97-F-21

Procedural Rationality and Inductive Learning I: Towards a Theory of Subjective Games

Hitoshi Matsushima, Faculty of Economics
University of Tokyo

June 1997

Procedural Rationality and Inductive Learning I: Towards a Theory of Subjective Games*

Hitoshi Matsushima+

Faculty of Economics, University of Tokyo

June 4, 1997

(The First Version: May 17, 1997)

---

* An earlier version of this paper was entitled "Procedural Rationality and Learning Behavior" which was written on the basis of the seminar at the Tokyo Center of Economic Research (TCER), Tokyo, January 1997. This paper was presented at the TCER International Summer Conference, Tokyo, May 1997, and the Eastern Meeting of the Japan Association of Economics and Econometrics, Shiga, May 1997. The brief summary of this paper was given by Section 5 of my servey article "Bounded Rationality in Economics: A Game Theorist's View," forthcoming in *the Japanese Economic Review* Vol. 48, No. 3, 1997. The author is grateful to Ichiro Ide and George Mailath for their helpful comments.

+ Faculty of Economics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan. Fax: (03) 3818-7082. E-mail: hitoshi@e.u-tokyo.ac.jp

Abstract

This paper investigates the situation of strategic conflict in which players have limited prior knowledge about the *objective game*, that is, they do not know their true, objective payoff functions, and therefore, have to formulate their own payoff functions based on their past *experiences* in a subjective way. In distinction with the objective game, we will define *the subjective game* by the combination of these subjective payoff functions and the sets of actions.

Most of real economic situations are complex and not even well-structured. A real economic agent spends most time to visualize and perceive the situation. Formulating the subjective game would be regarded as the most important step for an actual agent in reaching a decision. Despite its unquestionable importance, the investigation of the subjective game is at this time very immature. Especially, applied game theorists in the 1970's and 1980's have *never* dealt with the question of how players formulate the subjective game by assuming that the objective game is common knowledge among the players and assuming that players are ideally rational.

Since actual players are *boundedly rational* as Herbert A. Simon has stressed, they might formulate the subjective game which is essentially *different* from the objective game and *fail* to achieve a Nash equilibrium of the objective game. The main purpose of this paper is to give clear answers to several substantial questions such as what are the characteristics of the subjective games and the choices of actions. A player is modeled as an *inductive learning procedure* in a dynamic decision making which translates past experiences into subjective evaluations and decisions, and is mainly motivated by the maximization of the *subjective expected payoffs*. We will assume that, in every period, a player is *never* convinced that the situation is *recurrent*, and therefore, she can *not* establish a firm experience-based belief about the uncertain situation in which she will seldom waver when observing unlikely events.

By requiring a couple of plausible conditions on inductive learning procedures, we can derive the following drastic results in a wide class of environments including various recurrent and *non-recurrent* situations: The subjective game formulated in the long run belongs to an extremely restricted class of simplified games which are called *trivial games*. In a trivial game, there always exists the *unique* action profile which is both *strictly dominant* and *Pareto-efficient* among the set of pure action profiles. Moreover, players need not to be strategically sophisticated, because this strict dominance property holds *irrespective* of the details of its extensive form. Zero-sum games, prisoner-dilemma games, coordination games, stag-hunt games, and hawk-dove games are *not* trivial games, and therefore, are *never* perceived as subjective games. This strictly dominant action profile in the subjective game is *neither* a Nash equilibrium *nor* Pareto-efficient in the objective game. Of particular importance is that it is always equal to the *maximin action profile* in the objective game.

JEL Classification Numbers: C70, C90, D43, D80.
Keywords: inductive learning, procedural rationality, subjective game, maximin action. trivial game.

1. Introduction

This paper investigates the situation of strategic conflict in which players have limited prior knowledge about the *objective game*. Players know the set of actions, but do not know their true, objective payoff functions. Hence, players have to formulate their own payoff functions based on their *past experiences* in a subjective way. In distinction with the objective game, we will define *the subjective game* by the combination of these subjective payoff functions and the sets of actions.

Most of real economic situations, especially in the area of industrial organization, are complex and not even well-structured. A real economic agent spends most time to visualize and perceive the situation. Hence, as Thomas Schelling has already stressed in his celebrated book entitled "*The Strategy of Conflict*", formulating the subjective game would be regarded as the most important step for an actual agent in reaching a decision (Schelling (1960)). Selten (1978) has presented an informal model of the human reasoning process which takes into account the cognitive steps such as perception, problem solving, investigation, implementation, and learning, and has emphasized also that the step of perceiving the situation and formulating the subjective model is the most important[1].

Despite its unquestionable importance, the investigation of the subjective game is at this time very immature. Especially, applied game theorists in the 1970's and 1980's have *never* dealt with the question of how players formulate the subjective game: They have interpreted game theory in a *naïve* way that the objective game as a full description of a state of the physical world and a state of mind is assumed to be *common knowledge* among players. This naïve interpretation of game theory is strongly criticized, because players are sometimes required to be *ideally rational* in an extremely unrealistic way[2].

In the early 1950's, Herbert. A. Simon has introduced the concept of *bounded rationality* and emphasized that a real economic agent is not so rational as the

---

[1] See also van Damme (1995).

[2] For the criticisms on applied game theory and the naïve interpretation, see Fisher (1989), Pelzman (1991), Rubinstein (1991), van Damme (1995), Dekel and Gul (1997), and Matsushima (1997a). Rubinstein (1991) presented the *perceptive* interpretation of game theory as the alternative to this naïve interpretation, in which a combination of a game and a strategy profile is viewed as a *common perception* among players. Kaneko and Matsui (1996) investigated the situation in which players have their respective perceptions that may be incompatible each other.

neoclassical framework assumes[3]. Following the Simon's argument, it should be stressed that a real economic agent is not so ideally rational, or *substantively rational*, as applied game theorists assumes, but is *procedurally rational* in the sense that she perceives the situation and makes a decision as being compatible with the *cognitive* and *motivational* limits of rationality[4]. Such boundedly rational players might formulate the subjective game which is essentially *different* from the objective game, and might fail to achieve a Nash equilibrium of the objective game.

We have at this time very little knowledge about how the class of possible subjective games is restricted, in what way the subjective game is connected with the objective game, whether several famous games such as prisoner-dilemma games, coordination games, and hawk-dove games can be perceived as subjective games, and so on. This paper would be regarded as the first attempt to give clear answers to these questions, which are the most important ones in establishing a theory of subjective games.

In this paper, a player is modeled as an *inductive learning procedure* in the following way: Each player is repeatedly confronted with *randomly matched* opponents, plays games together with them, and chooses actions among the same set of actions. In every period, she can, not perfectly but *almost* perfectly, monitor which actions the randomly matched opponents have actually chosen, by observing the realization of a *random* signal. Before arriving at the current situation of conflict, she has accumulated the memories on the payoffs and the signals realized in the previous periods as her own experiences. On the basis of these experiences, she will formulate her own subjective payoff function, and also subjectively estimates the probability function on the set of the opponents' actions, according to which the current opponents will be anticipated to choose actions.

A player is motivated by *the maximization of the subjective expected payoff*, i.e., the

---

[3] See Simon (1976, 1982). Simon proposed that the rationality postulate of payoff-maximization in the neoclassical framework should be replaced by a more realistic *behavioral* hypothesis such as *satisficing with aspiration levels*, because the decisions in most economic environments are complicated enough to transcend the actual agent's cognitive capabilities. There are many papers discussing the importance of bounded rationality in economics and game theory. For example, Winter (1986), Binmore (1987, 1988), North (1990), Selten (1990, 1991), Aumann (1992), van Damme (1995), Rubinstein (1996), Conlisk (1996), Matsushima (1997b), and so on.

[4] The concepts of substantive rationality and procedural rationality were introduced by Simon (1976). Selten (1990) classified aspects of bounded rationality into the cognitive limits of rationality and the motivational limits of rationality, and emphasized the importance of the latter aspects.

sum of this subjective payoffs weighted by this subjective probabilities, as well as *the law of inertia*. Hence, in every period, she will not choose any action which is neither the same action as that chosen in the last period, nor one of the actions which maximize the subjective expected payoff.

We will assume throughout this paper that in every period, a player is *not* convinced that the situation is recurrent, and therefore, she can *not* establish a firm, experience-based, belief about the uncertain situation in which she will seldom waver when observing unlikely events. On this assumption, it is natural from the *psychological* aspects of human nature to require that a player will lay more stress on the *near* past experiences than the *far* past experiences. This requirement will be expressed by a condition on a learning procedure called *Uniform Adaptation* in Section 3: An inductive learning procedure satisfies Uniform Adaptation if there exists a *time-and-history independent* finite number of periods such that whenever a player has continued to monitor the same action profile and obtain some constant payoff for this number of periods, then she will always *equalize* the subjective payoff for this action profile with this constant payoff.

We will introduce another condition on a learning procedure called *Independence of Irrelevant Experiences*, which might be also plausible to require in a realistic context with respect to the psychological aspects of human nature. Independence of Irrelevant Experiences means that a player will *never* change her subjective payoff evaluation for an action profile as long as she does not actually monitor this action profile.

This paper examines a wide class of environments which includes various recurrent and *non-recurrent* situations. We nevertheless can derive the following very drastic consequences: The subjective game is essentially *different* from the objective game, and the subjective game always belongs to an extremely restricted class of simplified games which are called *trivial games*: A game $(N,(A_i,u_i)_{i\in N})$ is said to be trivial if for every $i\in N$, there exists an action $a_i^+ \in A_i$ such that

$$u_i(a_i^+,a'_{-i}) > u_i(a_i,a_{-i}) \text{ for all } a_{-i}\in A_{-i}, \text{ all } a'_{-i}\in A_{-i}, \text{ and all } a_i \neq a_i^+ .$$

Hence, in a trivial game, there always exists a unique action profile which is both *strictly dominant* and *Pareto-efficient* among the set of pure action profiles.

Most of the games which have been intensively studied such as zero-sum games, prisoner-dilemma games, coordination games, stag-hunt games, and hawk-dove games are *not* trivial games, and therefore, are *never* perceived as subjective games. This is in contrast with the applied game theory developed in the 1970's and 1980's in which various complex extensive form games have been contrived in order to classify situations in an extremely flexible way and have been required to be common knowledge among players.

