Sophisticated Players and Sophisticated Agents

Rustichini, A.

Publication date:
1998

Sophisticated Players and Sophisticated Agents

A. Rustichini ¤
CentER, Tilburg University
Warandelaan 2, P.O. Box 90153
5000 LE Tilburg
The Netherlands
email: aldo@kub.nl

August 1998

Abstract

A sophisticated player is an individual who takes the action of the opponents, in a strategic situation, as determined by decision of rational opponents, and acts accordingly. A sophisticated agent is rational in the choice of his action, but ignores the fact that he is part of a strategic situation.

We discuss a notion of equilibrium with sophisticated agents, we provide conditions for its existence, and argue that it di®ers systematically from the Nash equilibrium.

Keywords: Procedural Rationality, Sophisticated Agents.

JEL Classi¯cation: D81, D83.

1 Introduction

Many of the situations that real-life players face are complex. A situation may be complex, for example, because the number of players involved is very large. Or it may be di±cult because the game they are facing is relatively new to them, and they have not yet had the time to understand all the implications of each choice they can take. Finally this may happen because the game involves

¤I thank Eric van Damme, Ramon Marimon, Jean-Francois Mertens, Ariel Rubinstein, Harald Uhlig for very useful discussions on the topic of this paper, and an anonymous referee for very useful comments.

elements that they know little, or not at all. In this paper we have in mind in particular the first source of complexity.

A player who is facing such a complex strategic situation is facing a double order of difficulty. First, he has to consider how the action of the others is affecting his payoff for each action he takes. After he has done that, he has the problem of any decision-maker taken in isolation, of deciding what the best action is.

Let us focus on the first difficulty. If the number of players is large, then the attempt to construct a complete model of the situation may be prohibitive. The player should have a good idea of the set of actions available to all the others, of their preferences over outcomes, and of the strategy they adopt. If he does not have this information, then he should formulate an even more complex notion, a probability distribution over the different possibilities. Only after they have done this, he can proceed to the next step of deciding his best choice.

Introspection and anecdotal evidence suggest that real players behave differently. To this rather flimsy evidence recently a large body of research from experimental economics has added a more cogent support. The experimental literature on learning in games has grown very large in the last ten years, and has discussed an almost as large number of issues (for a review of this evidence, see [7]).

One particular aspect of this research interests us directly here. How do people adjust their learning procedures to the complexity of the problem they are facing? This question is explicitly addressed in some of the work in this literature (see for example [16], [17], [18], [24]). Many interpretations of the results, of theoretical explanations, have been presented and suggested.

Overall, the evidence collected in these works suggests the following working hypothesis. People in complex strategic situations try to simplify their decision task. They do this by first ignoring the fact that the consequences of their actions are influenced by other players. Rather, they focus on finding the best possible decision as if the environment that they are facing was produced by an exogenous process, rather than by intelligent agents.

The problem that players face in the present paper is complex because the game has a large number of players. There is little direct experimental evidence on this specific aspect (which is discussed in [19]), probably for the obvious reasons that large experiments are harder to organize. We will work on the assumption that the same results extend to this specific reason of complexity.

We may say that someone is a sophisticated player if he keeps fully into account the strategic nature of the situation. A sophisticated player does in fact build the complex, complete model of the entire game. He considers explicitly the actions of the others as produced by agents like him, who go through the same thought process as he does. On the other hand, we may say that someone is a sophisticated agent if he satisfies the more modest requirement of adopting

reasonable selection criteria of his procedures.

This distinction in economics is more than a century old. In walrasian equilibrium, consumers and firms are sophisticated agents (they maximize utilities and profits), but are not sophisticated players (because they take prices as given, when they might change them with their actions).

In fact, when the number of agents is finite, the walrasian equilibrium is not the Nash equilibrium of the economy. Economists know from the work of Aumann (see [3]) that the walrasian equilibrium may be, in a special case, the correct solution concept even in an economy where individuals are sophisticated players. It is enough that players are negligible. In this case, even if they think of it, their strategic behavior has no consequence and price-taking behavior is perfectly rational (or, in our terminology, sophisticated).

Economists however also accept and widely use the concept of walrasian equilibrium in models with a finite number of agents. A possible explanation for such a willingness to use a concept, which is not appropriate, might be that economists think the concept is appropriate for a different reason. Namely, the effect of strategic behavior is so small, compared to the complexity of the analysis needed to compute it, that it is reasonable for the agents in the economy use their limited resources for more productive ends.

In this paper we suggest that in complex situations people behave as naïve players and sophisticated agents, precisely as they do in the walrasian equilibrium. We then discuss the implications of this hypothes. Note that both aspects of the assumption that we make on the way people behave are important. Being a naïve player is an important weakening of the rationality requirement, but on the other hand being a sophisticated player is a strong restriction.

A difficulty in this project is the fact that a good, widely accepted concept of sophisticated player is still missing. We shall try to argue in the conclusion that to define a good concept of sophisticated player (and of the restriction that this imposes on his behavior) is still an open question, which is currently actively pursued. But it is time to be more specific.

A Simple Model

These concepts are discussed extensively in the recent paper by Osborne and Rubinstein ([23]). They may be well illustrated using a simple example of that paper that we recall here.

You enter a game played in a large society of players like you, and play for infinitely many periods. In each period, you are randomly matched with another player chosen out of the society.

You get the following payoff for your choice of A or B and the choice a or b of your fellow player for that period:

	a	b

$$\begin{array}{ccc} A & 2 & 4 \\ B & 3 & 1 \end{array} \qquad (1.1)$$

You are a representative agent, so the other player is facing the mirror image of your same problem. For future reference, let us note immediately that the mixed strategy Nash equilibrium of this game has proportions $(:75; :25)$ on the two actions.

A very myopic procedure

Now suppose you adopt the following procedure to choose between A or B. You try A in the first period, then you try B in the second, and keep a record of the payoff after each choice. After these first two rounds of experimentation, you choose forever the action that gave the highest payoff in the first two rounds of experimentation.

Suppose now that the population you are facing is split between a proportion α^1 of people who play a forever, and a proportion $1 - \alpha^1$ that plays b forever. What is the probability that you will choose A as your preferred action after the first two rounds? Note that in these two periods you may be matched with a sequence of players of type $(a; b)$, $(b; a)$ or $(b; b)$, and in these cases you will eventually choose A). Or, you will be matched with a sequence $(a; a)$, in which case you will eventually choose B. The second event happens with probability $(\alpha^1)^2$, so the first happens with probability $1 - (\alpha^1)^2$.

