

[THIS VERSION: September 27, 2005]

**GROWTH OF STRATEGY SETS, ENTROPY, AND  
NONSTATIONARY BOUNDED RECALL**

ABRAHAM NEYMAN<sup>†</sup> AND DALJIRO OKADA<sup>‡</sup>

---

<sup>†</sup> Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904,  
Israel.

EMAIL: aneyman@math.huji.ac.il.

<sup>‡</sup> Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, New Jersey  
08901-1248, USA.

EMAIL: okada@econ.rutgers.edu.

ABSTRACT. One way to express bounded rationality of a player in a game theoretic models is by specifying a set of feasible strategies for that player. In dynamic game models with finite automata and bounded recall strategies, for example, feasibility of strategies is determined via certain complexity measures: the number of states of automata and the length of recall. Typically in these models, a fixed finite bound on the complexity is imposed resulting in finite sets of feasible strategies. Also as a consequence, the number of distinct feasible strategies in any subgame, as well as the number of distinct strategies induced in the first  $T$  stages, is held finite.

In this paper, we initiate an investigation into a notion of feasibility that reflects varying degree of bounded rationality over time. Such concept must entail properties of a strategy, or a set of strategies, that depend on time. Specifically, we associate to each subset, say  $\Psi_i$ , of the full (theoretically possible) strategy set a function  $\psi_i$  from the set of positive integers to itself. The value  $\psi_i(t)$  represents the number of strategies in  $\Psi_i$  that are distinguishable in the first  $t$  stages. The set  $\Psi_i$  may contain infinitely many strategies, but it can differ from the fully rational case in the way  $\psi_i$  grows reflecting a broad implication of bounded rationality that may be alleviated, or intensified, over time.

We examine how the growth rate of  $\psi_i$  affects equilibrium outcomes of repeated games. In particular, we derive an upper bound on the individually rational payoff of the two-person repeated games where player 1, with a feasible strategy set  $\Psi_1$ , plays against a fully rational player 2. We will show that the derived bound is tight in that a specific, and simple, set  $\Psi_1$  exists that achieves the upper bound.

As a special case, we study repeated games with non-stationary bounded recall strategies where the length of recall is allowed to vary in the course of the game. We will show that a player with bounded recall can guarantee the minimax payoff of the stage game even against a player with full recall so long as he can remember, at stage  $t$ , at least  $K \log t$  stages back for some constant  $K > 0$ . Thus, in order to guarantee the minimax payoff, it suffices to remember only a vanishing fraction of the past.

*(Journal of Economic Literature Classification Number:)*

KEYWORDS: Bounded Rationality, Strategy Set Growth, Strategic Complexity, Nonstationary Bounded Recall, Repeated Games, Entropy

## 1. INTRODUCTION

Many social (economic, political etc.) interactions have been modeled as formal games. The idea that players in a game are rational is reflected in several aspects of the model as well as in the analysis performed (optimization, equilibrium). When a game theorist employs a particular solution concept, there is an implicit understanding that players optimize or find a best response to others' actions from the specified set of strategies. Aside from the assumption that the players can perform computations necessary for such tasks, it is also assumed that players can carry out any strategy in the specified strategy set should he choose to play it. While this latter assumption may seem innocuous in a model where a few strategies are available to each player<sup>1</sup>, e.g. prisoner's dilemma and the battle of the sexes, it may be criticized as being unrealistically rational in more complex models where theoretical definition of strategy leads to a strategy set that contains a large number of choices, many of which are impractically complex.

A case in point is models of dynamic interaction including repeated games in its most basic formulation. In repeated games, a strategy is a set of history-contingent plan of actions. Even when the underlying stage game contains only a few possible actions, the number and complexity of histories quickly grows as time passes. Consequently, the set of strategies contains a large number of elements, and many of them requires capability to process arbitrarily complex history for their implementation.