This strictly dominant action profile in the subjective game, $(a_i^+)_{i\in N} \in \times_{i\in N} A_i$, is regarded as the apparently obvious solution in noncooperative game theory, because the strict dominance property always holds *irrespective* of how the details of its extensive form such as the order of players' moves are specified, and therefore, players need *not* to be strategically sophisticated. This point also is in contrast with the applied game theory in which the predictions derived from the refinements of Nash equilibrium sometimes depend substantially on the very details of its extensive form[5].

Moreover, this action profile $(a_i^+)_{i\in N} \in \times_{i\in N} A_i$ is, in general, *neither* a Nash equilibrium *nor* Pareto-efficient in the objective game. Of particular importance is that this action profile is always equal to the *maximin action profile* in the objective game, where a maximin action is defined by the action which maximizes her own payoff minimized by the opponents' actions. Players nevertheless have lost the will to bridge the gap between the objective game and the subjective game in the long run.

For example, consider a *hawk-dove game* presented in Figure 1.1 as the objective game, where the mixed action assigning "dove" probability $1/3$ and "hawk" probability $2/3$ is the unique *evolutionary stable strategy* (ESS) addressed by Maynard Smith (1982) in the literature of *evolutionary game theory in biology*.[6] However, the subjective game formulated in the long run is approximated by Figure 1.2. Players comes to choose "dove" with probability 1 in the long run, which is the maximin action in the objective game. The action profile (dove, dove) is strictly dominant and Pareto-efficient in this subjective game, but is neither Nash equilibrium nor Pareto-efficient in the objective game.

---

[5] We must have the fact impressed on our mind that the attempts of applied game theory have been objected mainly because the flexibility in modeling extensive form games and the tight dependence of the predictions on the very details of the models make it impossible to judge which models are more appropriate and also make it difficult to get general insights with significant economic implications. See Fisher (1989), Pelzman (1991), and also van Damme (1994).

[6] See Hammerstein and Selten (1994) and van Damme (1987, Chapter 9).

	dove	hawk
dove	1, 1	0, 3
hawk	3, 0	-1, -1

Figure 1.1: Hawk-dove Game
as the Objective Game

	dove	hawk
dove	1, 1	0, L_4
hawk	L_1, 0	L_2, L_3

Figure 1.2: Trivial Game
as the Subjective Game
$(L_h < 0 \textit{ for } h = 1,2,3,4)$

The organization of this paper is as follows. In Section 2, we will explain the basic ideas in deriving several results of this paper by investigating a simple example of single-person decision making. In this example, it is shown that a decision maker (a manager) fails to maximize the objective expected payoff in the long run. In Section 3, we will present the formal model of the general single-person decision making problem, and will show that a decision maker comes to choose the maximin actions in the long run.

Section 4 is the main part of this paper, which considers the multi-person decision making problem with strategic conflict. It is shown that the subjective game is always a trivial game, and the maximin action profile in the objective game is strictly dominant and Pareto-efficient among the set of pure action profiles in the subjective game. Section 5 investigates several examples of the objective games, i.e., stag-hunt games and coordination games.

The results of this paper may be regarded as being *negative* by orthodox game theorists, because most of the famous games such as zero-sum games, prisoner-dilemma games, coordination games, stag-hunt games, and hawk-dove games can not be viewed as the subjective games. We, however, would like to stress that this paper would be regarded as the *benchmark* for the possible progress towards a theory of subjective games. Section 6 gives several discussions relevant to this issue.

Finally, Section 7 argues the relevance to evolutionary game theory in *economics*.

2. Formulating the Subjective Payoff Function: An Example

It would be helpful to start with the following example of recurrent situation in single-person decision making. A manager (a decision maker) repeatedly decides whether to enforce a *risky* project or a *safe* project. There are also two possible states of the world, i.e., state "*boom*" and state "*recession*". In every period, state "boom" occurs with a positive probability $p > 0$, whereas state "recession" occurs with a positive probability $1 - p > 0$. The probability p is time-and-history independent, and states are determined in a time-independent way. If the manager enforces the risky project and state "boom" occurs in a period, then she obtains one hundred dollars at the end of this period. If the manager enforces the risky project and state "recession" occurs, then she loses one hundred dollars. If she enforces the safe project, then she obtains zero dollar irrespective of which state actually occurs. We will assume $p > \frac{1}{2}$. The risky project is more profitable than the safe project, because the objective expected payoff for the risky project, $(2p - 1) \times 100$, is more than zero induced by the safe project (see Figure 2.1).

	boom $1 > p > \frac{1}{2}$	recession $0 < 1 - p < \frac{1}{2}$
risky	100	-100
safe	0	0

Figure 2.1: The Objective Payoff Function

The manager has limited prior knowledge about the structure of this decision making problem. Throughout this section we assume that the manager does *not* know the probability p. (If she is convinced that the probability of state "boom" is p, then she will always enforce the risky project, because by doing so she can maximize the expected payoff.) We also assume that the manager can *not* observe the realization of the states. (If she can observe the realizations of the states, then she eventually comes to learn p

approximated by the *relative frequency* of the realization of state "boom" in the previous periods.)[7]

From now on, we will investigate three cases. i.e., Cases 1, 2, and 3. Case 1 will suppose that the decision maker is always *convinced* that the situation is *recurrent*, whereas Cases 2 and 3 will suppose that she is *not* convinced of it. Moreover, Cases 1 and 2 will suppose that the manager a priori knows the true objective payoff function, whereas Case 3 will suppose that the manager a priori knows that the safe project always induces zero dollar but she does *not* know the state-dependent payoffs induced by the risky project.

Case 1 will give a scenario that the manager succeeds to enforce this profitable risky project in the long run. On the other hand, Cases 2 and 3 will give their respective scenarios that the manager *fails* to enforce this risky project. Among these cases, Case 3 is the most relevant to our concern.

CASE 1: Since the manager does not know p, she has to evaluate the subjective probability $\delta(t) \in [0,1]$ of the realization of state "boom" in every period t on the basis of the past experiences. The manager is *not* convinced that her initial probability evaluation $\delta(1)$ is correct and is willing to change the probability evaluation after experiencing unlikely events for a time being.

Since the manager is convinced that the situation is recurrent, it is natural to assume that she always puts the *same* stress on every past experience in evaluating the subjective probability $\delta(t)$. In this case, the influence of new experience on this evaluation gradually *weakens*, and in the long run the manager will establish a firm experience-based belief about this uncertain situation in which she will seldom waver when observing unlikely events.

For example, let $T(t)$ denote the set of all periods up to period t in which the decision maker has enforced the risky project, and let a subset $\overline{T}(t) \subset T(t)$ denote the set of all periods up to period t in which the manager has enforced the risky project and obtained one hundred dollars. Suppose that $\delta(t)$ is determined by

---

[7] According to the same idea as this parentheses, Milgrom and Roberts (1991) showed in a context of learning in games that if players can perfectly monitor the opponents' choices of actions, then they come to choose a *rationalizable* action profile in the long run. This result relies crucially on the *perfection* of the monitoring abilities in the strict sense. We will show in this paper that by adding a slight noise of observation, players instead come to choose the maximin action profile in the long run.

$$\delta(t) = \frac{|\overline{T}(t)|}{|T(t)|},$$

where $|B|$ is the number of the elements in the set B. The manager always equalizes the subjective probability $\delta(t)$ with the number of the previous periods in which the manager has obtained one hundred dollars divided by the number of the previous periods in which she has enforced the risky project.

Moreover, suppose that the manager will always *experiment* with the risky project at least with a small but positive probability $\xi > 0$. Hence, as time passed, the number of periods in which the manager has enforced the risky project increases without limits, and therefore, the influence of new experience on the probability evaluation $\delta(t)$ gradually becomes *negligible*. *The law of large numbers* says that it is almost certain in the long run that the probability evaluation $\delta(t)$ approximates the true probability p, which holds *irrespective* of how $\xi > 0$ is given. Hence, by choosing ξ close to zero, we can conclude that the manager who is mainly motivated by the maximization of the subjective expected payoff will succeed to maximize the objective expected payoff, i.e., to enforce the risky project.

CASE 2: We will suppose that the manager is *not* convinced that the situation is recurrent, and therefore, she may not put the *same* stress on every past experience in evaluating $\delta(t)$. It would be natural from the psychological aspects of human nature to assume that the manager always believes that the *near* past experiences are *more* tightly related to the current situation than the far past experiences, and therefore, she will put the *heavier* stress on the near past experiences than the *far* past experiences in evaluating the subjective probability $\delta(t)$. In this case, the influence of new experience on this evaluation is kept *non-negligible* in every period, and therefore, the manager can *not* establish a firm, experience-based, belief about the uncertain situation in which she will seldom waver when observing unlikely events.

Suppose that the manager believes at the very beginning of period 1 that the risky project is more profitable than the safe project, i.e., $\delta(1) > 1/2$, and therefore, enforces the risky project in period 1 because the subjective expected payoff $(2\delta(1) - 1) \times 100$ for the risky project is larger than zero induced by the safe project. However, if the manager continues to lose one hundred dollars for a long time, she will gradually decrease the subjective probability $\delta(t)$, make it less than $1/2$, and eventually stop to enforce the risky project. Assume that the probability of experimenting with the risky project $\xi > 0$ is very close to zero. Once the manager makes $\delta(t)$ less than $1/2$ and stops to enforce the risky project, it is unlikely for her to enforce the risky project again in the near future: After enforcing the safe project, the manager constantly obtains zero dollar and gains *no* informative experience concerning which state has actually occurred. Since the probability of experimenting is very small, almost no experiences in the future influence the subjective probability, and therefore, the manager will inevitably stick to this pessimistic evaluation for a very long time.

The important point is that the event that $\delta(t) < \frac{1}{2}$ always occurs with a probability more than some positive time-and-history independent real number, because the manager will put the heavier stress on the near past experiences. This, together with the assumption that the probability of experimenting is very small, urges the manager to stay "safe" for a very long time. This point is in contrast with the arguments in Case 1, where, as time passed, the subjective probability $\delta(t)$ gradually approaches the true probability $p > \frac{1}{2}$, and therefore, the probability that $\delta(t) < \frac{1}{2}$ approaches zero.