Now define a steady state value of α^1, and denote it by α^{1*}, as the value of α^1 that satisfies the following property. α^{1*}, which is the proportion of players using a, is also equal to the probability that you select A when facing such a population. From our previous remarks one can see that in our example 1.1 this value of α^1 is the solution of the equation $\alpha^1 = 1 - (\alpha^1)^2$, which is approximately :62, a value different from :75, which is the proportion in the mixed strategy Nash equilibrium.

This result of course depends on the assumption that you try each action once and then you make up your mind forever. Is this procedure reasonable? Consider the case in which you have chosen A in the first try, and received the value 4, and then have played B in the second and received a value 3. Following the procedure, you should then choose, forever, the action A. But at the steady state value $\alpha^1 = :62$, which is less than 2=3, the action A has an average value strictly less than 3. It seems hard to believe that if you are a sophisticated agent you will be unendingly faithful to your previous choice. Even if you continue to ignore the strategic aspects of the game. You might want to try again the action B that produced a value of 3 in the only time he gave it a chance, larger than what you are getting now on average. If you do not, it seems fair to say that you are a naïve player, and a naïve agent.

The k-sampling procedure

Clearly, limiting the experimental phase to two periods is a very simple way of dealing with the problem. A more sophisticated agent may want to try both options several times before he makes up his mind. The critical question is: \How many times?" A natural way of answering the question seems the following, also introduced in Osborne and Rubinstein ([23]).

Let us de¯ne a new procedure, call it k-procedure, which is very similar to the one described previously, with one di®erence: you try A and then B alternatively for k periods. Earlier we have seen an example of a 1-procedure. The steady state value of $®^{1¤}$ can be de¯ned in a similar way, and will in general depend on k. Osborne and Rubinstein show that as k tends to in¯nity, the corresponding value of $®^1$ tends to :75 which is the value of the Nash equilibrium.

Tentative conclusions

To summarize our discussion so far, we may try to draw the following two conclusions.

i. A society of naÅve players who are also naÅve agents converges to a distribution of actions that is di®erent from the distribution obtained at a Nash equilibrium. This is what happens in the case of the 1 procedure.

ii. As players become more and more sophisticated decision-makers, then the steady state distribution becomes more and more similar to the Nash equilibrium distribution. This is what happens with the k-procedure.

We want to argue here that while the ¯rst conclusion is true, the second is false. In addition we would argue that neither the 1 procedure nor the in¯nitely long k procedure are reasonable. We have seen previously why the 1 procedure is \ naÅve ". The k-procedure, for k large, is certainly more sophisticated than the 1-procedure. But it tells you to continue the experimentation no matter what evidence you have accumulated so far. For instance, even if one of the actions has persistently shown to be inferior to the others, you should keep trying it.

The procedure we are going to describe does keep into account the performance of an action along as the agent is experimenting. In a later section we even let agents use the optimal bayesian procedure. In both cases, we are going to show that there is a well-de¯ned equilibrium concept, which is the steady state of the process, and that the predictions of this equilibrium concept systematically deviate from the Nash equilibrium predictions.

Our argument involves, unfortunately, a relatively technical issue. We think this is a price worth payingy for two reasons. First, the issue has some general relevance. Second, we may trying to avoid it and let the mathematical di±culty

choose the model for us, by picking the one we can solve. But this would only constrain the theory to an unacceptably naïve level. Here is, in simple terms, what the technical issue is, and why we think it is of general importance.

A procedure that decides in finite, fixed time, a choice and settles down to it forever after is not likely to be very interesting. Typically, a procedure with a minimal degree of sophistication will require a period of experimentation which is not fixed, and depends on the history that the player has encountered. Suppose that we require that a choice is made, eventually. (If we decide not to make this restriction, the technical problem only gets worse). We now want to determine the steady state of a society where agents use this procedure. The steady state depends on the probability that any of the actions is chosen and a simple way to insure existence of a steady state is to insure that his probability is continuous on the process the player is facing. Here too, we may think of more general ways to insure existence but again the technical problem only gets worse.

The problem is that for procedures which are not too naïve the event where a specific action is chosen is a tail event: to decide if a point in the space belongs to it you have to look arbitrarily far in the future. Proving continuity of the probability of a tail event is relatively hard, and the lemma (5.3) in the appendix gives a way of doing it. The problem is general because it appears whenever we try to determine steady states of games where a population of players is experimenting with different strategies.

2 Naïve players, sophisticated agents

The game

Consider a game, which is a generalization of the situation described in the simple example. In this game, we have a large society, where players play for infinitely many periods against an opponent who has been randomly chosen out of the society. Each player has a set $I = \{1, \ldots, j, \ldots, n\}$ of actions he can take; the payoff when he takes the action i and the opponent the action j is $g^i(j)$.

We have already seen examples of procedures that a player may follow to decide the action to take in each period, the k-procedures. We now want to consider other possibilities. Here is a first example, that we are going to use to illustrate the main ideas. We are going to see a second example later.

Maximum average procedure

In a first experimental phase, each player tries each of the actions once, in a fixed sequence, and keeps a record of the payoffs he has received in this experimental phase. After that, he begins a second phase. Now before choosing

an action he computes the average payo® that each action has given so far. We de¯ne "average payo® to an action" to be the sum of the payo®s in the periods in which that action was chosen, divided by the number of these periods. Then he chooses for that period the action that gave so far the highest payo®, and randomizes uniformly to break ties. He continues this second phase forever. We call it the maximum average procedure.

This procedure may be considered similar to the in¯nite extension of the k-steps procedure: rather than sampling only ¯nitely many times for sure, you potentially sample the di®erent actions an in¯nite number of times. But the procedure has the additional °exibility of making the decision of sampling or not conditional on the previous results. Recall the possibility we have discussed earlier, that you have chosen A in the ¯rst try, and received the value 4, and then has played B in the second and received a value 3. If the procedure you are following tells you to stop experimenting and choose forever the action A, you get the steady state value $®^1 = :62$, where the action A has an average value strictly less than 3. The maximum average procedure would tell you to try B again.

We are now going to see that it gives a completely di®erent outcome.

Steady state values

The next element we need is a concept of equilibrium. We generalize the idea of the steady state value of $®^1$ presented in the simple example. Think what happens if the process the players are facing is really a process that chooses the action independently in each period, with a probability $® ´ (®^j)_{j2J}$ over the set J. The choice of actions by a player who is following the maximum average procedure will follow a rather complicated process. Let us assume, however, that eventually the action chosen settles down to a ¯xed action. The probability that an action j is eventually chosen clearly depends on ®. Formally, if we denote by c_t the choice of the agent at time t, the event where the choice eventually settles down to the action j is $flim_t c_t = jg$, and

$$P^®(flim_t c_t = jg)$$

is the probability of this event.

