

2. Set of U_C of the critical points of the utility defined by

$$U_C = \{x \in \Omega : \nabla U(x) = 0\}$$

is finite. Then $PM \subset LI$.

The first assumption states that there is no deterministic adjustment at the critical points of the utility. The second is a regularity assumption. It will always hold if the utility function is analytical. Theorem 1 states that LI is necessary for the process to be PM. It is, however, not sufficient. Moreover, a typical LI process is not PM. To formalize this idea I will restrict attention to the finite-dimensional subclasses of LI, PD and PDS. Theorem 2 states that a typical PD (PDS) process is neither PM nor IIA. Moreover, a typical process that is both PDS and IIA is not PM.

Theorem 2 *Let assumptions of Theorem 1 hold and also assume that the Hessian of the utility has full rank. Then for any $n > 1$*

1. $PD \cap PM$ can be embedded in PD as a submanifold of a lower dimension

2. $PDS \cap PM$ can be embedded in PDS as a submanifold of a lower dimension

3. PM $\subset$ IIA

4. PDS $\cap$ PM can be embedded in PDS $\cap$ IIA as a submanifold of a lower dimension

5. Moreover, if for any non-degenerate constant matrix C there exist i, k such that

$$\frac{\partial^2 U}{\partial x'_i \partial x'_k} \neq 0,$$

where

$$x' = Cx,$$

then PDS $\cap$ IIA can be embedded in PDS as a submanifold of a lower dimension.

The assumption in part 5 of the Theorem states that the utility is not additively separable and does not become additively separable after a non-degenerate linear transformation.

The main message of these Theorems is that the assumptions of the probability density of choices in the static approach are unlikely to hold. Hence, an explicit modelling of the dynamic adjustment process is needed. In doing so it may be useful to restrict attention to LI processes, or even to its finite-dimensional subclasses (for example, PD or PDS). However, doing so does not guarantee good properties for the steady state distribution. The next Section provides a proof of these Theorems, which proceeds through a sequence of lemmata that are of an independent interest.

5 PROOF OF THE MAIN RESULTS

In this Section I develop a sequence of lemmata that eventually lead to the proof of Theorems of the previous Section. I start with a characterization of the steady state distributions for the PM processes. Below I assume that the assumptions of Theorem 1 always hold without stating them explicitly.

Lemma 2 *A Markov process is payoff monotone if and only if for any Ω there exists a strictly increasing continuously differentiable function $g_\Omega : U(\Omega) \rightarrow R_+/\{0\}$ such that*

$$f_{s,\Omega}(x) = g_\Omega(U(x)). \tag{10}$$

Proof. Consider a rational continuous preference relation $\succeq$ defined by

$$(x \succeq y) \Leftrightarrow (U(x) \geq U(y)). \tag{11}$$

The payoff monotonicity implies that $f_{s,\Omega}(\cdot)$ is a utility function that represents preferences relation $\succeq$, which is also represented by $U(\cdot)$. Hence, there exists continuous strictly increasing function $g_\Omega : U(\Omega) \to R$ such that

$$f_{s,\Omega}(x) = g_\Omega(U(x)). \tag{12}$$

According to Lemma 1, $f_{s,\Omega}(\cdot)$ is positive on Ω, hence $g_\Omega(\cdot) > 0$.

To prove that $g_\Omega(\cdot)$ is differentiable let us consider $\overline{U}$, $\overline{U} + \delta U \in U(\Omega)$ and $\delta U \neq 0$. Then $\exists x, x + \delta x \in \Omega$ such that

$$U(x) = \overline{U},\ U(x + \delta x) = \overline{U} + \delta U. \tag{13}$$

Note that, since set U_C is finite, it is always possible to select x in such a way that $\nabla U(x) \neq 0$, which in turn allows to select δx such that $\langle \delta x, \nabla U(x) \rangle \neq 0$. Continuity of $U(\cdot)$ implies that

$$(\delta U \to 0) \Leftrightarrow (\delta x \to 0). \tag{14}$$

Let $\delta x = a \left\| \delta x \right\|$, where a is unit vector pointing in the direction δx. Then

$$\lim_{\delta U \to 0} \frac{g_\Omega(\overline{U} + \delta U) - g_\Omega(\overline{U})}{\delta U} = \frac{\langle \nabla f_{s,\Omega}, a \rangle}{\langle \nabla U, a \rangle}. \tag{15}$$

Equation (15) asserts that the limit on the left hand side exists, hence the function $g_\Omega(\cdot)$ is differentiable. Moreover, the derivative is continuous. Since the left hand side of (15) does not depend on a, so is the right hand side. Therefore, $\nabla f_{s,\Omega}$ is proportional to ∇U with the coefficient of proportionality $g'_\Omega(U(x))$.