The idea that the assumption of fully, or unbounded, rational players is unrealistic is not new (Simon (1955), Simon (1972), Aumann (1981), Aumann (1997)). There have been many attempts to model feasible (implementable) set of strategies that reflects some aspects of bounded rationality of players. Finite automata, bounded recall and Turing machines are a few of the approaches taken. These models are useful because they provide us with quantitative measures of complexity of strategies, e.g., the number of states of automata and the length of recall.<sup>2</sup>

---

<sup>1</sup>See however Anderlini (1990).

<sup>2</sup>Variants of complexity measure associated with a Turing machine include the number of bits needed to implement a strategy by Turing machines with bounded amount of tape (Stearns (1997)), and algorithmic or Kolmogorov complexity (Lacôte, Neyman).

Existing literature on bounded complexity in repeated games considers models where the complexity of strategies is fixed in the course of a long interaction. In the case of finite automata and bounded recall (e.g. Ben-Porath (1993), Lehrer (1988)), a single integer - the number of states or the length of recall - fully describes the set of feasible strategies. As a consequence, the set of feasible strategies, e.g. those implemented by finite automata of a fixed size, is finite. Moreover, the number of distinct feasible strategies in any subgame as well as the number of distinct strategies in the first  $T$  stages of the interaction is bounded. While this literature has supplied significant insights and formal answers to questions such as “when is having a higher complexity advantageous?” (op. cit.) or “when does bounded complexity facilitate cooperation?” (e.g. Neyman (1985), Neyman (1997)), we argue below that it would be fruitful to extend the analysis to include a salient feature and an implication of bounded rationality in dynamic decision making that are not captured by the existing approaches.

An important feature of an economic decision maker (consumer, firm, government, trade and labor union etc.) is described by its set of feasible decision rules. These rules, strategies or policies, are neither unimaginably complex or mindlessly simple. Nor is the set of feasible decision rules fixed over time. Technological progresses inevitably influence the sophistication and efficiency of handling information necessary to determine the behavior of these agents. Firms make investments in order to update technology and change organizational structure in an effort to improve their abilities to process information and arrive at better decisions, e.g. efficient allocation of resources within firms or streamlining of decision making in an uncertain, often strategic, environment. Such changes bring about the transformation of the set of possible decision rules over time.

A decision rule, in its abstract formulation, is a rule, or a function, that transforms information into actions. Information has to be, first of all, translated into a form (language) that can be communicated (in case of an organization or a team) and then interpreted by the decision maker. In economic applications, information relevant to decision making is often some data, signal, message and history etc., that modifies the decision maker’s perception about the environment including other decision makers’ actions and information they themselves possess. As the

ability to process such information improves (thereby making the agent capable of recognizing her environment in finer details), the decision maker would become able to implement more flexible and versatile decision rules.

Internet commerce offers one of many cases that are pertinent to this argument. Internet vendors collect detailed information on buying habits of consumers. An investment in technologies to collect such information, or expenditures on purchasing such information from a third party, enables the sellers to devise more flexible - customer specific - marketing strategies that would otherwise not be feasible. Since competing companies are likely to utilize similar information, this type of investment would also give the company an ability to gauge its rivals' strategic capabilities. Thus we need a model of an agent, or a player, whose strategic capabilities change over time.

As argued in the beginning, complexity of repeated games as a model of interactive decision making stems, in part, from the richness of strategies that the theory allows players to choose from. The number of theoretically possible strategies is double-exponential in the number of repetitions. This is due to the fact that the number of histories grows exponentially with the number of repetitions and also that we count strategies that map histories into actions in all possible ways. Some, in fact most,<sup>3</sup> strategies are too complicated to admit a short and practically implementable description: a short description of a strategy requires an efficient encoding of histories, but some histories may have no shorter descriptions than simply writing them out in their entirety. These considerations motivate research on bounded rationality in long-term interaction in general, and, various measures of complexity of implementing strategies and its effects on equilibrium outcomes in particular.