If we want a steady state to exist, our procedures need to satisfy two basic requirements. First, we need the player to choose an action eventually. Second, we need that the probability of an action being chosen to be continuous in the probability of the action of the opponent. More general concepts of equilibrium based on weaker conditions are certainly possible: but this seems a ¯rst acceptable step.

Choice procedures

A choice procedure is a sequence of functions that in each period decides the probability distribution on the next action, depending on the previous history of choices of the opponent. Formally, a choice procedure is

$$c \acute{} (c_1; : : : ; c_t; : : :); c_t : J^{t i 1} \; ! \; \¢(J) \text{ for every } t; \tag{2.2}$$

with c_1 a constant. We emphasize that the choice function maps from the history of actions of the opponent into the action, for that period, of the player. The maximum average procedure is clearly included in this de¯nition, as are the 1 and the k procedures.

Note that the function c_t is a complicated object, because it has built-in the record of the choices made previously by the player. Consider for instance the function c_t in the maximum average procedure applied to the simple example 1.1. Let c_1 = A, and $c_2(h)$ = B for sure after any history. Then $c_4(a; a; b)$ gives equal probability to A and B (because the two actions have given on that history the same average, namely 2, given the choices of the player in the ¯rst three periods), while $c_4(a; a; a)$ is B for sure.

We take the space of sequences of choices of the other player, $J^{\,1}$, as the probability space, endowed with the product ¾-¯eld. Every $® 2 \¢(J)$ induces a probability distribution $P^®$ on this space. This is our probability space. We denote:

$$C^j \acute{} f ! : c_t = j \text{ eventually } g \tag{2.3}$$

We take as part of the de¯nition of a choice procedure that it chooses an action eventually with probability one. That is we require for every procedure that

$$\sum_{j 2 J} P^®(f \lim_{t \, ! \, 1} c_t = j g) = 1; \text{ for every } ® 2 \¢(J): \tag{2.4}$$

We also say that a choice procedure is continuous if the probability of the choice of an action changes continuously with the probability $® 2 \¢(J)$, that is, formally, the function

$$® \; ! \; P^®(f \lim_{t \, ! \, 1} c_t = j g) \tag{2.5}$$

is continuous. As we have tried to argue earlier, these procedures are the natural candidates for the existence of equilibrium. A technical contribution of this paper is to determine conditions that give continuity of the procedures, and hence existence of equilibrium.

The issue is discussed in detail in lemma (5.3) in the appendix. There we give a precise formulation of the condition for continuity of choice procedures. In the following paragraphs, we discuss more informally the main ideas.

Continuity of choice procedures

Recall that, for a probability $\alpha \in \Delta(J)$, we have a probability P^{α} on the stochastic process of the actions of the opponent that the player is facing in each period. Now consider the more sophisticated procedure described earlier. In which in every period chooses the action that has given the highest average payo®. We claim that this procedure eventually chooses an action with probability one (so it is a choice procedure according to our de¯nition), and that the probability of the event where the player chooses a speci¯c action is continuous in ®. The claim is not easy to prove because this event is a tail event: it involves the behavior of the sequence of choices of the opponent far in the future.

It will turn out that it is critically important to de¯ne appropriately the distance between two sequence of choice of the opponent. The pointwise distance is a possible option. A precise de¯nition of this concept is recalled later (see (5.22)): what is important to know is that this distance gives relative more importance to the initial choices. This concept is too weak for our purposes: two sequences may be close in this distance and still have very di®erent averages, because the behavior at in¯nity may be very di®erent.

Other, stronger, distances are possible, but they may be too strong. To clarify this point an example may be useful. Consider, for $J = \{a; b\}$, the following two di®erent probabilities assigned to a: $\alpha = 0{:}5$ and $\bar{\alpha} = 0{:}499$. The two numbers α and $\bar{\alpha}$ are fairly close, and the probability of every event which is described by a ¯nite number of coordinates is close. But the measure on the in¯nite process induced by α is concentrated on the sequence with average equal to 0:5, while the one induced by $\bar{\alpha}$ is concentrated on those with average equal to 0:499. The two measures on the in¯nite process are supported on two disjoint sets. Even if we make $\bar{\alpha}$ arbitrarily close to ® we are not going to change this fact. An intermediate concept of distance is needed.

The key insight, due to Ornstein (see [20]), is that the two sequences with averages respectively 0:5 and 0:499 are close, if we measure their distance appropriately. More precisely, we may say that two sequences are close if the number of coordinates that we have to change to transform one sequence into the other is small. For instance, let us call a sequence an ®-sequence if the average number of a's in the sequence is ®. We only need to change 0:1 per cent of the coordinates to transform an ®-sequence into a $\bar{\alpha}$-sequence, and as $\bar{\alpha}$ gets closer to ® the percentage of changes tends to zero.

This concept of distance is formally introduced in the appendix (see (5.23). The basic property (that we prove in (5.38) of this distance is that it is su®-¯ciently strong to impose that two sequences which are close have averages which are close.

The continuity of the probability is related to the continuity of the function $c_1(!) \acute{} \lim_{t \to 1} c_t(!)$. Note that since c_1 is discretely valued, its continuity of

would imply that it is a constant function; so this concept is too strong. The lemma may be read to say that c_1 is continous but for a set of measure zero.

Proposition 2.1 i. the probability of eventually settling down to one of the actions is one; so almost surely the player does not oscillate forever between some of the actions;

ii. for every j, the probability $P^{\alpha}(\lim_t c_t = j)$ is a continuous function of α.

The function is continuous even at the point where we might find surprising that it is, namely at the points in which the expected payoff from two actions is equal. At such a point a small change in the probability α makes the expected payoff of one of the actions strictly larger than the value of the other. But the event we are considering looks at the limit behavior of the choice of actions. So a small advantage for one of the two actions might matter quite a lot, in term of the probability of eventually choosing that action. The proposition (2.1) shows, among other things, that it does not matter much. Of course this proposition also implies, by a standard application of Brouwer's fixed point theorem, that there exists a α^* such that

$$P^{\alpha^*}(\lim_t c_t = j) = \alpha^{*j} \tag{2.6}$$

for every j. This α^* is a steady state of the process, and we call it an equilibrium.

The equilibrium with sophisticated agents

We can now describe a steady state α^{1*} of the process where each player is following the maximum average procedure in the game 1.1. To find such α^{1*}, consider possible different values of α^1. When the value of α^1 is 1 almost all the players are choosing the action A: so a new player will end with probability 1 to the action B. At the other extreme, when α^1 is 0, a similar reasoning shows that the new player will converge with probability 1 to the action A. It is also easy to see that this function is decreasing. Since this function is also continuous by our proposition 2.1 there is some value, call it α^{1*}, for which the probability of converging to a is exactly α^{1*}.