Q. E. D.

Next, I am going to characterize the PM adjustment processes. As we will see, the payoff monotonicity implies some connection between the deterministic part of the adjustment process, μ, and its stochastic part, Γ. As a corollary, I will prove that any payoff monotone adjustment process is locally improving. The reverse, however, is not true. Moreover, I will describe a broad class of locally improving processes such that a generic process of this class is not payoff monotone.

Lemma 3 *Consider Markov process* (μ, Γ) *and assume that for* $\forall x \in \Omega$

$$(\mu(x) = 0) \Leftrightarrow (\nabla U(x) = 0). \tag{16}$$

The process is payoff monotone if and only if there exists a continuous func-

tion $c : U(\Omega) \rightarrow R_+/\{0\}$ *such that*

$$\mu = c(U)\Gamma\nabla U. \tag{17}$$

Moreover, if the Hessian of U *has full rank* $c(\cdot)$ *is differentiable.*

Condition (16) states that is there is no deterministic adjustment at the critical points of the utility function. That is, all such points would be steady states of the deterministic dynamics.

Proof. Suppose (17) holds. Define

$$\xi(z) = 2\int_0^z c(y)dy. \tag{18}$$

Then it is easy to check that

$$f_{s,\Omega}(x) = \frac{\exp(\xi(U(x)))}{\int_\Omega \exp(\xi(U(y)))dy} \tag{19}$$

is a normalized stationary solution of (5)-(6). According to Lemma 1, it is the unique normalized stationary solution. According to (18), $\xi'(\cdot) = c(\cdot) > 0$. Hence,

$$(f_{s,\Omega}(x_1) \geq f_{s,\Omega}(x_2)) \Leftrightarrow (U(x_1) \geq U(x_2)) \tag{20}$$

so the adjustment process is payoff monotone.

Now, suppose that the adjustment process is payoff monotone. Then, according to Lemma 2, there exists a continuously differentiable strictly increasing function $g_\Omega : R \to R_+/\{0\}$ such that

$$f_{s,\Omega}(x) = g_\Omega(U(x)). \tag{21}$$

Define vector

$$\kappa = g_\Omega(U(x))\mu(x) - \frac{g'_\Omega(U(x))}{2}\Gamma\nabla U(x). \tag{22}$$

Then (5)-(6) implies that vector κ satisfies

$$div\kappa = 0 \text{ on } \Omega, \tag{23}$$

$$\langle\kappa, n\rangle = 0 \text{ on } \partial\Omega. \tag{24}$$

Moreover, definition of payoff monotonicity implies that (23)-(24) should hold for any Ω. Hence, $\kappa = 0$. Now (22) implies that

$$\frac{g'_\Omega(U(x))}{2g_\Omega(U(x))} = \frac{\langle\mu(x), \nabla U(x)\rangle}{\langle\nabla U(x), \Gamma\nabla U(x)\rangle} \tag{25}$$

provided that $\nabla U(x) \neq 0$. The right hand side of (24) does not depend on Ω, hence the left hand side should also not depend on Ω. Introduce

$$c(z) = \frac{g'_\Omega(z)}{2g_\Omega(z)}. \tag{26}$$

Since set U_C is finite, for any $z \in U(\Omega)$ the exists $x \in \Omega$ such that $U(x) = z$ and $\nabla U(x) \neq 0$. Hence, $c(\cdot)$ is defined on $U(\Omega)$. According to Lemma 2, $c(\cdot) \geq 0$ and according to (16) and (25) $c(\cdot) \neq 0$. Hence, $c(\cdot) > 0$. Finally, putting $\kappa = 0$ in (22) and using the definition of $c(\cdot)$ we get

$$\mu(x) = c(U(x))\Gamma\nabla U(x). \tag{27}$$

Proof of differentiability of $c(\cdot)$ is similar to the proof of differentiability of $g_\Omega(\cdot)$ in Lemma 2 and is omitted.