Our aim in this paper is to take a first step into formalizing the idea of temporal change in the degree of bounded rationality and examining its consequences in long-term interactions. Thus, at the conceptual level, our motivation may be paraphrased as follows. Players with bounded rationality are limited by the set of feasible strategies, but computational resource available to the players may expand or contract over time. As a consequence, the limitation would vary over time and, in

---

<sup>3</sup>For instance, if a feasible set of strategies contains  $K$  distinct strategies, then one needs close to  $\log K$  bits (for sufficiently large  $K$ ) to encode most of them.

particular, there may not be a finite upper bound on complexity of strategies for the entire horizon of the game. Such consideration on a more general aspect of bounded rationality cannot be captured by a model with a finite set of feasible strategies. Thus we are led to considering in general a feasible set consisting of infinitely many strategies. The question that arises is then “What are the characteristics of an infinite strategy set that (1) may be derived from an explicit description (e.g. by means of a complexity measure) of a feasible strategy set, and, (2) can be used to provide bounds on equilibrium outcomes?”

A common feature of feasible strategy sets described by means of any complexity measure is that it contains a fewer elements than the fully rational case. As we take aim at a temporal aspect of an infinite strategy set, we consider here how the number of strategies induced in the first  $t$  stages of the game grows. Specifically, we associate to each subset, say  $\Psi_i$ , of the full (theoretically possible) strategy set a function  $\psi_i$  from the set of positive integers to itself. The value  $\psi_i(t)$  represents the number of strategies in  $\Psi_i$  that are distinguishable in the first  $t$  stages. The feasible strategy set  $\Psi_i$  may contain infinitely many strategies, but it can differ from the fully rational case in the way  $\psi_i$  grows reflecting a broad implication of bounded rationality that may vary over time.<sup>4</sup> To be more precise, for each  $t$ , let  $\Psi_i(t)$  be the projection of  $\Psi_i$  to the first  $t$  stages of the game. Then  $\psi_i(t)$  is the number of equivalence classes of strategies in  $\Psi_i(t)$ . If  $\Psi_i$  contains all theoretically possible strategies, then, as mentioned in the beginning,  $\psi_i(t)$  is double-exponential in  $t$ . Thus it is of interest to study how outcomes of repeated games are affected by various conditions on the rate of growth of  $\psi_i(t)$ .

Since no structure is imposed on the strategies that belong to  $\Psi_i$ , it appears to be difficult, if not impossible, to derive results purely on the basis of how  $\psi_i(t)$  grows. For this reason, and as a first undertaking in this line of research, we will study a simple case of two-person repeated games in which player 1 with a feasible set  $\Psi_1$  plays against fully rational player 2. The payoff in the repeated games is

---

<sup>4</sup>In this paper, the feasible set  $\Psi_i$ , and hence the growth of the function  $\psi_i$ , is exogenously given. We recognize the importance of studying models where players may invest in order to expand their strategic possibilities thereby endogenizing the growth rate of  $\psi_i$ . This certainly deserves further research. The work reported here provides limits to what can and cannot be achieved by such choice.

the long-run average of the stage payoffs. In this setup we will show that there is a continuous function  $U : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that player 1 cannot guarantee more than  $(\text{cav } U)(\gamma)$ , the concavification of  $U$ , whenever  $\psi_1(t)$  grows at most as fast as  $2^{\gamma t}$ . Moreover, this bound is tight. It will be seen that the function  $U$  is defined using the concept of entropy and has the property that  $U(0)$  is the maximin value of the stage game in pure actions and  $U(\gamma)$  is the usual minimax value for sufficiently large  $\gamma$ .