Now we argue that this value cannot be the probability that players use the action A in a mixed strategy Nash equilibrium, which is :75. In fact when the opponent uses the mixed strategy Nash equilibrium the expected payoff of the two actions is the same; so the probability of converging to A is almost equal to the probability of converging to B, and different from :75, which therefore cannot be the fixed point of the equation 2.6. In other words, for the probability of converging to A to be equal to :75, higher than the probability of converging

to B, it must be that the action A gives a higher expected payo® than the action B: and this cannot happen at the mixed strategy Nash equilibrium.

An explicit solution for the value of $®^{1¤}$ seems very hard, even for the simple game above. Monte Carlo computations show it to be roughly :665.[1] This value lies between the value of the mixed strategy equilibrium (which is :75) and the value in the one step procedure (which is :62). In addition, this value depends on the details of the procedure chosen. Take for instance this procedure: each agent tries each of the actions twice, rather than once, and then behaves exactly as before. The steady state value of $®^{1¤}$ for these procedures is di®erent.

We conclude putting on record our results so far:

Theorem 2.2 There is an equilibrium when agents use the maximum average procedure. In the game 1.1, the proportion of agents playing the two strategies is di®erent from that in the mixed strategy Nash equilibrium.

3 NaÄve Players, Bayesian Agents

One may argue that the procedure described is not obviously a sophisticated procedure. A simple way to answer this objection, however, is to consider the procedure which is sophisticated par excellence: the optimal policy in a two armed bandit problem.

The bandits problem

To de¯ne the Bayesian procedure, we may assume that each player thinks of the two actions he can take (either A or B) as producing a payo® independently from each other. This is of course a counterfactual belief: the two arms are not independent, since the distribution on the choice of the opponent of a versus b is the same in both cases. In terms of the distinction we have introduced, the player is not a sophisticated player, since he ignores this fact. He is however a sophisticated agent, since he follows the optimal policy for the two armed bandit problem.

The payo® he will receive in each period when he chooses the action A is determined by the draw of an independent random variable with unknown distribution. Again, this is only true from his point of view: but we are not going to iterate this proviso). This distribution can be parameterized simply by the probability of getting the payo® associated with the pair of actions $(A; a)$ of himself and of the opponent. The payo® matrix is:

	a	b
A	$g^1(1)$	$g^1(2)$

[1] The Matlab program that does this is available upon request. I thank Harald Uhlig for showing me how to do it.

$$B \quad g^2(1) \quad g^2(2) \tag{3.7}$$

where we assume

$$g^1(2) > g^2(1) > g^1(1) > g^2(2) \tag{3.8}$$

as it is in the simple example. Each player discounts by a factor equal to 1=2.

Equilibria

The set of equilibria depends of course on the initial belief. When the initial prior for all players has a continuous density, everywhere positive, then there is a natural candidate for the long-run equilibrium of this game. Each player has a belief concentrated on the mixed strategy Nash equilibrium of the stage game, say $®_N^{1¤}$, and players do in fact play the two actions in the proportion $®_N^{¤1}$. The Nash equilibrium is not the only possible outcome, however, as we are going to see immediately.

Simple beliefs

To keep the structure simple, we consider the case in which the prior of each player has a very simple form. He thinks that the true probability of facing the choice a of the opponent, when he plays A (and therefore the probability of a payo® $g^1(1)$) is either $¯^A$ or $°^A$, two numbers in $[0;1]$. He has a similar belief in the case of the action B. So his prior at t on the distribution on the action of the opponent and hence on payo®s for the arm i has the form $p_t^i±_{¯^i} + (1 ¡ p_t^i)±_{°^i}$, for $i = A;B$, where $±_{¯^i}$ denotes the Dirac measure at $¯^i$, and p_t^i is the probability he gives to the true parameter being $¯^i$. The true probability on the set $fa;bg$ is denoted by ®, depends on the distribution of players, and is the same on both arms A and B.

At the end of each period, players update in a Bayesian fashion the belief they have, as follows. If we denote the logarithm of the odds ratio

$$r_t^i \acute{} \log(\frac{p_t^i}{1 ¡ p_t^i}); i = A;B$$

then the posterior odds ratio after a choice of i, is of course:

$$r_{t+1}^i = \log(\frac{¯^i}{°^i}) + r_t^i; \text{ with probability } ® \tag{3.9}$$

(when a is observed) and

$$r_{t+1}^i = \log(\frac{1 ¡ ¯^i}{1 ¡ °^i}) + r_t^i; \text{ with probability } 1 ¡ ® \tag{3.10}$$

(when b is observed). We denote the expected payo® as

$$E_{-i}g^i \acute{} \bar{}^i g^i(1) + (1 \text{ ¡ } \bar{}^i)g^i(2); i = A; B; \tag{3.11}$$

and similarly for $E_{\circ i}g^i$. To make the problem interesting we assume that one of the arms is better than the other arm for some of the possible beliefs. Speci¯cally we assume that the ¯ belief gives a higher expected value for both arms:

$$E_{-A}g^1 > E_{\circ A}g^1; E_{-B}g^2 > E_{\circ B}g^2: \tag{3.12}$$

Also we assume that at the belief $(p^A; p^B) = (1; 0)$ the A arm has an expected value higher than B, and that the reverse holds at $(p^A; p^B) = (0; 1)$:

$$E_{-A}g^1 > E_{\circ B}g^2; \; E_{-B}g^2 > E_{\circ A}g^1: \tag{3.13}$$

Optimal Policy

The optimal policy in this problem may be characterized by a function ¾ of the two values $(r^A; r^B)$, taking value in the set $fA; Bg$. This simply follows from the fact that the optimal policy is stationary. Also a complete description of the relevant state (the belief over the distribution of outcomes) is provided by the pair $(p^A; p^B)$ and therefore, equivalently, by the pair $(r^A; r^B)$.

Acually, thanks to the Gittins-Jones dynamic allocation indices result (see [12], [30], and [4]), more is known. There exist two functions $¿^i; i = A; B$ from R_+ to R, such that

$$¾(r^A; r^B) = \text{ argmax }_{i2fA;Bg} ¿^i(r^i): \tag{3.14}$$

The functions $¿^i$ are the dynamic allocation indices for the two arms: the Gittins Jones theorem says that the optimal policy consists in choosing in each period the arm that has the index with maximum value. Finally, by theorem 3.1 of Berry and Fristedt ([4], page 1095: the theorem is stated for Bernoulli processes, but the proof can be extended to our case), the functions $¿^A$ and $¿^B$ are increasing.