Q.E.D.

An easy corollary of Lemma 3 is that any payoff monotone process is locally improving.

Corollary 1 *If Markov process* (μ, Γ) *is payoff monotone and (16) holds it is locally improving.*

Proof. According to Lemma 3, we can write

$$\mu(x) = c(U(x))\Gamma\nabla U(x). \tag{28}$$

for some positive real valued function $c(\cdot)$. Therefore,

$$\langle \mu(x), \nabla U(x)\rangle = c(U(x))\langle \nabla U(x), \Gamma\nabla U(x)\rangle \geq 0. \tag{29}$$

Hence, the process is locally improving.

Q. E. D.

This completes the proof of Theorem 1. The reverse to the Corollary 1, however, is not true. Indeed, consider the following example.

Example 1. *Assume that the choices made by the individual follow the stochastic process:*

$$dx = \nabla U(x)dt + \Lambda dW. \tag{30}$$

Here $U(\cdot)$ *is twice continuously differentiable function, which is interpreted as a utility function of the individual,* Λ *is* $n \times n$ *matrix of full rank and* $W = (W_1, ..., W_n)$ *is a vector of independent standard Wiener processes. Note* *that the probability density of choices generated by process (30) is governed by (5)-(6) with and* μ *and* Γ *given by*

$$\begin{aligned} \mu(x) &= \nabla U(x) \\ \Gamma &= \Lambda^T \Lambda. \end{aligned}$$

Note that (16) trivially holds for this process. The first term in (30) corresponds to the gradient dynamics and says that the individuals adjust their choices in the direction of the maximal increase of their utility. The second term states that this adjustment is subject to a random error or experimentation. These errors are uncorrelated in time, though correlation among different components of x *is permitted and is given by the matrix* $\Gamma = \Lambda^T \Lambda$.

Let us assume $n = 2$, *and put*

$$\begin{aligned} U(x_1, x_2) &= u_1(x_1) + u_2(x_2), \\ \Lambda &= \begin{matrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{matrix} . \end{aligned}$$

Then

$$f_{s,\Omega}(x) = C_\Omega \exp(\frac{u_1(x_1)}{\sigma_1^2} + \frac{u_2(x_2)}{\sigma_2^2}).$$

Consider two choice vectors $x = (x_1, x_2)$ *and* $y = (y_1, y_2)$*. Assume* $u(x_1) = u(x_2) = 5$, $u(y_1) = 4$, $u(y_2) = 8$, $\sigma_1^2 = 1$, $\sigma_2^2 = 10$*. Then* $u(x) < u(y)$ *but* $f(x) > f(y)$*.*

Part 1 of Theorem 2 claims that the situation illustrated by Example 1 is quite generic. To formalize this idea, consider a class PD of locally improving Markov adjustment processes (μ_B, Γ). By the definition of a PD process we can write

$$\mu_B = B \nabla U$$

for some positive definite matrix B with constant coefficients. (Matrix B is called positive definite if for any $z \in R^n/\{0\}$ $\langle z, Bz \rangle > 0$. Note that matrix B is not required to be symmetric.) Such an adjustment process can be considered to be an element of $R^{n^2+n(n+1)/2}$ (n^2 is the number of independent elements in matrix B, while $n(n+1)/2$ is the number of the independent elements in the symmetric matrix Γ). Endowing $R^{n^2+n(n+1)/2}$ with Lebesque measure we can make the class PD a measure space.

Our next goal is to prove that for $n > 1$ PM meets PD by a submanifold of a lower dimension. Which means that a generic PD process is not PM, where genericity is understood is both topological and measure theoretic sense. First, we need to prove a technical result. To formulate it, let $x_0 \in \Omega$

and

$$I = \{x \in \Omega : U(x) = U(x_0)\} \tag{31}$$

be the indifference surface passing through x_0.