As a concrete case of an infinite feasible strategy set arising from a complexity consideration, we will study nonstationary bounded recall strategies. This is a model of a player whose memory of the past varies over time and hence it is an extension of classical stationary bounded recall strategies. As a direct consequence of a theorem mentioned above, we will show that a player with nonstationary bounded recall can only guarantee the maximin payoff in pure actions of the stage game if the size of his recall is less than  $K_0 \log t$  at stage  $t$  for some constant  $K_0 > 0$ . In addition, we will show that there is a constant  $K_1 > K_0$  such that, so long as his recall is at least  $K_1 \log t$  at stage  $t$ , the minimax value of the stage game can be guaranteed. Hence, in order to secure the value of the stage game even against a player with full recall, one needs to remember a long enough yet still only a vanishing fraction of the past history.

In order to avoid possible confusion, we point out that, as is standard in the literature, we consider mixed strategies so long as its support lies in the set of feasible pure strategies. A possible interpretation of mixed strategies in games in general is that they are distributions of pure strategies in a population of potential players. In the context of games we analyze in this paper, a fully rational player faces one of the players randomly drawn from this population. Thus a mixed strategy of her opponent reflects the uncertainty that she faces as to which feasible pure strategy is employed by this particular opponent.

We will set the notation used throughout the paper and formalize the idea of the growth of strategy sets in Section 2. Some examples, including nonstationary bounded recall strategies, will also be discussed in this section. Section 3 contains some results on the values of two-person repeated games where a player with bounded rationality plays against a fully rational player. As mentioned above these

results are based purely on the rate of growth of strategy sets regardless of which strategies they contain. In Section 4, nonstationary bounded recall strategies are examined.

## 2. GROWTH OF STRATEGY SETS

Let  $G = (A_i, g_i)_{i \in I}$  be a finite game in strategic form. The set of player  $i$ 's mixed actions is denoted by  $\Delta(A_i)$ . Henceforth we refer to  $G$  as a stage game.

In the repeated version<sup>5</sup> of  $G$ , written  $G^*$ , a pure strategy of a player is a rule that assigns an action to each history. A history by definition is a finite string of action profiles (including the null string which is denoted by  $\epsilon$ ). Thus the set of all histories is  $A^* = \bigcup_{t=0}^{\infty} A^t$  where  $A = \prod_{i \in I} A_i$  and  $A^0 = \{\epsilon\}$ . A pure strategy of player  $i$  is a mapping

$$\sigma_i : A^* \rightarrow A_i.$$

Let  $\Sigma_i$  be the set of all pure strategies of player  $i$ . The set of mixed strategies of player  $i$  is denoted by  $\Delta(\Sigma_i)$ .

We say that two pure strategies of player  $i$ ,  $\sigma_i$  and  $\sigma'_i$ , are equivalent up to the  $t$ -th stage if, for every profile of other players' strategies  $\sigma_{-i}$ , the sequence of action profiles induced by  $(\sigma_i, \sigma_{-i})$  and  $(\sigma'_i, \sigma_{-i})$  are identical up to, and including, stage  $t$ . If two strategies are equivalent up to the  $t$ -th stage for every  $t$ , then we simply say they are equivalent. Equivalence between two mixed strategies is defined similarly by comparing the induced distributions over sequence of action profiles.

Let us denote by  $m_i$  the number of actions available to player  $i$ , i.e.,  $m_i = |A_i|$ , and let  $m = \prod_{i \in I} m_i = |A|$ . We note first that the number of strategies available to player  $i$  in the first  $t$  stages of a repeated game is<sup>6</sup>

$$m_i^{m^0} \times \cdots \times m_i^{m^{t-1}} = m_i^{\frac{m^t - 1}{m - 1}}.$$

This number is double exponential in  $t$ .

Suppose that player  $i$  has access to a set of strategies,  $\Psi_i \subset \Sigma_i$ . This would be the case, for example, when there is limitations on some aspects of complexity

<sup>5</sup>In this paper we consider the most basic model of repeated games, i.e., ones with complete information and perfect monitoring.