This follows because the dynamic allocation index is increasing in the distribution over the arm, when the distribution is given the ¯rst order stochastic dominance partial order, and the distribution is increasing in p^i and therefore in r^i because of the inequality 3.13.

This remark together with (3.14) implies

$$\text{if } ¾(r^A; r^B) = A \text{ and } \hat{r}^A > r^A; \text{ then } ¾(\hat{r}^A; r^B) = A; \tag{3.15}$$

and

$$\text{if } ¾(r^A; r^B) = B \text{ and } \hat{r}^B > r^B; \text{ then } ¾(r^A; \hat{r}^B) = B: \tag{3.16}$$

Also note that from (3.8) and (3.12) we can infer that

$$\bar{}^A < \circ^A; \; \bar{}^B > \circ^B: \tag{3.17}$$

A simple example

For concreteness, a simple numerical example may help to clarify. Let

$$g^1(1) = g^2(2) = 0; g^1(2) = g^2(1) = 1; \tag{3.18}$$

and

$$^{-A} = 0; \circ^A = 1=3; ^{-B} = 1; \circ^B = 1=3: \tag{3.19}$$

Then [2] the functions $¿^i$ of (3.14) are:

$$¿^A(r) = \frac{2 + 4e^r}{3 + 4e^r}; ¿^B(r) = \frac{1 + 5e^r}{3 + 5e^r};$$

and the optimal policy is

$$\text{choose B if and only if } r^B \ ¸ \ ¼(r^A); \tag{3.20}$$

where $¼(r)$ is the solution of $e^{¼(r)} = (8=5)e^r + 3=5$, so $¼$ is increasing, tends to $\log(3=5)$ as $r \ ! \ ¡1$, and to $r + \log(8=5)$ as $r \ ! \ +1$.

The limit choice

From the above characterization of the optimal policy we know how the process of choices of the agent will evolve over time. If the optimal choice in the ¯rst period is, say, A he will keep choosing A, and only update the probability p^A. In the meantime, the value of p^B_t is constant and equal to p^B_0 until, if ever, the value $p^A_{t_0}$ falls below a critical level, which depends only on p^B_0. At that point the process will start again, with $p^A_{t_0}$ playing the role that p^B_0 was playing in the previous phase. Recall that because of (3.17), the updating described by (3.9) and (3.10) implies that r^A increases after an observation of b and r^B after an observation of a. Let now $t_1; t_2; : : : ; t_i; : : :$ be the sequence of switching times, that is:

$$c_t \ 6= \ c_{t\,i\,1} \text{ if and only if } t = t_i \text{ for some } i:$$

It is clear that

$$\text{for every } i; \text{ if } c_{t_i} \ 6= \ j; \text{ then } p^j_{t_i} < p^j_{t_{i+2}}; \text{ and } r^j_{t_i} < r^j_{t_{i+2}}; \tag{3.21}$$

that is the probability assigned to the "good" belief ¯ can only decrease at the switch between policies. To see this, suppose that at t_i the player switches to B, that is $¿(r^B_{t_i}) > ¿(r^A_{t_i})$. As long as he uses A, r^A_t remains equal to $¿(r^A_{t_i})$,and so when he switches back to A it must be that $¿(r^A_{t_i}) > ¿(r^B_{t_{i+1}})$, hence $¿(r^B_{t_i}) > ¿(r^B_{t_{i+1}})$. An implication of this is that for every ® the choice of the player becomes constant in ¯nite time. We put this result on record as:

[2] For details, see the example 4.3, page 1100 of Berry and Fristedt [4], and also the example 6.1.1, page 143 of Berry and Fristedt [5].

Proposition 3.1 *There is an equilibrium when agents have simple beliefs.*

The proof is in section 5. The key idea is again the behavior of the probability that a player will eventually choose A as a function $f(\alpha)$. It is easy to see that in the extreme values $f(0) = 1$ and $f(1) = 0$. So we have only to show continuity of the function in the intermediate values. This again follows because the choice procedure is continuous.

4 Conclusions

We have presented and developed, for general games, a distinction between sophisticated players and sophisticated agents. In economic theory this distinction is at the basis of the concept of walrasian equilibrium for competitive economies, and hence at the basis of most of our working conceptual structure. We have argued that in a world where agents are naïve players but sophisticated agents there is also a consistent notion of equilibrium. This equilibrium can be shown to exist. The equilibrium is different from the Nash equilibrium of the corresponding game, even if the procedures that agents use are "reasonable".

It is clear that a fundamental question which is preliminary to this line of research is \Which procedures are reasonable?" or in our terminology, \Who is a sophisticated agent?" For competitive economies, utility maximization and profit maximization by price-taking individuals seems a widely accepted concept of sophistication. In the case of general games, the question is still open. But the question seems so fundamental that it should probably be addressed before or at least at the same time, as an attempt is made to conjecture a new solution concept.

A comparison with the theory of games among sophisticated players may be useful. This theory is in turn based on a complete theory of utility and decision-making of fully rational agents. For example, the first three section of the fundamental book of Von Neumann and Morgenstern ([29]) deal with utility theory, and then proceed, on that basis, to the development of the game theory.

Should the theory of players with limited rationality follow the same way? In particular should we require a theory of individual rationality for sophisticated agents that precedes the theory of rationality in games? Perhaps not: but it certainly seems necessary at least to check if the procedures that the agents in our model are supposed to follow are plausible.

Some elements in this direction already exist. Recently many attempts have appeared that try to provide a similar foundation for the decision theory of agents with limited rationality. Some of these attempts explicitly formulate, and discuss the idea that we have used here of \procedures". We may try here to quote a few, like [1], [2], [6], [8], [9], [10], [11], [14], [15], [26], [27], [28]: but

the literature is very large, and growing. Some of these contributions explicitly examine criteria for a sound de¯nition of \reasonable procedure".

Finally, we are not arguing here that we should maintain the assumption of fully rational agents in all cases. Instead, we are claiming that, ¯rst, it is important to keep a precise distinction between sophisticated players and agents, and second that we still have to give a precise content to the assumption of sophisticated agents.

5 Appendix

In the appendix we present the proof of the proposition (2.1). The proof involves techical notions that we recall brie°y. The entire appendix is self-contained, if the reader is prepared to believe some basic results.