Lemma 4 *Let the Hessian of utility be non-degenerate Then there exist n different points $x_1, .., x_n \in I$ such that vectors $\nabla U(x_1), ..., \nabla U(x_n)$ are linearly independent.*

Proof. First observe that if n vectors $b_1, .., b_n$ ($b_i \in R^n$) are linearly independent then $\exists \delta > 0$ such that for any $\varepsilon_1, ..., \varepsilon_n$ ($\varepsilon_i \in R^n$, $\|\varepsilon_i\| < \delta$) vectors $b_i + \varepsilon_i$ are also linearly independent (this follows from the fact that the determinant of the matrix formed by n vectors in R^n continuously depends on its columns). Since the Hessian of U has full rank the indifference surface is not a hyperplane, therefore for $\forall \varepsilon > 0$ there exist n different points $x_1, .., x_n \in I$ such that vectors $(x_i - x_0)$ are linearly independent and$\|x_i - x_0\| < \delta$. Using the full rank assumption again one concludes that vectors b_i defined by

$$b_i = D^2U(x_0) \cdot (x_i - x_0) + \nabla U(x_0)$$

are linearly independent. But

$$\nabla U(x_i) = b_i + o(\varepsilon).$$

Hence, according to the observation made in the start of the proof one can choose δ small enough to ensure that $\nabla U(x_i)$ are linearly independent.

Q. E.D.

Now we are ready to prove the following Lemma.

Lemma 5 *Assume $n > 1$ and the Hessian of utility has full rank. Then $PD \cap PM$ can be embedded in PM as a submanifold of a lower dimension. In particular, this implies that the Lebesque measure of the PM processes in class PD is zero and that the set of the PM processes is nowhere dense in PD.*

Proof. According to Lemma 3, for each payoff monotone process in class PD we can write

$$B\nabla U(x) = c(U(x))\Gamma\nabla U(x). \tag{32}$$

for some positive real valued function $c(\cdot)$. Fix $x_0 \in \Omega$ and let $x_1, .., x_n \in I$ be such that $\nabla U(x_1), ..., \nabla U(x_n)$ are linearly independent. Such $x_1, .., x_n$ exist by Lemma 4. Let us introduce the following notation

$$\begin{aligned} U(x_0) &= U,\ U_{kj} = \frac{\partial U(x_k)}{\partial x_j} && (33) \\ y^i_j &= b_{ij} - c(U)\gamma_{ij}, && (34) \end{aligned}$$

where b_{ij} and γ_{ij} are matrix elements of matrices B and Γ respectively. Then for a fixed i

$$\sum_{j=1}^{n} U_{kj} y^i_j = 0. \qquad (35)$$

Since $\nabla U(x_1), ..., \nabla U(x_n)$ are linearly independent, the unique solution of (36) is $y^i_j = 0$. Since this is true for every i, (35) implies

$$B = c(U)\Gamma. \qquad (36)$$

Since both B and Γ are constant matrices $c(U) = c > 0$ is also a constant. This means that the set of payoff monotone processes is given by $B = c\Gamma$, which is a smooth manifold of dimension $1 + n(n+1)/2 < n^2 + n(n+1)/2$, provided $n > 1$. (A point on the manifold can be uniquely determined by $n(n+1)/2$ elements of matrix Γ and c). Therefore, this set is nowhere dense in PD and has Lebesque measure zero.

Q.E.D.

This completes the proof of part 1 of Theorem 2. The proof of part 2 is almost verbatim the same and is omitted. Now let us return to Example 1. An interesting corollary of Lemma 4 given by the following Proposition.

Proposition 1 *Assume the Hessian of the utility has a full rank. Process (30) is payoff monotone if and only if* $\Gamma = \sigma^2 I$*, where* I *is the identity matrix.*

Proof. According to Lemma 3, process (30) is payoff monotone if and only if for any $x \in \Omega$ vector $\nabla U(x)$ is an eigenvector of matrix Γ. Moreover, eigenvalue corresponding to this eigenvector depends on x only through the utility level $U(x)$. Lemma 4 implies that a symmetric matrix Γ has n linear independent eigenvalues corresponding to the same eigenvalue, hence $\Gamma = \sigma^2 I$. If $\Gamma = \sigma^2 I$ then the steady state is given by

$$f_{s,\Omega} = \frac{\exp(\frac{2U(x)}{\sigma^2})}{\int\limits_{\Omega} \exp(\frac{2U(y)}{\sigma^2})dy}.$$

It is, clearly, payoff monotone.