<sup>6</sup>The number of reduced strategies available to player  $i$  in the first  $t$  stages is  $m_i^{\frac{m^t - 1}{m - 1}}$  where  $m_{-i} = \prod_{j \neq i} m_j$ .



of his strategies. For each positive integer  $t$ , let  $\Psi_i(t)$  be formed by identifying strategies in  $\Psi_i$  that are equivalent up to the  $t$ -th stage.<sup>7</sup> Let  $\psi_i(t)$  be the number of elements in  $\Psi_i(t)$ . Any consideration on strategic complexity gives rise to some strategy set  $\Psi_i$  and thus limitation on the rate of growth of  $\psi_i(t)$ . For example, if player  $i$ 's feasible strategies are described by finite automata with a fixed number of states, then  $\Psi_i$  is a finite set and  $\Psi_i(t) = \Psi_i$  for all sufficiently large<sup>8</sup>  $t$ . In this case  $\psi_i(t) = O(1)$ . Below we illustrate some examples of feasible strategy sets with various rate of growth of  $\psi_i(t)$ .

**Example 1.** In this example we provide a framework for nonstationary bounded recall strategies that we will examine in detail in Section 4. Recall that a stationary bounded recall strategy of size  $k$  is a strategy that depends only on at most the last  $k$ -terms of the history. More precisely, for each pure strategy  $\sigma_i \in \Sigma_i$ , define a strategy  $\sigma_i \bar{\wedge} k : A^* \rightarrow A_i$  by

$$(\sigma_i \bar{\wedge} k)(a_1, \dots, a_t) = \begin{cases} \sigma_i(a_1, \dots, a_t) & \text{if } t \leq k, \\ \sigma_i(a_{t-k+1}, \dots, a_t) & \text{if } t > k. \end{cases}$$

The set of stationary bounded recall strategies of size  $k$  is denoted by  $\bar{\mathbf{B}}_i(k)$ , i.e.

$$\bar{\mathbf{B}}_i(k) = \{\sigma_i \bar{\wedge} k : \sigma_i \in \Sigma_i\}.$$

It is clear that the number of distinct strategies, i.e. the number of equivalence classes, in  $\bar{\mathbf{B}}_i(k)$  is at most the number of distinct functions from  $\bigcup_{\ell=0}^k A^\ell$  to  $A_i$  which is of the order  $m_i^{O(m^k)}$ .

Now consider a function  $\kappa : \mathbb{N} \rightarrow \mathbb{N} \cup \{0\}$  with  $\kappa(t) \leq t - 1$ . For each  $t \in \mathbb{N}$ , the value  $\kappa(t)$  represents the length of recall at stage  $t$ . A  $\kappa$ -recall strategy of player  $i$  is a pure strategy that plays like a stationary bounded recall strategy of size  $k$  whenever  $\kappa(t) = k$  regardless of the time index  $t$ . Formally, for each  $\sigma_i \in \Sigma_i$  define a strategy  $\sigma_i \wedge \kappa : A^* \rightarrow A_i$  by

$$(\sigma_i \wedge \kappa)(a_1, \dots, a_t) = \sigma_i(a_{t-\kappa(t)+1}, \dots, a_t).$$

<sup>7</sup>If two strategies in  $\Psi_i$  are equivalent, then they are never distinguished in  $\Psi_i(t)$  for any  $t$ . So the reader may consider  $\Psi_i$  to be the set of equivalence classes of strategies.

<sup>8</sup>In fact, this holds for all  $t \geq k^2$  and  $2^{c_1 k \log k} \leq |\Psi_i| \leq 2^{c_2 k \log k}$  where  $k$  is the bound on the number of states of automata and  $c_1$  and  $c_2$  are positive constants.