More precisely, there exists a positive integer s^* such that for *every* period t, if the manager continues to enforce the risky project and lose one hundred dollars from period t through period $t + s^* - 1$, then she always evaluates $\delta(t + s^*)$ less than $\frac{1}{2}$ in period $t + s^*$, *irrespective* of what are the history of the past experiences up to period t. (This will be expressed in Section 3 by *Uniform Adaptation*.) In every period, the event that the manager continues to lose one hundred dollars for the next s^* periods will occur at least with a positive probability $(1-p)^{s^*} > 0$, which means that it is certain that this event will eventually occur, provided the manager continues to enforce the risky project. Hence, it is almost certain in the long run that the manager continues to enforce the safe project for a very long time.

From now on, for the simplicity of the arguments, the probability of experimenting with the risky project is assumed to be *zero*. Thus, it is certain that the manager eventually comes to decide *not* to enforce the risky project *forever*.

For example, let $\theta \in [0,1]$ denote the *discount factor*, and suppose that $\delta(t)$ is determined by

$$\delta(t) = \frac{\sum_{\tau \in \bar{T}(t)} \theta^{t-\tau}}{\sum_{\tau \in T(t)} \theta^{t-\tau}}.$$

This implies that for every $\tau < t$, the manager will put the $\frac{1}{\theta}$ times heavier stress on the experience in period τ than the experience in period $\tau - 1$. We must note that if $\theta = 1$, this evaluation rule is equivalent to that presented in Case 1. If $\theta < 1$, then any integer $s^* > \log_\theta(\frac{1}{2})$ satisfies that whenever the manager continues to enforce the risky project and lose one hundred dollars for s^* periods, then $\delta(t+s^*) < \frac{1}{2}$ holds. (Because

$$\delta(t+s^*) = \frac{\sum_{\tau \in \bar{T}(t+s^*)} \theta^{t+s^*-\tau}}{\sum_{\tau \in T(t+s^*)} \theta^{t+s^*-\tau}} \leq \frac{\sum_{\tau=s^*+1}^{\infty} \theta^{\tau}}{\sum_{\tau=1}^{\infty} \theta^{\tau}} = \theta^{s^*} < \frac{1}{2}.)$$

Hence, according to this evaluation rule, the manager eventually comes to enforce the safe project forever.

The failure to enforce the risky project relies crucially on the assumption that when enforcing the safe project the manager gets *no* informative experience concerning *which state has actually occurred*. From now on, we will assume that in every period the manager always observes the realization of some *random signal* which is tightly related to the realization of the state: If state "boom" occurs, then the manager observes signal "B" with probability $1-\varepsilon > 0$ and signal "R" with probability $\varepsilon > 0$ irrespective of whether to enforce the risky project or not. Similarly, if state "recession" occurs, then the manager observes signal "B" with probability $\varepsilon > 0$ and signal "R" with probability $1-\varepsilon > 0$. By choosing ε sufficiently close to zero, the manager can, not perfectly but *almost* perfectly, monitor the realized states. We will assume that the manager is *convinced* that the observation of signal "B" (the observation of signal "R") almost surely means the realization of state "boom" (the realization of state "recession"). Hence, the manager eventually comes to learn p approximated by the relative frequency of observing signal "B" in the past experiences, and therefore, succeeds to enforce the risky project with almost certainty (see Figure 2.2).

	Signal "B"	Signal "R"
boom	$p(1-\varepsilon)$	$p\varepsilon$
recession	$(1-p)\varepsilon$	$(1-p)(1-\varepsilon)$

Figure 2.2: The Probability Structure

CASE 3: The last statement in Case 2 relies crucially on the assumption that the manager a priori *knows* the objective payoff function. In Case 3 we will suppose that the manager knows that the safe project always induces zero dollar but she does not know the state-dependent payoffs induced by the risky project. Moreover, suppose that the manager does *not* even know that the payoff for the risky project is determined only by which state actually occurs. We will show below that the manager fails to enforce the risky project in the long run, even though she can almost perfectly monitor the true states.

In every period t, the manager has to subjectively evaluate the payoffs $v^{(B)}(t)$ and $v^{(R)}(t)$ for the risky project associated with states "boom" and "recession" respectively, as well as $\delta(t)$. Assume that at the beginning of period 1, the manager evaluates $v^{(B)}(1) > 0$ and $v^{(R)}(1) > 0$, and therefore, enforces the risky project irrespective of how the initial probability evaluation $\delta(1)$ is given. The manager is not convinced that these evaluations are correct and is willing to change them after experiencing unlikely events. For every period t, if the manager continues to enforce the risky project, observe signal "B", and lose one hundred dollars for a large finite number of periods, say s' periods, from period $t+1$, then she evaluates $v^{(B)}(t+s') < 0$, and evaluates $\delta(t+s')$ close to unity in period $t+s'$. Hence, the manager's subjective expected payoff for the risky project in period $t+s'$, $\delta(t+s')v^{(B)}(t+s') + (1-\delta(t+s'))v^{(R)}(t+s')$, becomes less than zero, and therefore, the manager will stop to enforce the risky project. We must note that the payoff evaluation $v^{(R)}(t+s')$ for state "recession" is kept equal to $v^{(R)}(t)$ evaluated in period t, because the manager has never observed signal "R", and therefore, never gained experiences relevant to state "recession" during these periods. We must note also that this event will occur at least with a positive probability $\{(1-p)\varepsilon\}^{s'} > 0$.

Next, if the manager continues to observe signal "R" for a large finite number of periods, say s'' periods, from period $t+s'+1$, then she makes $\delta(t+s'+s'')$ close to zero in period $t+s'+s''$. Hence, the manager's subjective expected payoff for the risky project in period $t+s'+s''$ becomes more than zero, and therefore, the manager will start to enforce the risky project again. We must note that this event will occur at least with a positive probability $\{p\varepsilon + (1-p)(1-\varepsilon)\}^{s''} > 0$.

Furthermore, if the manager continues to enforce the risky project, observe signal "R", and lose one hundred dollars for a large finite number of periods, say s''' periods, from period $t+s'+s''+1$, then she evaluates $v^{(R)}(t+s'+s''+s''') < 0$. We must note that the payoff evaluation $v^{(B)}(t+s'+s''+s''')$ for state "boom" in period $t+s'+s''+s'''$ is kept equal to $v^{(B)}(t+s') < 0$ evaluated in period $t+s'$, because the manager has never observed signal "B" from period $t+s'+1$ through period $t+s'+s''+s'''$. Hence, the manager's subjective expected payoff for the risky project in period $t+s'+s''+s'''$

becomes less than zero, and therefore, the manager will stop to enforce the risky project again. Of particular importance is that from period $t+s'+s''+s'''+1$ the manager never decides to enforce the risky project: The manager has come to believe that the payoff for the risky project is always *negative*, irrespective of which state occurs, i.e.,

$$v^{(B)}(t+s+s'+s''+s''') < 0 \textit{ and } v^{(R)}(t+s+s'+s''+s''') < 0.$$

Since she gains no informative experience concerning *how the payoff for the risky project is related to the realization of state* when enforcing the safe project, she inevitably sticks to those pessimistic payoff evaluations. This event will occur at least with a constant positive probability $\{(1-p)(1-\varepsilon)\}^{s'''} > 0$.

From the above observations, we have shown that in every period, the event that the manager comes to make both payoff evaluations negative and stop to enforce the risky project will occur at least with a positive probability

$$\{(1-p)\varepsilon\}^{s'}\{p\varepsilon+(1-p)(1-\varepsilon)\}^{s''}\{(1-p)(1-\varepsilon)\}^{s'''} > 0,$$

which implies that this event will eventually occur. Hence, the manager in the long run comes to evaluate the payoffs for the risky project negative, and decide not to enforce the risky project forever. The subjective payoff function which the manager formulates in the long run is approximately described by Figure 2.3.

	boom (signal "B")	recession (signal "R")
risky	L_1	L_2
safe	0	0

Figure 2.3: The Objective Payoff Function
($L_h < 0$ *for* $h = 1,2$)

In the next section, we will extend the results obtained in this section to the more general class of recurrent and *non-recurrent* environments with *multiple* risky actions.

3. Maximin Actions

In this section, we will consider *a long-run single-person decision making problem* by $D \equiv (A, \Omega, \Phi, H, u^{(\cdot)}, p^{(\cdot)}, q^{(\cdot)})$ which is defined in the following way: A is the finite set of *actions*. Ω is the finite set of *states*. Φ is the finite set of *signals*. A decision maker repeatedly chooses actions among A infinitely many times. In each period $t \geq 1$, the decision maker chooses an action $a(t) \in A$, and a state $\omega(t) \in \Omega$ is realized. We will assume that the decision maker can *not* observe the realization of state $\omega(t)$. At the end of each period $t \geq 1$, the decision maker obtains a payoff $v(t) \in R$ and observes the realization of a *random* signal $\phi(t) \in \Phi$ which is related to the state $\omega(t)$.

Let h^0 be *the null history*, let $H^0 \equiv \{h^0\}$, and let $u^{(h^0)}: A \times \Omega \to R$ be the payoff function in period 1. Recursively, for every $t \geq 1$, let $h^t \equiv (a(\tau), \omega(\tau), v(\tau), \phi(\tau))_{\tau=1}^t$ be *a history up to period* t and $u^{(h^t)}: A \times \Omega \to R$ be the payoff function in period $t+1$ provided the history h^t up to period t is realized, where we assume that for every $\tau \in \{1, \dots, t\}$,

$$v(\tau) = u(a(\tau), \omega(\tau)).$$

Let H^t be the set of all histories h^t up to period t, and let $H \equiv \bigcup_{t=0}^{\infty} H^t$. In every period t, the decision maker obtains the payoff $v(t) = u^{(h^{t-1})}(a(t), \omega(t))$ according to the history-dependent payoff function $u^{(\cdot)}$.[8]

Let $p^{(h^{t-1})} \equiv p^{(h^{t-1})}(\cdot | a(t)): \Omega \to R_+$ be a history-dependent conditional probability function on Ω, and let $q^{(h^{t-1})} \equiv q^{(h^{t-1})}(\cdot | a(t), \omega(t)): \Phi \to R_+$ be a history-dependent conditional probability function on Φ, where

$$\sum_{\omega \in \Omega} p^{(h^{t-1})}(\omega | a(t)) = 1 \text{ and } \sum_{\phi \in \Phi} q^{(h^{t-1})}(\phi | a(t), \omega(t)) = 1.$$

Given that $h^{t-1} \in H^{t-1}$ was realized and $a(t) \in A$ was chosen in period t, a state $\omega(t) \in \Omega$ is realized with probability $p^{(t-1)}(\omega(t) | a(t))$, and the decision maker observes a signal $\phi(t) \in \Phi$ with probability $q^{(h^{t-1})}(\phi(t) | a(t), \omega(t))$.