Q.E.D.

Recall that a Markov adjustment process satisfies the IIA property if the ratio of the probability that the choice is in an $\varepsilon-$ ball centered at the point x_1 to the probability that it is in an $\varepsilon-$ ball centered at point x_2 does not depend on whether some other choice z is available, up to the order $o(\varepsilon)$. The following result holds.

Lemma 6 *A Markov adjustment process satisfies IIA if and only if the Jacobi matrix of the vector field* $\Gamma^{-1}\mu(x)$, $D(\Gamma^{-1}\mu(x))$, *is symmetric for* $\forall x \in \Omega$.

Proof. Let us introduce vector j by the formula:

$$j(x) = -\nabla U(x) f(x) + \frac{1}{2}\Gamma \nabla f(x). \tag{37}$$

Then, in the steady state $j(x)$ should solve the following boundary problem

$$div(j(x)) = 0 \tag{38}$$

$$\langle j(x), n(x) \rangle = 0 \text{ on } x \in \partial\Omega. \tag{39}$$

The distribution f is then determined by the system of first-order partial differential equations:

$$j(x) = -\mu(x) f(x) + \frac{1}{2}\Gamma \nabla f(x). \tag{40}$$

The IIA property implies that a change in Ω will result in multiplication of f, and hence of j, by a constant, that is $j_{new} = Cj_{old}$. This relation should hold at each point, which belongs to the intersection of the new and the old choice sets. Hence j_{new} should solve the same boundary problem, but on a different domain. The only vector j that would solve (39)-(40) for any domain is $j = 0$. Hence IIA, together with the definition of j, implies that the steady state density $f(x)$ solves the system

$$\mu(x)f(x) - \frac{1}{2}\Gamma\nabla f(x) = 0. \tag{41}$$

or

$$\frac{1}{2}\nabla \ln f(x) = \Gamma^{-1}\mu(x). \tag{42}$$

The Jacobi matrix of the left hand side of (43) is the Hessian matrix of $\ln f(x)$. Since, according to Lemma 1, $f(x)$ is positive and twice continuously differentiable this matrix is symmetric, so the Jacobi matrix of the right hand side should also be symmetric.

To prove the reverse, assume that the Jacobi matrix of $\Gamma^{-1}\mu(x)$ is symmetric and define $f(x)$ to be the solution of (43). According to the Frobenuous theorem, the solution exists and is unique up to a multiplicative constant.

It is easy to see that $f(\cdot)$ such defined solves (5)-(6). The constant is chosen from the normalization condition.

Q. E. D.

Lemma 6 shows that IIA requires some connection between the deterministic and stochastic part of Markov process to hold. This connection does not have any *a priori* economic justification and we should not expect it to hold in general. Moreover, as I will show below, IIA does not hold for a generic Markov process from some broad class of the payoff monotone processes. To see this consider a class PDS of Markov adjustment processes, which is obtained from PD assuming that B is symmetric. I can be naturally embedded in $R^{n(n+1)}$. Let us endow this set with Lebesque measure. Then the following result holds.

Lemma 7 *Assume $n > 1$ and for any non-degenerate constant matrix C there exist i, k such that*

$$\frac{\partial^2 U}{\partial x_i' \partial x_k'} \neq 0, \tag{43}$$

where

$$x' = Cx. \tag{44}$$

Then $PDS \cap IIA$ can be embedded into PDS as a submanifold of lower dimen-

sion. In particular, this implies that the Lebesque measure of the IIA processes in class PD is zero and that the set of the IIA processes is nowhere dense in PD.

Proof. Since both matrices Γ^{-1} and B are positive definite, there exists a non-degenerate constant matrix C such that both $C^T\Gamma^{-1}C$ and C^TBC are diagonal, with all diagonal entries strictly positive (Gantmakher, 1989). Let us denote the i^{th} diagonal element of Γ^{-1} as $1/\sigma_i^2$ and the i^{th} diagonal element of B as b_i. Let $x' = Cx$. Then, according to Lemma 5, the process satisfies IIA if and only if

$$(\frac{b_i}{\sigma_i^2} - \frac{b_k}{\sigma_k^2})\frac{\partial^2 U}{\partial x_i' \partial x_k'} = 0. \tag{45}$$

Now (46) and (44) imply that there exists a pair of indices i, k such that

$$\frac{b_i}{\sigma_i^2} = \frac{b_k}{\sigma_k^2}. \tag{46}$$

This means that the the set of processes for which IIA holds is a smooth manifold with dimension at least by one smaller then $n(n+1)$ and therefore, the set of such processes has Lebesque measure zero and is nowhere dense in *PDS.*

Q.E.D.