Observe that in this definition player  $i$  must take the same action at stage  $t$  and  $t'$  where  $\kappa(t) = \kappa(t') = k$  so long as he observes the same sequence of action profiles in the last  $k$  stages. Thus, the set of  $\kappa$ -recall strategies is

$$\mathbf{B}_i(\kappa) = \{\sigma_i \wedge \kappa : \sigma_i \in \Sigma_i\}.$$

Set  $\Psi_i = \mathbf{B}_i(\kappa)$ . Then from its definition it is clear that there is a canonical embedding of  $\Psi_i$  into  $\prod_{k \in \kappa(\mathbb{N})} \bar{\mathbf{B}}(k)$  as well as a canonical embedding of  $\Psi_i(t)$  into  $\prod_{k \in \kappa(\{1, \dots, t\})} \bar{\mathbf{B}}(k)$  for each  $t$ . Hence

$$\psi_i(t) \leq \prod_{k \in \kappa(\{1, \dots, t\})} m_i^{m^k} \leq m_i^{c m^{\bar{\kappa}(t)}}$$

for some constant  $c$  (in fact,  $c = m/(m-1)$ ) where  $\bar{\kappa}(t) = \max_{s \leq t} \kappa(s)$ .

**Example 2.** A strategy of player  $i$  is said to be oblivious (O'Connell and Stearns (1999)) if it depends only on the history of his own actions. That is,  $\sigma_i : A^* \rightarrow A_i$  is oblivious if  $\sigma_i((a_{i1}, a_{-i1}), \dots, (a_{it}, a_{-it}))$  is independent of  $a_{-i1}, \dots, a_{-it}$ . The set of oblivious strategies of player  $i$  is denoted by  $\mathbf{O}_i$ . Every oblivious strategy induces a sequences of player  $i$ 's actions. Also, any sequence of player  $i$ 's actions can be induced by an oblivious strategy. So the set of equivalence classes of strategies in  $\mathbf{O}_i$  can be identified with the set of sequences of player  $i$ 's actions,  $A_i^\infty$ . Hence if  $\Psi_i = \mathbf{O}_i$ , then  $\Psi_i(t)$  is identified with  $A_i^t$  and so  $\psi_i(t) = m_i^t$ . For each sequence  $\mathbf{a} = (a_{i1}, a_{i2}, \dots) \in A_i^\infty$ , we denote by  $\sigma_i(\mathbf{a})$  the oblivious strategy that takes action  $a_t$  at stage  $t$  regardless of the past history.

In all the examples that follow, consider a two person game in which each player has two actions,  $A_1 = A_2 = \{0, 1\}$ .

**Example 3.** For each integer  $k \geq 0$ , define a strategy  $\sigma_1^{(k)}$  as follows. For each history  $h$ , let  $N(1|h)$  be the number of times player 2 chose action 1 in  $h$ .

$$\sigma_1^{(k)}(h) = \begin{cases} 1 & \text{if } N(1|h) \geq k \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\Psi_1 = \{\sigma_1^{(0)}, \sigma_1^{(1)}, \dots\}$ . Then  $\Psi_1(t) = \{\sigma_1^{(0)}, \dots, \sigma_1^{(t-1)}\}$  and  $\psi_i(t) = t$ .

**Example 4.** A prefix of a history  $h = (h_1, \dots, h_t)$  is any of its initial segment  $h' = (h_1, \dots, h_s)$ ,  $s \leq t$ . A set of histories  $L \subset \bigcup_{t=1}^{\infty} H_t$  is said to be prefix-free

if no element of  $L$  is a prefix of another. For each positive integer  $t$ , let  $L(t) = L \cap (H_1 \cup \dots \cup H_{t-1})$ ;  $L(t)$  is prefix-free and  $L(t) \subset L(t+1)$ . Define a strategy  $\sigma_1^L$  as follows.

$$\sigma_1^L(h_1, \dots, h_t) = \begin{cases} 1 & \text{if } (h_1, \dots, h_s) \in L \text{ for some } s \leq t, \\ 0 & \text{otherwise.} \end{cases}$$

This is a generalization of the trigger strategy:  $\sigma_1^L$  takes action 1 forever as soon as a history in  $L$  occurs. Let  $\mathcal{L}$  be the class of all prefix-free sets of histories. Take a subset  $\mathcal{M}$  of  $\mathcal{L}$  and define  $\Psi_1$  to be the set of player 1's strategies  $\sigma_1^M$  with  $M \in \mathcal{M}$ . Let us examine  $\Psi_1(t)$  and  $\psi_1(t)$ .