Assume that for every $a \in A$, there exists a real number $\underline{u}(a) \in R$ which satisfies

$$\underline{u}(a) = \min_{\omega \in \Omega} u^{(h^t)}(a, \omega) \text{ for all } t \geq 0 \text{ and all } h^t \in H^t.$$

That is, $\underline{u}(a)$ is the *minimal payoff* for action $a \in A$. Define

$$\underline{u} \equiv \max_{a \in A} \underline{u}(a).$$

---

[8] In this paper, we will sometimes refer to the case of history-independent payoff function as a recurrent situation, and the case of history-dependent payoff function as a non-recurrent situation.

Definition 1: An action $a^* \in A$ is a *maximin action* if

$$\underline{u}(a) = \underline{u}.$$

A maximin action maximizes the minimal payoff $\underline{u}(a)$ with respect to pure action $a \in A$.[9] The set of maximin actions is denoted by $A^* \subset A$. The set of maximin actions A^* does *not* depend on the probability structure $(p^{(\cdot)}, q^{(\cdot)})$. In the example of Section 2 the safe project is the only maximin action.

We will present a technical condition on D as follows.

Condition 1 (Uniform Positive Lower Bound for $(p^{(\cdot)}, q^{(\cdot)})$): There exists a positive real number $\varepsilon > 0$ such that for every $t \geq 1$, every $h^{t-1} \in H^{t-1}$, and every $(a, \omega, \phi) \in A \times \Omega \times \Phi$,

$$p^{(h^{t-1})}(\omega|a) \geq \varepsilon \text{ and } q^{(h^{t-1})}(\phi|a, \omega) \geq \varepsilon.$$

Condition 1 implies that $p^{(\cdot)}$ and $q^{(\cdot)}$ have *the full supports* with *a time-and-history independent positive lower bound.*

A decision maker is modeled as *an inductive learning procedure* which is defined by (d, Γ), where $\Gamma \equiv (\Gamma^{(a)})_{a \in A}$, and $\Gamma^{(a)} \equiv ((v^{(a,\phi)})_{\phi \in \Phi}, \delta^{(a)})$: The set of *mixed actions* is denoted by $\Delta(A)$. The decision maker chooses among A according to *a decision rule* $d: H \rightarrow \Delta(A)$. For every $t \geq 1$ and every $h^{t-1} \in H^{t-1}$, the decision maker chooses each action $a \in A$ with probability $d(h^{t-1})(a)$. Since the decision maker can not observe the states, it is clear that $d(h^{t-1})$ is *independent* of $(\omega(\tau))_{\tau=1}^{t-1}$ for all $t \geq 1$ and all $h^{t-1} \in H^{t-1}$.

The set of all probability functions on Φ is denoted by $\Delta(\Phi)$. *An evaluation rule for an action* $a \in A$ is defined by $\Gamma^{(a)} = ((v^{(a,\phi)})_{\phi \in \Phi}, \delta^{(a)})$, where $v^{(a,\phi)}: H \rightarrow R$ and $\delta^{(a)}: H \rightarrow \Delta(\Phi)$. According to $v^{(a,\phi)}$, the decision maker *subjectively* evaluates the payoff $v^{(a,\phi)}(h^{t-1})$ which she obtains when choosing $a(t) = a$ and observing $\phi(t) = \phi$ in period t, provided h^{t-1} was realized. According to $\delta^{(a)}$, the decision maker anticipates that she observes a signal $\phi(t) = \phi$ with probability $\delta^{(a)}(h^{t-1})(\phi)$ when choosing $a(t) = a$, provided $h^{t-1} \in H^{t-1}$ was realized. *An evaluation rule* is defined by $\Gamma \equiv (\Gamma^{(a)})_{a \in A}$.

For every $a \in A$, every $t \geq 1$, and every $h^{t-1} \in H^{t-1}$, *the subjective expected payoff for an action* $a \in A$ is defined by

---

[9] The definition of maximin action in this paper is different from the definition in the textbook: The latter is in terms of mixed action, whereas the former is in terms of pure action.

$$V^{(a)}(h^{t-1}) \equiv \sum_{\phi\in\Phi} \delta^{(a)}(h^{t-1})(\phi) v^{(a,\phi)}(h^{t-1}),$$

which is the sum of the evaluations $v^{(a,\phi)}(h^{t-1})$ with respect to $\phi \in \Phi$ each weighted by the subjective probability $\delta^{(a)}(h^{t-1})(\phi)$.

Definition 2: A decision rule d is *consistent with* an evaluation rule Γ if for every $t \geq 1$ and every $h^{t-1} \in H^{t-1}$,

$$[V^{(a)}(h^{t-1}) \geq V^{(a')}(h^{t-1}) \text{ for all } a' \in A] \Rightarrow [d(h^{t-1})(a) > 0],$$

and

$$[a \neq a(t-1) \text{ and } V^{(a)}(h^{t-1}) < V^{(a')}(h^{t-1}) \text{ for some } a' \in A] \Rightarrow [d(h^{t-1})(a) = 0].$$

The consistency of (d, Γ) implies that the decision maker maximizes the subjective expected payoff with a positive probability, and she *never experiments* with any action which neither maximizes the subjective expected payoff nor is the same as the action chosen in the last period.

We will present two conditions on a decision rule d as follows.

Condition 2 (Law of Inertia): For every $t > 1$ and every $h^{t-1} \in H^{t-1}$,

$$d(h^{t-1})(a(t-1)) > 0.$$

Condition 3 (Uniform Positive Lower Bound for d): There exists a positive real number $\varepsilon > 0$ such that for every $t \geq 1$, every $h^{t-1} \in H^{t-1}$, and every $a \in A$,

$$[d(h^{t-1})(a) > 0] \Leftrightarrow [d(h^{t-1})(a) \geq \varepsilon].$$

Condition 2 implies that the decision maker chooses the same action as that chosen in the last period with a positive probability. Condition 3 is a technical condition similar to Condition 1.

We will present three conditions on an evaluation rule Γ as follows.

Condition 4 (Minimal Evaluation): For every $t \geq 1$, every $h^{t-1} \in H^{t-1}$, and every $(a, \phi) \in A \times \Phi$,

$$v^{(a,\phi)}(h^{t-1}) \geq \underline{u}(a).$$

Condition 5 (Independence of Irrelevant Experiences): For every $t \geq 1$, every $h^t \in H^t$, and every $(a, \phi) \in A \times \Phi$,

$$[(a(t), \phi(t)) \neq (a, \phi)] \Rightarrow [v^{(a,\phi)}(h^t) = v^{(a,\phi)}(h^{t-1})].$$

Condition 6 (Uniform Adaptation): There exists a positive integer s^* such that for every $(a,\phi) \in A \times \Phi$, every $t > s^*$ and every $h^{t-1} \in H^{t-1}$, if

$$(a(\tau), v(\tau), \phi(\tau)) = (a, \underline{u}(a), \phi) \text{ for all } \tau = t - s^*, \ldots, t-1,$$

then

$$v^{(a,\phi)}(h^{t-1}) = \underline{u}(a), \text{ and } \delta^{(a')}(h^{t-1})(\phi) = 1 \text{ for all } a' \in A.$$

Condition 4 implies that the evaluations for an action are *never* less than the minimal payoff for this action. Condition 5 implies that the evaluation for a combination of an action and a signal is influenced *only* by the experiences which the decision maker gains when *actually* choosing this action and observing this signal. Condition 6 implies that when continuing to choose an action $a \in A$ and observe the minimal payoff $\underline{u}(a)$ and a signal $\phi \in \Phi$ for a *time-and-history independent* finite number of periods, say, s^* periods, the decision maker always comes to make the payoff evaluation for (a, ϕ) equal to this minimal payoff and also comes to believe that she certainly observes this signal ϕ *irrespective of* which action she actually chooses.

Remark: The difference from the *Bayesian framework* pioneered by Savage (1954) and Harsanyi (1967, 1968) is important. The Bayesian framework assumes that a decision maker knows the objective model, whereas our theory does not. Harsanyi has applied the Bayesian framework to a quite wide class of multi-person decision making problems with uncertainty, and advocated the doctrine that every situation of *incomplete information* can be described by a state of nature in a well-defined *Bayesian game* which is assumed *common knowledge* among players. This "Harsanyi doctrine" is sometimes strongly criticized, because it relies heavily on the unrealistic assumption that players are *ideally rational*.[10] Since an actual economic agent is not so rational, her subjective model is essentially different from the objective one and she does *not* even know the entire structure of her own subjective model. The current paper characterizes the subjective model not from the view-points of rationality but from the *psychological* aspects of human natures, and views the decision maker as the *myopic* expected-payoff maximizer

---

[10] Harsanyi has emphasized that the common knowledge assumption of a Bayesian game and the common prior assumption is automatically satisfied if players are ideally rational. However, it is at present pointed out that the Harsanyi doctrine is not necessarily justified only by ideal rationality. See Dekel and Gul (1996).

within her own subjective world.[11]

Definition 3: For every $t' \geq 1$ and every $t > t'$, a history up to period t, $h^t \in H^t$, is *reachable from* a history up to period t', $h^{t'} \in H^{t'}$, *with respect to a decision rule* d if

$$d(h^{\tau-1})(a(\tau)) > 0 \text{ for all } \tau \in \{t'+1,\dots,t\}.$$

Theorem 1: *Suppose that* D *satisfies Condition 1,* d *satisfies Conditions 2 and 3,* Γ *satisfies Conditions 4, 5 and 6, and* d *is consistent with* Γ. *Then, the following two properties hold.*

(i) For every $\xi \in (0,1]$, *there exists a positive integer* s *such that for every* $t \geq s$, *it holds at least with probability* $1-\xi$ *that for every* $a \in A$ *which is not a maximin action,*

$$d(h^t)(a) = 0, \textit{ and } v^{(a,\phi)}(h^t) < \underline{u} \textit{ for all } \phi \in \Phi.$$