The d distance

As we mentioned, the distance de¯ned in (5.23) below is a basic concept in the Ornstein theory (see [20] and [22]). We introduce two distances on J^1. The ¯rst:

$$D(!;!^0) \acute{} \sum_{fi:!_i \neq !^0_i g} \pm^i \tag{5.22}$$

with $\pm$ 1=2. This last restriction is not important, we only assume it to make the comparison with the second distance easier. The second:

$$d(!;!^0) \acute{} \sup_{n,1} \frac{1}{n}]fi : 1 \quad i \quad n; !_i \neq !^0_i g: \tag{5.23}$$

The topology induced by d is strictly stronger than the one induced by D. In fact, the inequality (5.24) below holds. To see this, consider the problem:

$$\max_{(!;!^0)} D(!;!^0)$$

subject to:

$$d(!;!^0) \quad a$$

which has for solution a pair $(!;!^0)$ which are di®erent at a period equal to the integer part of 1=a, for which the D distance is equal to

$$\sum_{i=1}^{1} \pm^{\frac{1}{a}t} = \frac{\pm^{\frac{1}{a}}}{1 \, ¡ \, \pm^{\frac{1}{a}}}$$

which is less than a. Hence:

$$D(!;!^0) \quad d(!;!^0) \text{ for every } !;!^0: \tag{5.24}$$

The D-convergence is the pointwise convergence, the d-convergence is strictly stronger. A function which is D-continuous is d-continous, and a set which is D-open is d-open, but not viceversa.

A good feeling for the di®erence between the two distances may be given by the following observation, that will also be useful in the following. De¯ne for every j, t and ! the frequency of j as

$$f^j_t(!) \acute{}]fi : 1 \quad i \quad t; !_i = jg \tag{5.25}$$

If $!$ and $!^0$ are close in the D-distance, the two frequencies are not necessarily close. However, it is easy to see that:

$$\begin{aligned} j\,]fi:1 \quad i \quad t;!_i = jg \;\; i \;\;]fi:1 \quad i \quad t;!_i^0 = jg\,j \\ j \sum_{i=1}^{\infty} 1_{fjg}(!_i) \;\; i \sum_{i=1}^{\infty} 1_{fjg}(!_i^0)\,j \\ \sum_{i=1}^{\infty} j\,1_{fjg}(!_i) \;\; i \;\; 1_{fjg}(!_i^0)\,j \\ = \quad]fi:1 \quad i \quad t;!_i \neq !_i^0g \end{aligned} \tag{5.26}$$

and therefore:

$$j\,f_t^j(!) \;\; i \;\; f_t^j(!^0\,j \quad d(!;!^0): \tag{5.27}$$

For these two metrics on J^1 we can de¯ne two di®erent criteria of weak convergence of measures on J^1. We say that a sequence of measures $f¸_kg_{k=1;:::}$ on J^1 converges D-weakly to $¸$ if and only if $\int_{J^1} f(!)¸_k(d!)$ converges to $\int_{J^1} f(!)¸(d!)$ for all the D-continuous functions, and similarly for the d-weak convergence. A sequence of measures which d-weakly converges also D-weakly converges.

Finitely Determined Processes

A Bernoulli process on J^1 is the stochastic process induced by a sequence of i.i.d draws. So for every $® 2 ¢(J)$ there is an associated Bernoulli process, with measure $P^®$.

For any measure $¸$ over J^1 we denote by $h(¸)$ the entropy of $¸$. For a precise de¯nition of entropy see for example Petersen ([25]), chapter 5. The only fact we need here is that the entropy of a Bernoulli process induced by $® 2 ¢(J)$ is

$$i \sum_{j2J} ®_j \log ®_j \tag{5.28}$$

(see Petersen, ([25]), example 3.4, page 245). The key concept, introduced by Ornstein, is the following:

De¯nition 5.1 A measure $¸$ on J^1 is ¯nitely determined if for any sequence $f¸_kg_{k=1;:::}$ the two conditions

i. $¸_k$ D-weakly converges to $¸$ and

ii. $h(¸_k) \;!\; h(¸)$

imply that $¸_k$ d-weakly converges to $¸$.

For Bernoulli processes, a basic fact is (see [21]):

Theorem 5.2 Bernoulli processes are finitely determined.

For every finite subset I of the non-negative integers, and a finite vector $(!_i^0; i \in I)$ we define the cylinder

$$C(!_i^0; i \in I) \doteq \{! : !_i = !_i^0; i \in I\}$$

From the definition of D-weak convergence,

$$\nu_k \text{ D } \mu \text{ weakly converges to } \nu \text{ iff } \nu_k(C) \to \nu(C) \text{ for any cylinder } C: \quad (5.29)$$

We conclude from (5.28) and (5.29) that if a sequence θ_k converges to θ then the two conditions of the theorem (5.2) are satisfied, and therefore:

$$P^{\theta_k} \text{ d } \mu \text{ weakly converges to } P^{\theta}: \quad (5.30)$$

We now apply this concept to our problem.

Continuous choice procedures

Lemma 5.3 Assume that for every $j \in J$

$$C^j = C_0^j \cup C_1^j; \quad (5.31)$$

where

i. the set C_0^j has P^{θ}-measure zero for every θ;

ii. C_1^j is d-open,

then the function $\theta \to P^{\theta}(\{\lim_{t \to 1} c_t = j\})$ is continuous for every j.

Proof. Note first that if $f_j; j \in J$ is a vector of functions which are lower-semicontinuous and

$$\sum_{j \in J} f_j = 1; \quad (5.32)$$

then each f_j is continuous. In fact, for each $k \in J$, the function $\sum_{j \neq k} f_j$ is lower-semicontinous, so $1 - \sum_{j \neq k} f_j$ is upper-semicontinuous, but this is f_k. So in order to prove our claim it is enough to prove (since the condition (5.32) is satisfied by assumption (2.4)), and that each function is lower-semicontinous. But we know from (5.30) that P^{θ_k} d-weakly converges to P^{θ}, and therefore, by a basic property of weak convergence,

$$\liminf_k P^{\theta_k}(C^j) \geq P^{\theta}(C^j) \text{ for every } j: \quad (5.33)$$