It is easy to verify that, for any  $L$  and  $M$  in  $\mathcal{L}$ ,  $\sigma_1^L$  and  $\sigma_1^M$  are equivalent up to the  $t$ -th stage whenever  $L(t) = M(t)$ . Then we have<sup>9</sup>  $\psi_1(t) \leq |\mathcal{M}(t)|$  where  $\mathcal{M}(t) = \{M(t) : M \in \mathcal{M}\}$ .

Examples of  $\mathcal{M}$  can be constructed so that the corresponding function  $\psi_1(t)$  is, e.g.,  $O(t^p)$  for a given  $p \geq 1$ , or  $O(2^{\alpha t})$  for  $0 < \alpha < 1$ .

### 3. GAMES AGAINST A FULLY RATIONAL PLAYER

We now derive a few consequences of bounded rationality implied by a growth rate of  $\psi_i(t) = |\Psi_i(t)|$ . We emphasize that the nature of the feasible strategy set  $\Psi_i$  is completely arbitrary. It may include infinitely many strategies and also the strategies that cannot be represented by any finite state machines or finitely bounded recall.

Various forms of the folk theorem in repeated games assert that any feasible payoff vector that gives each player at least his individually rational payoff can be an equilibrium outcome of the repeated game. Often, the individually rational payoff level for player  $i$  is the minimax value, say  $v^i$ , of the stage game. This is so, for example, in the standard repeated game played by fully rational players (Aumann and Shapley (1994), Fudenberg and Maskin (1986)) because (1) for any strategy profile of the other players, player  $i$  has a strategy that yields him at least  $v^i$  in the long run, and (2) the other players do have strategies that hold player  $i$ 's long run payoff to  $v^i$ . It is thus important to note that individually rational payoff for a player depends on what strategies are available to him as well as to others.

---

<sup>9</sup>Some histories are not compatible with the strategy, hence the inequality.

In models of repeated games where the sets of feasible strategies are specified via bounds on some complexity measure, and therefore they differ from the fully rational case, it is essential to know the relationship between the complexity bounds and individually rational payoffs before proceeding to the question of equilibria. In fact, once individually rational payoffs are characterized, and strategies that achieve such payoffs are found, versions of the folk theorem follow in a relatively straight forward manner (Lehrer (1988), Ben-Porath (1993)). The reader will see that this is the case in the next section on nonstationary bounded recall.

Thus, our focus in this, and the next, section will be what payoff a player with bounded rationality, implied by a specified rate of growth of  $\psi_i$ , can guarantee or defend in a repeated game. As we mentioned in the introduction, we study a benchmark case for which we can obtain concrete results in this abstract setting: two-person repeated games where a player with bounded rationality plays against a fully rational player. We point out that our results apply to any measure of strategic complexity that gives rise to a feasible strategy set satisfying our condition on the rate of growth  $\psi_1(t)$ .

We shall follow the following notational rule. Actions of player 1 and 2 in the stage game are denoted by  $a$  and  $b$ , respectively, and their strategies in the repeated game are denoted by  $\sigma$  and  $\tau$ , respectively, with sub or superscripts and other affixes added as necessary. The payoff function of player 1,  $g_1$ , will be denoted simply by  $g$ . Let  $w$  be player 1's maximin payoff in the stage game where max and min are taken over the pure actions:  $w = \max_{a \in A_1} \min_{b \in A_2} g(a, b)$ . This is the worst payoff that player 1 can guarantee himself for sure in the stage game. Also, let  $v$  be the minimax payoff to player 1:  $v = \min_{b \in \Delta(A_2)} \max_{a \in A_1} g(a, b) = \max_{a \in \Delta(A_1)} \min_{b \in A_2} g(a, b)$ . For a pair of repeated game strategies  $(\sigma, \tau) \in \Sigma_1 \times \Sigma_2$ , we write  $g_T(\sigma, \tau)$  for the player 1's average payoff in the  $T$ -th stage.