(ii) For every $t' \geq 1$ *and every* $h^{t'} \in H^{t'}$, *if for every* $a \in A$ *which is not a maximin action,*

$$d(h^{t'})(a) = 0, \textit{ and } v^{(a,\phi)}(h^{t'}) < \underline{u} \textit{ for all } \phi \in \Phi,$$

then, for every $t > t'$, *every* $h^t \in H^t$ *which is reachable from* $h^{t'}$, *and every* $a \in A$ *which is not a maximin action,*

$$d(h^t)(a) = 0, \textit{ and } v^{(a,\phi)}(h^t) < \underline{u} \textit{ for all } \phi \in \Phi.$$

Theorem 1 implies that almost certainly in the long run, the decision maker comes to choose the maximin actions and the payoff evaluations for the maximin actions are greater than the payoff evaluations for all actions that are not maximin.[12]

The proof of Theorem 1 is summarized as follows: Consider an action $a \in A$ which is not a maximin action, i.e., $\underline{u}(a) < \underline{u}$. Condition 6 (Uniform Adaptation) says that when continuing to choose this action $a \in A$, the decision maker eventually makes $v^{(a,\phi)}(h^{t-1})$ less than $\underline{u}$ for every signal $\phi \in \Phi$. After that, the decision maker never chooses this

---

[11] In this paper, an inductive learning procedure is assumed to be exogenously given. Several papers such as Abreu and Rubinstein (1988) derived decision rules endogenously by regarding boundedly rational players as the payoff-maximizers with the explicit constraints on the limitation of information processing ability. This approach, however, is sometimes criticized, because computing an optimal rule with the constraints of bounded rationality is much more difficult than computing without constraints. See Gilboa (1988), Lipman (1991), Conlisk (1996) and Matsushima (1997a).

[12] We must note that this does *not* imply that the decision maker is motivated by choosing maximin actions in the subjective game.

action as long as the subjective expected payoffs for all of the other actions become less than $\underline{u}$. However, Condition 4 (Minimal Evaluation) says that the subjective expected payoff for any maximin action is never less than $\underline{u}$. Hence, the manager will not choose any action which is not a maximin action in the long run.

Proof of Theorem 1: First of all, we will prove Part (ii) of this theorem. Fix $t' \geq 1$ and $h^{t'} \in H^{t'}$ arbitrarily, and suppose that for every $a \in A$ which is not a maximin action,

$$d(h^{t'})(a) = 0 \text{ and } v^{(a,\phi)}(h^{t'}) < \underline{u} \text{ for all } \phi \in \Phi.$$

Suppose that $h^{t'+1}$ is reachable from $h^{t'}$. Then, $a(t'+1)$ must be a maximin action, and therefore, Condition 5 (Independence of Irrelevant Experiences) says that for every $a \in A$ which is not a maximin action,

$$v^{(a,\phi)}(h^{t'+1}) = v^{(a,\phi)}(h^{t'}) < \underline{u} \text{ for all } \phi \in \Phi.$$

From Condition 4 (Minimal Evaluation) and the consistency of d with Γ, one gets that every $a \in A$ which is not a maximin action neither is the same as $a(t'+1)$ nor maximizes the subjective expected payoff, and therefore, satisfies

$$d(h^{t'+1})(a) = 0.$$

Recursively, for every $t > t'$ and every $h^t \in H^t$ which is reachable from $h^{t'}$, we can check similarly that for every $a \in A$ which is not a maximin action,

$$d(h^t)(a) = 0 \text{ and } v^{(a,\phi)}(h^t) < \underline{u} \text{ for all } \phi \in \Phi.$$

Next, we will prove Part (i) of this theorem. Fix $\xi \in (0,1]$ arbitrarily. From Part (ii), all we have to do is to prove that it holds at least with probability $1-\xi$ that there exists a positive integer s such that for every $a \in A$ which is not a maximin action,

$$d(h^s)(a) = 0 \text{ and } v^{(a,\phi)}(h^s) < \underline{u} \text{ for all } \phi \in \Phi.$$

Let $m^* \equiv |\Phi|$, $k^* \equiv |A|$, and denote $\Phi = \{\phi^1,\dots,\phi^{m^*}\}$ and $A = \{a^1,\dots,a^{k^*}\}$. Fix $t \geq 1$ and $h^t \in H^t$ arbitrarily.

We will define m^*k^* actions $a^{[y,z]} \in A$, $y \in \{1,\dots,m^*\}$, $z \in \{1,\dots,k^*\}$, below. Define

$$a^{[1,1]} \equiv a(t).$$

Recursively, for every $z \in \{2,\dots,k^*\}$, define $a^{[1,z]}$ as follows: Fix $k \in \{1,\dots,k^*\}$ arbitrarily. If for every $z' < z$,

$$a^k \neq a^{[1,z']} \text{ and } v^{(a^k,\phi^1)}(h^t) \geq \underline{u}(a^{[1,z']}),$$

for every $a \in A$ such that $a \neq a^{[1,z']}$ for all $z' < z$,

$$v^{(a^k,\phi^1)}(h^t) \geq v^{(a,\phi^1)}(h^t),$$

and for every $k' < k$ such that $a^{k'} \neq a^{[1,z']}$ for all $z' < z$,

$$v^{(a^k,\phi^1)}(h^t) > v^{(a^{k'},\phi^1)}(h^t),$$

then

$$a^{[1,z]} \equiv a^k.$$

If there is no such $k \in \{1,\dots,k^*\}$, then

$$a^{[1,z]} \equiv a^{[1,z-1]}.$$

For every $y \in \{2,\dots,m^*\}$, we define

$$a^{[y,1]} \equiv a^{[y-1,k^*]}.$$

Recursively, for every $y \in \{2,\dots,m^*\}$ and every $z \in \{2,\dots,k^*\}$, $a^{[y,z]}$ is defined as follows: Fix $k \in \{1,\dots,k^*\}$ arbitrarily. If for every $z' < z$,

$$a^k \neq a^{[y,z']} \text{ and } v^{(a^k,\phi^y)}(h^t) \geq \underline{u}(a^{[y,z']}),$$

for every $a \in A$ such that $a \neq a^{[y,z']}$ for all $z' < z$,

$$v^{(a^k,\phi^y)}(h^t) \geq v^{(a,\phi^y)}(h^t),$$

and for every $k' < k$ such that $a^{k'} \neq a^{[y,z']}$ for all $z' < z$,

$$v^{(a^k,\phi^y)}(h^t) > v^{(a^{k'},\phi^y)}(h^t),$$

then

$$a^{[y,z]} \equiv a^k.$$

If there is no such $k \in \{1,\dots,k^*\}$, then

$$a^{[y,z]} \equiv a^{[y,z-1]}.$$

Let s^* be the integer presented in Condition 6 (Uniform Adaptation), and let $\hat{k} \in \{1,\dots,k^*\}$ be the integer such that $a^{\hat{k}}$ is a maximin action but for every $k < \hat{k}$, a^k is not a maximin action. We will define $(a(\tau), v(\tau), \phi(\tau))_{\tau=t+1}^{t+m^*k^*s^*+1}$ as follows: Define

$$a(t + m^*k^*s^* + 1) = a^{\hat{k}},$$

$$v(t + m^*k^*s^* + 1) = \underline{u},$$

and

$$\phi(t + m^*k^*s^* + 1) = \phi^1.$$

For every $y \in \{1,\dots,m^*\}$, every $z \in \{1,\dots,k^*\}$, and every $s \in \{1,\dots,s^*\}$, define

$$a(t + (y-1)k^*s^* + (z-1)s^* + s) = a^{[y,z]},$$

$$v(t + (y-1)k^*s^* + (z-1)s^* + s) = \underline{u}(a^{[y,z]}),$$

and

$$\phi(t + (y-1)k^*s^* + (z-1)s^* + s) = \phi^y.$$

Let

$$H(h^t) \equiv \{h^{t+m^*k^*s^*+1} \in H^{t+m^*k^*s^*+1} : h^t \text{ is the sub-history of } h^{t+m^*k^*s^*+1} \text{ and } (a(\tau), v(\tau), \phi(\tau))_{\tau=t+1}^{t+m^*k^*s^*+1} \text{ is defined above}\}.$$

Conditions 5 (Independence of Irrelevant Experiences) and 6 (Uniform Adaptation) say that for every $h^{t+m^*k^*s^*+1} \in H(h^t)$, every $y \in \{1,\dots,m^*\}$, and every $z \in \{1,\dots,k^*\}$, if $(a,\phi) = (a^{[y',z']}, \phi^{y'})$ for some $y' \leq y$ and some $z' \in \{1,\dots,k^*\}$, then

$$v^{(a,\phi)}(h^{t+yk^*s^*+zs^*}) = \underline{u}(a),$$

and if a is not a maximin action and $\phi = \phi^{y'}$ for some $y' \leq y$, then

$$v^{(a,\phi)}(h^{t+yk^*s^*+zs^*}) < \underline{u}.$$

Hence, it holds for every $h^{t+m^*k^*s^*+1} \in H(h^t)$ that if a is a maximin action, then

$$v^{(a,\phi)}(h^{t+yk^*s^*+zs^*}) = \underline{u} \text{ for all } \phi \in \Phi,$$

and if a is not a maximin action, then

$$v^{(a,\phi)}(h^{t+m^*k^*s^*+1}) < \underline{u} \text{ for all } \phi \in \Phi,$$

and

$$d(h^{t+m^*k^*s^*+1})(a) = 0.$$

Here, this last equality is derived from the consistency of d with Γ and the fact that $a(t+m^*k^*s^*+1)$ is a maximin action.

Conditions 1 (Uniform Positive Lower Bound for $(p^{(\cdot)}, q^{(\cdot)})$), Condition 2 (Law of Inertia), and the consistency of d with Γ imply that every $h^{t+m^*k^*s^*+1} \in H(h^t)$ is reachable from h^t. Let $\varepsilon > 0$ be a positive real number which satisfies the inequalities in Conditions 1 and 3 (Uniform Positive Lower Bound). We must note that the probability conditional on h^t that $H(h^t)$ occurs is at least $\varepsilon^{m^*k^*s^*+1} > 0$, irrespective of how t and h^t are given. Hence, for every $\xi \in (0,1]$, by choosing ε small enough to satisfy $\varepsilon^{m^*k^*s^*+1} < \xi$, it holds at least with probability $1-\xi$ that there exists a period t such that $H(h^t)$ occurs in period $t+m^*k^*s^*+1$.