■

Proof of proposition 2.1

Now we are ready for the ¯nal details of the proof of proposition (2.1). Each player has a set $J \acute{} f1; : : : ; j; : : : ; ng$ of actions that he can take in each period. A state is any possible sequence of actions that the opponent he faces in each period takes. So the set $-$ of states is the set of sequences of elements in J:

$$- \acute{} J^{1};$$

with a generic element $!$, and $!_t$ the state at time t. A given $® 2 ¢(J)$ induces the probability $P^{®}$ on the product space. We denote by $X_t(!)$ the value of a random variable when the time is t and the state is $!$. The payo® from the action j at time t when the state is $!$ is

$$g_t^j(!);$$

the choice of action at t is $c_t(!) 2 J$; the sum of payo®s from the action j is

$$G_t^j(!) \acute{} \sum_{fs2f1;::::;tg;c_s=jg} g_s^j(!);$$

the number of times the action j has been chosen is

$$n_t^j(!) \acute{}]fs 2 f1; : : : ; tg; c_s = jg$$

and the average payo® from j is the ratio:

$$a_t^j(!) \acute{} \frac{G_t^j(!)}{n_t^j(!)}:$$

In each period the action that maximizes the average payo® is:

$$c_t(!) \acute{} \operatorname{argmax}_{j2J} a_t^j(!) \tag{5.34}$$

For every subset $I \mu J$, the value of the best average among actions in I is

$$a_t^I(!) \acute{} \max_{j2I} a_t^j(!); \tag{5.35}$$

so that $a_t^J(!)$ is the best overall average. As usual, we use the supercript ¡ j to denote the set of all elements but j.

Let $t_1; t_2; : : : ; t_i; : : :$ be the sequence of switch times, that is:

$$c_t \neq c_{t\,i\,1} \text{ if and only if } t = t_i \text{ for some } i:$$

Then

$$\text{the function } i \;!\; a_{t_i} \text{ is decreasing :} \tag{5.36}$$

This is clear: consider for any i the time interval between t_{i-1} and t_i. At t_i the action with the best average changes because its average falls below the maximum of the averages of the other actions. In turn, the average of these actions was constant in the interval between t_{i-1} and t_i, and equal to the value at t_{i-1}. So the best average at t_i is less than the best average at t_{i-1}. Let

$$C^j \equiv \{\omega : c_t(\omega) = j \text{ for all but finitely many } t's\}$$

and $C \equiv \bigcup_{j \in J} C^j$.

We begin with the the proof of the first claim: C has probability 1. First consider the case where θ is such that the expected payoff from each action is different. The supposition that the choice of action does not converge leads to a contradiction. At least two actions have to be chosen infinitely many times, but the by the strong law of large numbers the average payoff converges to the expected payoff, and these are different for the two actions. So the one with the least payoff will not be chosen eventually.

Now consider the case in which the expected payoff from two or more actions is equal. Suppose that the choice of action is not eventually constant. By the previous argument, we may assume that two or more actions with the same expected payoff are chosen infinitely many times. Again the sequence of the averages is converging to the expected payoff on each of these actions. But from (5.36) we know that this sequence is decreasing, for each of these actions, to the expected payoff. This can only happen with probability zero, since the averages of a sequence of i.i.d. takes almost surely values on both sides of the expected value.

Now the second part: take a sequence $\theta_k \in \Phi(J)$ converging to θ. We claim that for every j,

$$\lim_k P^{\theta_k}(C^j) = P^{\theta}(C^j). \tag{5.37}$$

For each j, the set C^j is the union of the set C_1^j of ω's for which the action j is chosen, and the limit average is strictly larger than the maximum of the averages of the other actions, and the complement in C^j, denoted by C_0^j. Note that these sets are the same, independently of the probability P^θ. We claim that

i. the set C_1^j is d-open set, and

ii. the set C_0^j which has P^θ-probability zero for any $\theta \in \Phi(J)$.

In C_0^j the limit of the average is equal to the average of some of the other actions, and so with probability zero it assumes values always larger than its limit. For the first claim, take any point $\omega \in C_1^j$, and note that:

$$|\, t[a_t^j(\omega') - a_t^j(\omega)] \,| = \Big| \sum_{i : \omega_i' \neq \omega_i} g_i^j(\omega') - g_i^j(\omega)) \Big|$$

$$]\{i : !_i^0 \neq !_i\}M$$
$$td(!^0; !)M \tag{5.38}$$

where M is twice the maximum of the payo®, and therefore

$$\sup_{\{!^0 : d(!^0;!) \leq ½;\, t \geq 1\}} |a_t^j(!^0) - a_t^j(!)| \leq ½M;$$

so for a su±ciently small ½ the set of elements at distance ½ from $!$ is contained in C_1^j. Now the conclusion follows from lemma (5.3). ■

Proof of proposition 3.1

Denote by $X^i; i = A; B$ the random variables taking the value $\log(\frac{-_i}{\circ_i})$ with probability ® and $\log(\frac{1-_i}{1-\circ_i})$ with probability $1 - ®$ respectively: this is the value of the change in r_t^i in each period. Also let $M \geq |X^i|$ for both i. $E_® X^i$ is the expected value of X^i for a given value of ®. Recall that (see (3.17))

$$-_A < \circ_A; \quad -_B > \circ_B;$$

so as a function of ®, $E_® X^A$ ($E_® X^B$) is decreasing (increasing, respectively), with a positive (negative) value at 0 and negative (positive) value at 1, and is zero at two critical values of ®, that we denote by $®^i; i = A; B$:

$$E_{®^i} X^i = 0; i = A; B; \tag{5.39}$$

and $\underline{®} \equiv \min\{®^A; ®^B\}$, $\bar{®} \equiv \max\{®^A; ®^B\}$. To ¯x ideas, we assume that

$$E\circ_A g^1 > E\circ_B g^2; \tag{5.40}$$

so that at $(p^A; p^B) = (0; 0)$, the action A is chosen. The case with the opposite inequality is symmetric. Since from (3.13) B is chosen at $(p^A; p^B) = (0; 1)$, there is a value $\hat{p}^B$ at which the optimal policy switches from A to B. This implies that the optimal policy has the form: choose B if and only if $r^B \geq ¼(r^A)$, where

$$¼ \text{ is increasing, and } \lim_{r \to 1} ¼(r) = \log(\frac{\hat{p}^B}{1 - \hat{p}^B}): \tag{5.41}$$

(The behavior of ¼ as $r \to 1$ depends on whether A or B is chosen at $(p^A; p^B) = (1; 1)$. For completeness we note that in the ¯rst case ¼ tends to in¯nity as r approaches a ¯nite value of r^A, and in the second it tends to a ¯nite value. It is easy to see that these cases are all possible.)

We now analyze the behavior of the function $f : [0; 1] \to [0; 1]$ that describes the probability that the action A is eventually chosen.

For $\alpha \in [0, \alpha^B]$, $f(\alpha) = 1$. In fact both arms cannot be tried infinitely many times (suppose they do: then the two values of r_t^i both converge to -1, where A is chosen by (5.41), a contradiction), and the chosen arm cannot be B, because of (5.41) and for these values of α, $0 \geq E_\alpha X^B$.