This completes the proof one part 5 of Theorem 2. As one can see from the proof, assumptions (44)-(45) can be weakened. Indeed, it is sufficient to require them to hold only for C that brings Γ^{-1} and B to a diagonal form, rather than for any non-degenerate C. Economically, assumption (44) says that the utility is not additively separable in the components of vector x'. If it is separable IIA will hold for any process in PDS.

Results obtained so far show that both the payoff monotonicity and IIA do not hold for a generic locally improving processes. However, the payoff monotonicity is strictly stronger assumption then IIA. Indeed, the following result holds.

Lemma 8 *Assume the Hessian of the utility function is non-degenerate. Then $PM \subset IIA$.*

Proof. According to Lemma 3 payoff monotonicity implies that

$$\mu(x) = c(U(x))\Gamma\nabla U(x) \tag{47}$$

for some differentiable function $c : R \rightarrow R_+$. But then the matrix element

$D(\Gamma^{-1}\mu)_{ij}$ is given by

$$\frac{\partial(\Gamma^{-1}\mu)_i}{\partial x_j} = c'(U(x))\frac{\partial U}{\partial x_i}\frac{\partial U}{\partial x_j} + c(U(x))\frac{\partial^2 U}{\partial x_i \partial x_j}. \tag{48}$$

Hence, the matrix is symmetric and the process satisfies IIA.

Q.E.D.

This completes the proof of part 3 of Theorem 2. The following lemma completes the proof of Theorem 2.

Lemma 9 *Assume that $n > 1$ and Hessian of the utility has a full rank. Then PDS $\cap$ PM can be embedded in PDS $\cap$ IIA as a submanifold of a lower dimension.*

Proof. Following the same logic as in the proof of Lemma 5, it is easy to see the set $PM \cap PDS$ has dimension $1 + n(n+1)/2$. On the other hand, as one can deduce from the proof of Lemma 6, set $IIA \cap PDS$ has dimension at least $n(n+1) - (n-1) = n^2 + 1$. Lemma 8 implies that $PM \cap PDS \subset IIA \cap PDS$. Therefore, $PM \cap PDS$ can be embedded into $IIA \cap PDS$ as a submanifold of a lower dimension.

Q. E. D.

To conclude, I have shown that for a sufficiently broad class of LI processes a generic process does not satisfy IIA and a generic process that satisfies IIA is not PM. Note also that an IIA process need not be LI.

6 DISCUSSION AND CONCLUSIONS

Two kinds of theories of boundedly rational behavior are possible. Static theories focus on stationary behavior and do not include any explicit mechanism for temporal change. As in rational choice theory, they embody something of a subject's cognitive analysis of the choice problem. Dynamic theories, on the other hand, explicitly model the fine-grain adjustments made by the subjects in response to their recent experiences.

Both types of theories originated in mathematical psychology. Static theories, first considered by Luce (1959), were based on the axiomatic approach to the characterization of the choice probabilities. Dynamic learning models where pioneered by Bush and Mosteller (1955). In these models learning is modelled as a Markov process on the choice set.

The main contribution of this paper is to argue that the axioms of the static approach are not supported by a generic dynamic adjustment procedure. Therefore, when studying boundedly rational behavior, it would be desirable to start with explicit formulation of the learning process.

In the Introduction I mentioned some applied papers that used the static approach. It would be interesting to study to what extend the results of these papers are robust to explicit dynamic modelling.

REFERENCES

Alchian, A. A.: Uncertainty, evolution, and economic theory. Journal of Political Economy, **LVIII**, 211-221 (1950).

Anderson S. P., Goeree, J. K., Holt C. A.: Rent seeking with bounded rationality: An analysis of all-pay auction. Journal of Political Economy, **106**, 828-853 (1998).