From these observations, one gets that for every $\xi \in (0,1]$, it holds at least with probability $1-\xi$ that there exists a period t such that if a is a maximin action, then

$$v^{(a,\phi)}(h^{t+yk^*s^*+zs^*}) = \underline{u} \text{ for all } \phi \in \Phi,$$

and if a is not a maximin action, then

$$v^{(a,\phi)}(h^{t+m^*k^*s^*+1}) < \underline{u} \text{ for all } \phi \in \Phi, \text{ and } d(h^{t+m^*k^*s^*+1})(a) = 0.$$

This, together with Part (ii), implies that Part (i) holds.

Q.E.D.

4. Subjective Games and Trivial Games

This section is the main part of this paper. We will investigate the situation of strategic conflict which is described by an n-person noncooperative game $G \equiv (N, A_1, \ldots, A_n, u_1, \ldots, u_n)$ as the objective game, where $N = \{1, \ldots, n\}$ is the finite set of players, A_i is the finite set of actions for player i, and $u_i: \times_{j \in N} A_j \to R$ is the payoff function for player i. Players a priori know the sets of actions $A_1, \ldots, A_n$, but have no prior knowledge about the payoff functions $u_1, \ldots, u_n$. Each player i instead has her own *subjective payoff function* $\hat{u}_i: \times_{j \in N} A_j \to R$. In distinction from *the objective game* G, *the subjective game* is defined by $\hat{G} = (N, A_1, \ldots, A_n, \hat{u}_1, \ldots, \hat{u}_n)$.

Players also have no prior knowledge about which actions the others will choose. Each player i instead has her own *subjective probability function* $\rho_i^{(a_i)}: A_{-i} \to R_+$ on the set of the other players' action profiles $A_{-i} \equiv \times_{j \neq i} A_j$, where $\sum_{a_{-i} \in A_{-i}} \rho_i^{(a_i)}(a_{-i}) = 1$. According to $\rho_i^{(a_i)}$, player i who chooses an action $a_i \in A_i$ expects the opponents to choose an $(n-1)$-tuple of actions $a_{-i} \in A_{-i}$ with probability $\rho_i^{(a_i)}(a_{-i})$.

We assume that each player i does not even know that the game with which she is confronted is a *simultaneous-move* game: She may believe incorrectly that the opponents choose actions *after* observing her choice of action. This is why we will allow player i's subjective probability function $\rho_i^{(a_i)}$ to depend on her own choice of action a_i.

Let $\Delta(A_i)$ be the set of all mixed actions for player i. Player i chooses an action among A_i according to a mixed action denoted by $\sigma_i \in \Delta(A_i)$. If player i maximizes the subjective expected payoff, then σ_i must satisfy

$$[\sigma_i(a_i) > 0] \Rightarrow [\sum_{a_{-i} \in A_{-i}} \rho_i^{(a_i)}(a_{-i}) \hat{u}_i(a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} \rho_i^{(a_i')}(a_{-i}) \hat{u}_i(a_i', a_{-i})$$

$$\text{for all } a_i' \in A_i].$$

The main purpose is to characterize the combination of the subjective game, the subjective probability functions, and the mixed actions $(\hat{G}, (\rho_i^{(\cdot)}, \sigma_i)_{i \in N})$, and to provide clear answers to several important questions such as in what way the subjective game is connected with the objective game. On the basis of the analysis of the long-run single-person decision making problems in Section 3, we will specify $(\hat{G}, (\rho_i^{(\cdot)}, \sigma_i)_{i \in N})$ in the following way: Each player i is *randomly matched* with $n-1$ opponents infinitely many times. In every period, player i plays an n-person noncooperative game which has the same sets of actions as G, together with the randomly matched opponents in this period.

Similarly to Section 3, a long-run single-person decision making problem for player i is defined by $D_i \equiv (A_i, \Omega_i, \Phi_i, H_i, u_i^{(\cdot)}, p_i^{(\cdot)}, q_i^{(\cdot)})$, where it is assumed that

$$\Omega_i = A_{-i},$$

and the associated maximin set of actions A_i^* is *a singleton,* that is,

$$A_i^* = \{a_i^*\}.$$

In a period t, player i chooses an action $a_i(t) \in A_i$, obtains the realized payoffs $v_i(t) \in R$ and signals $\phi_i(t) \in \Phi_i$, but can not observe the opponents' actions $\omega_i(t) \in A_{-i}$, where $v_i(t) = u_i^{(h^{t-1})}(a_i(t), \omega_i(t))$. In particular, the state $\omega_i(t)$ realized in each period t represents *the actions chosen by the opponents with whom player* i *is matched in this period.*

Similarly to Section 3, define $h_i^{t-1} = (a_i(\tau), \omega_i(\tau), v_i(\tau), \phi_i(\tau))_{\tau=1}^{t-1}$, H_i^{t-1}, and H_i. If a history $h_i^{t-1} \in H_i^{t-1}$ for player i up to period t was realized and player i chose an action $a_i(t) \in A_i$, then the randomly matched opponents in period t will be anticipated to choose $\omega_i(t) = a_{-i}$ with probability $p_i^{(h_i^{t-1})}(a_{-i}|a_i(t))$, and player i observes $\phi_i(t) = \phi_i$ with probability $q_i^{(h_i^{t-1})}(\phi_i|a_i(t), \omega_i(t))$.

Furthermore, similarly to Section 3, each player i is modeled as an inductive learning procedure (d_i, Γ_i), where $\Gamma_i \equiv (\Gamma_i^{(a_i)})_{a_i \in A_i}$: Player i chooses actions among A_i according to a decision rule $d_i : H \to \Delta(A_i)$, where $d_i(h_i^t)$ is independent of $(\omega_i(\tau))_{\tau=1}^{t-1}$ because player i can not observe the opponents' actions. For every $t \geq 1$ and every $h_i^{t-1} \in H_i^{t-1}$, player i chooses an action $a_i(t) = a_i$ with probability $d_i(h_i^{t-1})(a_i)$.

Let $\Gamma_i^{(a_i)} = ((v_i^{(a_i, \phi_i)})_{\phi_i \in \Phi_i}, \delta_i^{(a_i)})$ be player i's evaluation rule for $a_i \in A_i$, and let $\Gamma_i = (\Gamma_i^{(a_i)})_{a_i \in A_i}$ be player i's evaluation rule. Here, we assume that $|\Phi_i| \geq |A_{-i}|$ for all $i \in N$, and let $\pi_i : A_{-i} \to \Phi_i$ be a one-to-one function. By observing the signal $\phi_i(t) = \pi_i(\omega_i(t))$ in period t, player i is *convinced* that the opponents have almost surely chosen the actions $\omega_i(t) \in A_{-i}$. Hence, player i will regard the payoff evaluation $v_i^{(a_i, \pi_i(a_{-i}))}(h_i^{t-1})$ as the payoff which she can almost surely obtain when she chooses action $a_i \in A_i$ and the opponents chooses $a_{-i} \in A_{-i}$. Moreover, player i who chooses $a_i \in A_i$ anticipates that the randomly matched opponents in a period t will choose $a_{-i} \in A_{-i}$ with probability $\delta_i^{(a_i)}(h^{t-1})(a_{-i})$.

We will suppose that all players in N happen to meet together and play the objective game G in a period S, given an arbitrary profile of histories $(h_i^{S-1})_{i \in N}$ up to period $S-1$. We will specify the combination of the subjective game, the subjective probability functions, and the mixed actions, $(\hat{G}, (\rho_i^{(\cdot)}, \sigma_i)_{i \in N})$, by

$$\hat{u}_i(a_i, a_{-i}) \equiv v_i^{(a_i, \pi_i(a_{-i}))}(h_i^{S-1}) \text{ for all } i \in N, \text{ all } a_i \in A_i, \text{ and all } a_{-i} \in A_{-i},$$

$$\rho_i^{(a_i)}(a_{-i}) \equiv \delta_i^{(a_i)}(h_i^{S-1})(a_{-i}) \text{ for all } a_{-i} \in A_{-i},$$

and

$$\sigma_i(a_i) \equiv d_i(h_i^{S-1})(a_i) \text{ for all } a_i \in A_i.$$

We will require each player i's inductive learning procedure (d_i, Γ_i) to satisfy Conditions 1 through 6 as well as the consistency of (d, Γ).

Remark: Condition 1 implies that for every $j \neq i$ and every $a_j \in A_j$, there always exists a positive proportion of the population who will choose this action $a_j \in A_j$ when being matched with player i. One interpretation of Condition 1 is that there exists a positive proportion who are programmed to always choose $a_j \in A_j$ by birth. The other interpretation is that in every period, there always exists a positive proportion who were just born and begin by choosing actions at random.

Definition 4: A subjective game $\hat{G}$ is *trivial* if for every $i \in N$, there exists $a_i^+ \in A_i$ such that

$$\hat{u}_i(a_i^+, a'_{-i}) > \hat{u}_i(a_i, a_{-i}) \text{ for all } a_{-i} \in A_{-i}, \text{ all } a'_{-i} \in A_{-i}, \text{ and all } a_i \neq a_i^+.$$

We must note that for every $i \in N$, the action $a_i^+ \in A_i$ presented in Definition 4 is *strictly dominant*, i.e.,

$$\hat{u}_i(a_i^+, a_{-i}) > \hat{u}_i(a_i, a_{-i}) \text{ for all } a_{-i} \in A_{-i} \text{ and all } a_i \neq a_i^+,$$

and the n-tuple of the actions $(a_i^+)_{i\in N} \in \underset{i\in N}{\times} A_i$ is *Pareto-efficient* among the set of *pure* action profiles in a trivial subjective game $\hat{G}$, i.e., there exists no $(a_i)_{i\in N} \in \underset{i\in N}{\times} A_i$ such that $(a_i^+)_{i\in N} \neq (a_i)_{i\in N}$ and

$$\hat{u}_i(a_i^+, a_{-i}^+) \geq \hat{u}_i(a_i, a_{-i}) \text{ for all } i \in N.$$

We will provide the main theorem of this paper as follows.