We then prove that when $\alpha \in (\alpha^B, 1]$, so that $E_\alpha X^B > 0$, the function f is continuous. We use the lemma (5.3). Let

$$L = \{\omega : \lim_{t \to \infty} f_t^j(\omega) \text{ for } j = A, B\},$$

and

$$C_1^j = C^j \cap L; C_0^j = C^j \setminus L.$$

From the ergodic theorem, for every α $P^\alpha(L) = 1$, so $P^\alpha(C_0^j) = 0$. So we have to prove that the set of ω in C_1^B is open. So take an ω such that B is eventually chosen. Since for any other ω',

$$|r_t^B(\omega) - r_t^B(\omega')| = \left|\sum_{s=1}^{M(\{i : 1 \leq i \leq t, \omega_i \neq \omega_i'\})} (X_s^B(\omega) - X_s^B(\omega'))\right| \tag{5.42}$$

then

$$\left|\frac{r_t^B(\omega)}{t} - \frac{r_t^B(\omega')}{t}\right| \leq d(\omega, \omega'). \tag{5.43}$$

Also since $\omega \in L$, for some a

$$\lim_t \frac{r_t^B(\omega)}{t} = a \tag{5.44}$$

But if ω' is also in L and d-close to ω the corresponding limit of the frequency must be close (recall (5.27)). Then for for any ε, for some t large enough

$$\left|\frac{r_t^B(\omega')}{t} - a\right| < \varepsilon + Md(\omega, \omega') \tag{5.45}$$

and therefore:

$$r_t^B(\omega') \geq (a - \varepsilon - Md(\omega, \omega'))t \tag{5.46}$$

and therefore if $r_t^B(\omega)$ is larger than the value then so is $r_t^B(\omega')$, hence B is chosen at ω', hence the set of ω where B is chosen is open. ∎

References

[1] Arthur, W. B., (1991), "Designing Economic Agents that Behave like Human Agents: A Behavioral Approach to Bounded Rationality", American Economic Review Papers and Proceedings, 81, 353-359.

[2] Arthur, W. B., (1993), "On Designing Economic Agents that Behave like Human Agents" Journal of Evolutionary Economics, 3, 1-22.

[3] Aumann, R. J., (1964), \Markets with a continuum of traders" Econometrica, 32, 39-50.

[4] Berry, D. A. and Fristedt, B. (1979), \Bernoulli One-armed Bandits-Arbitrary Discount Sequences" The Annals of Statistics, 7, 5, 1086-1105.

[5] Berry, D. A. and Fristedt, B., (1985), Bandit Problems, Chapman and Hall, London

[6] Bärgers, T. and R. Sarin, (1997), \Learning Through Reinforcement and Replicator Dynamics", Journal of Economic Theory, 77, 1-14.

[7] Camerer, C., (1997), \Progress in Behavioral Game Theory" Journal of Economic Perspectives, 11, 167-188

[8] Chen, Hsiao-Chi, Friedman, J. W. and Thisse, J. F.,(1997), \Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach" Games and Economic Behavior, 18, 1, 32-54.

[9] Cross, J. G., (1983), A Theory of Adaptive Economic Behavior, Cambridge University Press, Cambridge, UK.

[10] Easley, D. and Rustichini, A, (1995), \Choice Without Beliefs", Econometrica, forthcoming.

[11] Fudenberg D., and Levine, D., (1995), "Universal Consistency and Cautious Fictitious Play", Journal of Economic Dynamics and Control, 19, 1065-1089.

[12] Gittins, J. C. and Jones, D. M., (1974), \A dynamic allocation index for the sequential design of experiments" in Progress in Statistics, editor: J. Gani, 241-266, North-Holland, Amsterdam.

[13] Gittins, J. C., (1979), \Bandit processes and dynamic allocation indices" Journal of the Royal Statistical Society, Sr. B, 41, 148-164.

[14] Gilboa, I. And Schmeidler, D., (1996), \Case-Based Decision Theory" Quarterly Journal of Economics, 110, 605-639.

[15] Lettau, M., and Uhlig, H., (1995), \Rules of Thumb and Dynamic Programming" CentER, Tilburg University Discussion Paper.

[16] Merlo, A. and Schotter, A., (1994), \An Experimental Study of Lerning in One and Two-Person Games" CV Starr Center Discussion Paper, RR] 94-17.

[17] Mookerjee, D. and Sopher, B, (1994) \Learning Behavior in Experimental Matching pennies Games" Games and Economic Behavior, 7, 62-91.

[18] Mookerjee, D. and Sopher, B, (1997) \Learning and Decision Costs in Experimental Constant-Sum Games" Games and Economic Behavior, 19, 97-132.

[19] Murnigham, J. K. and Roth, A. E., (1980), \The E®ect of Group Size and Communication Availability on Coalition Bargaining in a Veto Game", Journal of Personality and Social Psychology, 39, 92-103

[20] Ornstein, D. S., (1970), \Bernoulli shifts with the same entropy are isomorphic", Advances in Mathematics, 4, 337-352.

[21] Ornstein, D. S., (1970), \Imbedding Bernoulli shifts in °ows", in Contributions to Ergodic Theory and Probability, Lecture Notes in Mathematics, 160, Springer Verlag, New York, 178-218.

[22] Ornstein, D. S., (1974), Ergodic Theory, Randomness, and Dynamical Systems Yale Mathematical Monographs 5, Yale University Press, New Haven, Connecticut.

[23] Osborne, M. J. and Rubinstein, A., (1997), \Games with Procedurally Rational Players", American Economic Review, forthcoming.

[24] Partow, Z. and Schotter, A., (1993), \Does Game Theory Predict Well for the Wrong Reasons: an Experimental Investigation" CV Starr Center Discussion Paper, RR] 93-46.

[25] Petersen, K., (1983), Ergodic theory Cambridge Studies in Advanced Mathematics, 2, Cambridge University Press, Cambridge, UK.

[26] Rustichini, A. (1998), \Optimal Properties of Stimulus-Response Learning Models", forthcoming in Games and Economic Behavior.

[27] Schlag, K.H., (1998), "Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-Armed Bandits", Journal of Economic Theory, 78, 130-156.

[28] Sahrin, R. and Vahid, F., (1997), \Payo® Assessments without Probabilities: A Simple Dynamic Model of Choice", Texas AM Discussion Paper.

[29] Von Neumann, J. and Morgenstern, O., (1944), Theory of Games and Economic Behavior, Princeton University Press.

[30] Whittle, P., (1980), \Multi-armed bandits and the Gittins index" Journal of the Royal Statistical Society Ser. B, 42, 143-149.