Anderson S. P., Goeree, J. K., Holt C. A.: Stochastic game theory: adjustment to equilibrium under bounded rationality. Working Paper, University of Virginia, (1999).

Anderson S. P., Goeree, J. K., Holt C. A.: Minimum-effort coordination games: stochastic potential and logit equilibrium. Games and Economic Behavior, **34**, 177-199 (2001).

Arkes, H. R., Hammond, K. R.: Judgment and decision making: an interdisciplinary reader. Cambridge: Cambridge U. Press, 1986.

Arrow, K. J. and L. Hurwicz. : Stability of gradient process in n-person games, Journal of Society of Industrial and Applied Mathematics, **8**, 280-94 (1960).

Basov, S.: Bounded rationality, reciprocity, and their economic consequences.

Ph.D. Thesis, The Graduate School of Arts and Sciences, Boston University 2001.

Bush R.: and Mosteller, F.: Stochastic models for learning. New York: Wiley 1955.

Chen, H. C., Friedman, J. W., Thisse, J. F.: Boundedly rational Nash equilibrium: a probabilistic choice approach, Games and Economic Behavior, 18, 32-54 (1997).

Conlisk, J.: Why bounded Rationality? Journal of Economic Literature, **XXXIV**, 669-700, (1996).

Estes, W. K.: Towards statistical theory of learning. Psychological Review, **5**, 94-107 (1950).

Foster, D., Young, P.: Stochastic evolutionary game theory. Theoretical Population Biology, **38**, 219-232 (1990).

Friedman, D., Yellin, J.: Evolving landscapes for population games, University of California Santa Cruz, mimeo (1997).

Friedman, D.: The evolutionary game model of financial markets, Quantitative Finance, **1**, 177-185, (2000).

Fudenberg, D., Harris, C.: Evolutionary dynamics with aggregate shocks. Journal of Economic Theory, **57**, 420-41 (1992).

Gantmakher, F. R: The theory of matrices. New York : Chelsea, 1989.

Hogarth, R.: Judgment and choice: Psychology of decision. New York: Wiley, 1980.

Ito, S.: Diffusion equation. American Mathematical Society, 1992.

Kagel, J. H., Roth, A. E.: Handbook of experimental economics. Princeton: Princeton University Press 1995.

Kahneman, D., Slovic P., Tversky A.: Judgment under uncertainty: heuristic and biases. Cambridge: Cambridge U. Press 1982.

Kanan, D.: An Introduction to stochastic processes. North Holland, Elsevier Inc. 1979.

Kandori, M., Mailath, G., Rob, R.: Learning, mutation and long run equilibria in games. Econometrica, **61**, 29-56, (1993).

Luce, R. D.: Individual choice behavior, New York: Wiley 1959.

McKelvey, R. D., Palfrey, T. R.: Quantal response equilibria for normal form games. Games and Economic Behavior, **10**, 6-38, (1995).

McKelvey, R. D., Palfrey, T. R.: Quantal response equilibria for extensive form games. Experimental Economics, **1**, 9-41, (1998).

Nelson, R. R., Winter, S. G.: An evolutionary theory of economic change, Cambridge: Harvard University Press 1982.

Nisbett, R., Ross L.: Human inference: Strategies and shortcomings in the social judgment. Englewood Cliffs: Prentice-Hall 1980.

Offerman, T., Schram, A., Sonnemans, J.: Quantal response models in step-level public good games. European Journal of Political Economy, **14**, 89-100, (1998).

Oxtoby, J.C.: Measure and category. New York, Springer-Verlag, 1980.

Payne, J. W., Bettman, J. R., Johnson E. J.: Behavioral decision research: a constructive processing perspective. Annual Review of Psychology **43**, 87-131 (1992).

Rogers L. C. G., and Williams, D.: Diffusions, Markov processes and martingales. New York: Wiley 1994.

Pitz, G., Sachs N. J.: Judgment and decision: Theory and application. Annual Review of Psychology, **35**,139-63 (1984).

Simon, H. A:. Administrative behavior; a study of decision-making processes in administrative organization. New York: Macmillan 1957.

Young, P.: The evolution of conventions, Econometrica, **61**, 57-84, (1993).