Theorem 2: *Suppose that for every $i \in N$, $D = D_i$ satisfies Condition 1, $d = d_i$ satisfies Conditions 2 and 3, $\Gamma = \Gamma_i$ satisfies Conditions 4, 5 and 6, and $d = d_i$ is consistent with $\Gamma = \Gamma_i$. Then, for every $\xi \in (0,1]$, there exists a positive integer S^+ such that if $S \geq S^+$, then it holds with at least probability $1 - \xi$ that*

(i) the subjective game $\hat{G}$ is trivial,

(ii) for every $i \in N$, the action a_i^+ in Definition 4 is equivalent to the unique maximin action a_i^*,

(iii) for every $i \in N$, every $a_{-i} \in A_{-i}$, every $a'_{-i} \in A_{-i}$, and every $a_i \neq a_i^*$,

$$\hat{u}_i(a_i^*, a'_{-i}) \geq \underline{u}_i \equiv \min_{a''_{-i} \in A_{-i}} u_i(a_i^*, a''_{-i}) > \hat{u}_i(a_i, a_{-i}),$$

and

(iv) $\sigma_i(a_i^*) = 1$.

Proof of Theorem 2: Theorem 1 says that for every $\xi \in (0, 1]$, there exists a positive integer S^+ such that if $S \geq S^+$, then it holds at least with probability $1 - \xi$ that for every $i \in N$,

$$d_i(h_i^S)(a_i^*) = 1 \text{, and } v_i^{(a_i,\phi_i)}(h_i^S) < \underline{u}_i \text{ for all } a_i \neq a_i^* \text{ and all } \phi_i \in \Phi_i .$$

From the definition of the subjective game $\hat{G}$ and Condition 4 (Minimal Evaluation), one gets that for every $a_{-i} \in A_{-i}$, every $a'_{-i} \in A_{-i}$, and every $a_i \neq a_i^*$,

$$\hat{u}_i(a_i^*, a'_{-i}) = V_i^{(a_i^*)}(h_i^S) \geq \underline{u}_i > V_i^{(a_i)}(h_i^S) = \hat{u}_i(a_i, a_{-i}) .$$

This implies that $\hat{G}$ is trivial (Property (i)), and for every $i \in N$, the action a_i^+ presented in Definition 4 is equivalent to the maximin action a_i^* (Property (ii)). Since $\underline{u}_i = \min_{a''_{-i} \in A_{-i}} u_i(a_i^*, a''_{-i})$ for all $i \in N$, Property (iii) holds. It is straightforward from the definition of σ_i that Property (iv) holds.

Q.E.D.

From Theorem 2, we can obtain the following remarkable properties: The subjective game is totally *different* from the objective game. The subjective game is always *trivial*, and therefore, several famous games which have been intensively studied in the literature of game theory such as zero-sum games, prisoner-dilemma games, stag-hunt games, hawk-dove games, and coordination games can *not* be perceived as subjective games because these are not trivial games. The maximin action profile $(a_i^*)_{i \in N} \in \times_{i \in N} A_i$ in the objective game is *Pareto-efficient* among the set of pure action profile and *strictly dominant* in the subjective game. Of particular importance is that the strict dominance property holds *irrespective* of how the order of players' move is specified: In a trivial game $\hat{G}$, it holds that for every $i \in N$ and every function $\alpha : A_i \to A_{-i}$,

$$\hat{u}_i(a_i^*, \alpha(a_i^*)) > \hat{u}_i(a_i, \alpha(a_i)) \text{ for all } a_i \neq a_i^* .$$

This means that choosing the maximin action $a_i^* \in A_i$ is always *the best response* for player i against all correlated strategies for the opponents which may *condition* on player i's choice of action.

5. Several Examples

In this section, we will investigate several example of the objective games such as hawk-dove games, stag-hunt games, and modified coordination games.

Hawk-Dove Game: We have already examined the hawk-dove game in the introduction (see Figure 1.1 and 1.2). Here, (dove, dove) is the maximin action profile in the hawk-dove game as the objective game.

Stag-Hunt Game: In the stag hunt game as the objective game presented in Figure 3.1, (stag, stag) is the *payoff-dominant* Nash equilibrium.[13] If $x < 3/2$, then (stag, stag) is also the *risk-dominant equilibrium* in the sense of Harsanyi and Selten (1988).[14] If $x > 3/2$, then (hare, hare) is the risk-dominant equilibrium. The emergence of the risk-dominant equilibrium has been studied in the literature of evolutionary game theory (see Kandori et al. (1993) and Young (1993)). The subjective game in the long run is approximately expressed by Figure 3.2, where (hare, hare) is strictly dominant and also Pareto-efficient among the set of all mixed and correlated action profiles. As a result, players behave as being extremely *risk-averse*.

---

[13] A Nash equilibrium $(a_i) \in \times_{i \in N} A$ in $G = (N, (A_i, u_i))$ is *payoff-dominant* if for every Nash equilibrium $(a_i') \in \times_{i \in N} A_i$ and every $i \in N$, $u_i((a_j)) \geq u_i((a_j))$.

[14] A Nash equilibrium (a, a) in a 2×2 symmetric game such that $A_1 = A_2 = \{a, b\}$ is *risk-dominant* if $u_1(b,b) - u_1(a,b) < u_1(a,a) - u_1(b,a)$.

	stag	hare
stag	3, 3	0, x
hare	x 0	x, x

Figure 3.1: Stag Hunt Game
as the Objective Game
$(0 < x < 3)$

	stag	hare
stag	$L_1, \quad L_4$	$L_3, \quad x$
hare	$x, \quad L_2$	$x, \quad x$

Figure 3.2: The Subjective Game
$(L_h < x \; for \; h = 1,2,3,4)$

Coordination Game: In the *pure* coordination game as the objective game presented in Figure 4, there are multiple Nash equilibria, that is, (c,c), (d,d,), and the mixed action profile assigning c probability $1/3$ and d probability $2/3$. Since there are also multiple maximin actions (action "c" and action "d"), Theorem 2 gives no insight on the analysis of this objective game.

	c	d
c	2, 2	0, 0
d	0, 0	1, 1

Figure 4: Pure Coordination Game
as the Objective Game

Next, in the *modified* coordination game which is presented in Figure 5.1, there are also multiple Nash equilibria, i.e., (c,c), (d,d) and the mixed action profile which assigns c probability $\frac{1+2\beta-\gamma}{3-\gamma}$ and d probability $\frac{2-2\beta}{3-\gamma}$. Clearly, (c,c) is the payoff-dominant equilibrium. If $\beta-\gamma>1$, then (d,d) is the risk-dominant equilibrium, whereas if $\beta-\gamma<1$, then (c,c) is the risk-dominant equilibrium. According to the argument of Section 4, if $\beta>\gamma$, then the associated subjective game is approximated by Figure 5.2, in which the payoff-dominated equilibrium (d,d) in the objective game is strictly dominant. On the other hand, if $\beta<\gamma$, then the associated subjective game is approximated by Figure 5.3, in which the payoff-dominant equilibrium (c,c) in the objective game is strictly dominant. Here, we must note that if $0<\beta-\gamma<1$, then the strictly dominant equilibrium in the subjective game is not equal to the risk-dominant equilibrium in the objective game.

In a modified coordination game as the objective game, the *"spiteful"* action in the sense that a player's choosing this action makes the opponent's payoff worse than her own payoff will survive in the long run as the strictly dominant action in the associated subjective game.

	c	d
c	2, 2	γ, β
d	β, γ	1, 1

Figure 5.1: Modified Coordination Game
as the Objective Game
($\beta < 1, \gamma < 1$)

	c	d
c	L_1, L_4	L_3, β
d	β, L_2	1, 1

Figure 5.2: The Subjective Game
($\gamma < \beta < 1, L_h < \beta$ *for* $h = 1,2,3,4$)

	c	d
c	2, 2	γ, L_4
d	L_1, γ	L_2, L_3

Figure 5.3: The Subjective Game
($\beta < \gamma < 1$, $L_h < \gamma$ *for* $h = 1,2,3,4$)

6. Discussions

Several results in this paper might be regarded as being *negative* by the orthodox game theorists, because these imply that most of the famous games such as zero-sum games, prisoner-dilemma games, coordination games, stag-hunt games, and hawk-dove games can not be viewed as the subjective games. We, however, would like to emphasize that this paper should be regarded as the *benchmark* for the possible progress toward a theory of subjective games in the future.

We have assumed that a player is not convinced that the situation is recurrent, and she can not establish a firm experience-based belief about the situation in which she will seldom waver when observing unlikely events. This point was well expressed by Condition 6 (Uniform Adaptation). A real economic agent, however, sometimes establishes a firm experience-based belief as the *undoubted self-evident knowledge* which essentially regulates how to visualize the uncertain situation. In order to show that various types of games can be viewed as the subjective games, it would be inevitable to weaken Condition 6 and clarify what kind of experience-based beliefs will be regarded as being self-evident in the long run.

We can provide a scenario that players succeed to view any *non-trivial* game as the subjective game: Let us reconsider Case 3 in the example of recurrent situation presented in Section 2. Differently from Section 2, we will suppose here that the decision maker *is* convinced that the situation is recurrent. Then, in the similar way to the latter half of Case 1, one can get that in the long run the decision maker succeeds to make the subjective payoff function approximate the objective payoff function and to maximize the *objective* expected payoff with almost certainty. It might be easy for sophisticated readers to conjecture that this argument is also applicable to general class of multi-person recurrent situations, and that players who are convinced that the situation is recurrent will eventually come to equalize the subjective game with the objective game *irrespective* of how the objective game is given.

I think, however, that the next round in the theory of subjective game will be to investigate *non-recurrent* situations much more intensively from the view-points of bounded rationality, instead of recurrent situations: Actual players sometimes *categorize* various situations into a *limited* number of groups each of which are viewed as a particular form of subjective games such as trivial games, prisoner-dilemma games, coordination games, zero-sum games, hawk-dove games, stag-hunt games, and so on. These subjective games are *not* necessarily the same as the objective games. It would be quite substantial to answer the question of how players categorize situations, which is the matter of the more precise understanding of inductive learning.

We have assumed in this paper that a player is motivated by the maximization of the subjective expected payoff as well as the law of inertia, and have assumed that the probability of experimenting is negligible. However, as Selten (1978, 1990) have stressed, a real economic agent has *multiple* motivations some of which urge her to choose an action which neither maximizes the subjective expected payoff nor is the same as the action chosen in the last period. What motivates a player is undoubtedly one of the most important issues in the study of procedural rationality, which might also depend on the past experiences.[15]

The companion paper, Matsushima (1997b), reconsidered the example presented in Section 2, by assuming that the manager (the decision maker) behaves according to a particular form of *Markovian* learning procedure, and that she *experiments* with the risky project with a non-negligible probability, but that the probability of experimenting *decreases* as she continues to experience the loss of one hundred dollars. It was shown that the manager sometimes comes to believe as 'undoubted self-evident knowledge' that the risky project should *not* be enforced in the long run, even though this risky project is actually profitable.

We have assumed in the main body of this paper that a player can, not perfectly but almost perfectly, monitor the realized states, or the realized actions chosen by the opponents. This implies that players need to spend *very long time* in perceiving the situation as a trivial game, which leads us to pay attention to the question of what about *the middle run*. It might be conjectured that similarly to the case of perfect monitoring, players continue to monitor the realized states *correctly* until the middle run. Hence, in the middle run of a recurrent situation, players may succeed to make their subjective payoff functions equalize with the objective ones and to maximize their objective expected payoffs.

We have assumed in Section 3 that for every action $a \in A$, there exists the minimal payoff $\underline{u}(a) \equiv \min_{\omega \in \Omega} u^{(h^t)}(a, \omega)$ common to all histories. This assumption is quite restrictive in a class of situations, because the minimal payoff $\min_{\omega \in \Omega} u^{(h^t)}(a, \omega)$ may be *history-dependent*, and the set of the actions which maximizes the minimal payoff may also be *history-dependent* in general. The results of this paper can not necessarily be applied to these cases without substantial modifications.

---

[15] We have also assumed in the paper that players are *myopic*. It might be interesting to consider players who maximize their *long-run* subjective expected payoffs.

Finally, Theorem 1 in Section 3 says that if the set of maximin actions is a *singleton*, then the action which the decision maker chooses in the long run is *uniquely* determined. However, if there are *multiple* maximin actions, there still exists a variety of the decision maker's possible long run behaviors which satisfy the conditions required in this theorem. The other companion paper, Matsushima (1997c), provided an additional condition on a learning procedure, which guarantees the *uniqueness* of the decision maker's choice of action in the long run, even though there are multiple maximin actions.

7. Evolution and Learning

The recent studies of *evolutionary game theory in economic environments* are related to this paper.[16] Players are modeled as inductive learning procedures in order to learn to anticipate how the opponents will behave. Players may not even know the structure of the objective game. However, as we have already stressed in the introduction, most papers do not have dealt with how to learn to perceive the subjective game. Many papers in this literature have explored conditions under which the realizations of rational solution concepts such as rationalizability, Nash equilibrium, and its refinements can be *justified* also in evolutionary contexts.

Unlike our work, most papers in this literature have investigated only recurrent situations, and have assumed that a player can observe the actions chosen and the payoff obtained by *the other players* who did not match her but played the same game with their respective randomly matched opponents. This assumption plays the crucial role in justifying these rational solution concepts: A player always gets informative empirical data concerning *any* action, by observing the payoff obtained by some of the other players who *did* choose this action. This prevents the player from sticking to the improper pessimistic evaluations for the actions which are not maximin.[17]

Fudenberg and Kreps (1988) considered learning behaviors in extensive form games *without* this assumption, and explained the similar point to ours that players may choose as the stationary point a strategy profile which is *not* a Nash equilibrium, because they have serious lack of experiences on the situations which take place after some players' deviation.

Börgers and Sarin (1996) reconsidered a version of the learning model addressed by Bush and Mosteller (1951) in a game-theoretic context, and showed that in a *continuous time limit* this learning model converges to the *replicator dynamics* of evolutionary game theory, which makes players maximize the objective expected payoffs in the long run. This convergence property, however, is rather exceptional and relies crucially on the assumption that players adjust *very slowly*. This assumption inevitably makes this

---

[16] For the surveys of this literature, see Fudenberg and Levine (1995), Kandori (1997), Matsui (1995), Weibull (1996), Vega-Redondo (1996), Samuelson (1997), van Damme (1994) and so on.

[17] It might be also interesting to consider the case in which a player can not observe the choices of actions and the payoffs obtained by the other players, but can hear the *opinions* of the other players about how they subjectively evaluate these actions.

convergence take place in the *ultra-long* run.[18]

Blume and Easley (1992) studied an evolutionary model of single-person decision making with payoff uncertainty and showed that *the expected log-fitness maximizer* dominates the population in the long run, which differs from the expected fitness maximizer. Their result also expresses the similar point to ours that the actions having a possibility to result in very low payoffs tend to be weeded out even though these expected payoff are the highest.

---

[18] See Arthur (1993) and Marimon (1997, Section 3.3).

References

Abreu, D. and A. Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica* 56, 1259-1282.

Arthur, B. (1993): "On Designing Economic Agents that Behave Like Human Agents," *Journal of Evolutionary Economics* 3, 1-22.

Aumann, R. (1992): "Rationality and Bounded Rationality," Nancy L. Schwartz Lecture, Kellogg School of Management, Northwestern University.

Aumann, R. and A. Brandenburger (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica* 63.

Binmore, K. (1987, 1988): "Modeling Rational Players I and II," *Economics and Philosophy* 3 and 4, 179-214 and 9-55.

Blume, L. and D. Easley (1992): "Evolution and Market Behavior," *Journal of Economic Theory* 58, 9-40.

Börgers, T. and R. Sarin (1996): "Learning through Reinforcement and Replicator Dynamics," mimeo.

Bush, R. and F. Mosteller (1951): "A Mathematical Model for Simple Learning," *Psychological Review* 58, 312-323.

Conlisk, J. (1996): "Why Bounded Rationality ?," *Journal of Economic Literature* 36, 669-700.

Dekel, E. and F. Gul (1996): "Rationality and Knowledge in Game Theory," in Kreps, D. and K. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications Vol. 1*, Cambridge University Press.

Fisher, F. (1989): "Games Economists Play: A Noncooperative View," *RAND Journal of Economics* 20, 113-137.

Fudenberg, D. and D. Kreps (1988): "A Theory of Learning, Experimentation, and Equilibrium in Games," mimeo.

Fudenberg, D. and D. Levine (1995): Theory of Learning in Games, mimeo.

Gilboa, I. (1988): "The Complexity of Computing Best Response Automaton in Repeated Games," *Journal of Economic Theory* 45, 342-352.

Hammerstein, P. and R. Selten (1994): "Game theory and Evolutionary Biology," in Aumann, R. and S. Hart (eds.), *Handbook of Game Theory with Economic Applications, Vol. 2*, Amsterdam: Elsevier.

Harsanyi, J. (1967, 1968): "Games with Incomplete Information Played by Bayesian Players," *Management Science* 14, 159-182, 320-334, 486-502.

Harsanyi, J. and R. Selten (1988): *A General Theory of Equilibrium Selection in Games*, Cambridge: MIT Press.

Kandori, M. (1996): "Evolutionary Game Theory in Economics," in Kreps, D. and K. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications Vol. 1*, Cambridge University Press.

Kandori, M., G. Mailath, and R. Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica* 61, 29-56.

Kaneko, M. and A. Matsui (1997): "Inductive Game Theory: Discrimination and Prejudices Part I," Discussion paper No. 711, University of Tsukuba.

Lipman, B. (1991): "How to Decide How to Decide How to ... : Modeling Limited Rationality," *Econometrica* 59, 1105-1126.

Marimon, R. (1997): "Learning from Learning in Economics," in Kreps, D. and K. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications Vol. 1*, Cambridge University Press.

Matsui, A. (1995): "On Cultural Evolution: Social Norms, Rational Behavior, and Evolutionary Game Theory," mimeo.

Matsushima, H. (1997a): "Bounded Rationality in Economics: A Game Theorist's View," Discussion Paper 97-F-10, University of Tokyo, forthcoming in *Japanese Economic Review* Vol. 48, No. 3.

Matsushima, H. (1997b): "Procedural Rationality and Inductive Learning II: Involving Experimentation," mimeo (in preparation).

Matsushima, H. (1997c): "Procedural Rationality and Inductive Learning III: Strict Maximin Actions," mimeo (in preparation).

Maynard Smith, J. (1982): *Evolution and the Theory of Games*, Cambridge University Press.

North, D. (1990): *Institutions, Institutional Change and Economic Performance*, Cambridge University Press.

Pelzman, S. (1991): "The Handbook of Industrial Organization: A Review Article," *Journal of Political Economy* 99, 201-217.

Rubinstein, A. (1991): "Comments on the Interpretation of Game Theory," *Econometrica* 59, 909-924.

Samuelson, L. (1997): *Evolutionary Games and Equilibrium Selection*, Cambridge: MIT press.

Savage, L. (1954): *The Foundation of Statistics*, New York: Wiley.

Schelling, T. (1960): *The Strategy of Conflict*, Cambridge: Harvard University Press.

Selten, R. (1978): "The Chain Store Paradox," *Theory and Decision* 9, 127-159.

Selten, R. (1990): "Bounded Rationality," *Journal of Institutional and Theoretical Economics* 146, 649-658.

Selten, R. (1991): "Evolution, Learning and Economic Behavior," *Games and Economic Behavior* 3, 3-24.

Simon, H. (1976): "From Substantive to Procedural Rationality," in S. J. Latsis ed. *Methods and Appraisal in Economics*, Cambridge University Press.

Simon, H. (1982): *Models of Bounded Rationality,* MIT Press.

van Damme, E. (1987): *Stability and Perfection of Nash Equilibria*, Berlin: Springer Verlag, Chapter 9.

van Damme, E. (1994): "Game Theory: The Next Stage," mimeo.

Vega-Redondo (1996): *Evolution, Games, and Economic Behavior*, Oxford University Press.

Young. P. (1993): "The Evolution of Conventions," *Econometrica* 61, 57-84.

Weibull, J. (1996): *Evolutionary Game Theory,* Cambridge: MIT Press.

Winter, S. (1986): "Comments on Arrow and Lucas," *Journal of Business* 59, 427-434.