**3.1. Finite Strategy Set.** If the feasible strategy set  $\Psi_1$  is a finite set, then it is obvious that player 2, with entire strategy set  $\Sigma_2$  available to her, can construct a strategy, say  $\tau_2^*$ , which eventually plays best replies to the stage actions chosen by player 1 and thus gives him at most  $w$  at each stage thereafter.<sup>10</sup> Therefore

---

<sup>10</sup>Ben-Porath (1993)[Lemma 1, Theorem 1] provides an explicit construction of such strategy when  $\Psi_1$  arises from complexity bound in terms of finite automata.

$g_t(\sigma, \tau^*)$  converges to  $w$  for every  $\sigma \in \Psi_1$ . The first theorem provides its speed of convergence. It has appeared in Neyman and Okada (2000b) in a study of nonzero-sum two person finitely repeated games with finite automata. In order to make this paper self-contained, and, since a particular strategy constructed in its proof will be used in the proof of the second theorem, we will provide the proof.

**Theorem 1.** *For every finite subset  $\Psi_1$  of  $\Sigma_1$  there exists  $\tau^* \in \Sigma_2$  such that for all  $\sigma \in \Psi_1$*

$$g_T(\sigma, \tau^*) \leq w + \|g\| \frac{\log_2 |\Psi_1|}{T} \quad \text{for all } T = 1, 2, \dots$$

where  $\|g\| = 2 \max\{|g(a, b)| : a \in A_1, b \in A_2\}$ .

**Proof:** For each history  $h = (h_1, \dots, h_{t-1})$ , where  $h_s = (a_s, b_s) \in A_1 \times A_2$ , let  $\Psi_1^h$  be the set of strategies in  $\Psi_1$  that are compatible with  $h$ . That is,

$$\Psi_1^h = \{\sigma \in \Psi \mid \sigma(\epsilon) = a_1 \text{ and } \sigma(h_1, \dots, h_{s-1}) = a_s \text{ for all } s = 2, \dots, t-1\}.$$

For each  $a \in A_1$  let  $\Psi_1^{h,a}$  be the set of strategies in  $\Psi_1^h$  that takes the action  $a$  given the history  $h$ , i.e.,

$$\Psi_1^{h,a} = \{\sigma \in \Psi_1^h \mid \sigma(h) = a\}.$$

Let  $a(h) \in A_1$  be such that  $|\Psi_1^{h,a(h)}| \geq |\Psi_1^{h,a}|$  for all  $a \in A_1$ . Now define  $\tau^*$  by  $\tau^*(h) = \operatorname{argmin}_{b \in A_2} g(a(h), b)$ .

Clearly,  $\{\Psi_1^{h,a} \mid a \in A_1\}$  is a partition of  $\Psi_1^h$ . If  $a \neq a(h)$ , then  $|\Psi_1^{h,a}|$  is at most one half of  $|\Psi_1^h|$ . This implies that

$$(1) \quad |\Psi_1^{(h_1, \dots, h_{t-1}, h_t)}| \leq \frac{|\Psi_1^{(h_1, \dots, h_{t-1})}|}{2}$$

whenever  $h_t \neq (a(h), \cdot)$ .

Fix  $\sigma \in \Psi_1$  and let  $(h_1, h_2, \dots)$ , where  $h_t = (a_t, b_t)$ , be the play generated by  $(\sigma, \tau^*)$ . Let  $I_t = \mathbf{1}(a_t \neq a(h_1, \dots, h_{t-1}))$  where  $\mathbf{1}(\cdot)$  is the indicator function. Then (1) implies that

$$\sum_{t=1}^T I_t \leq \log_2 |\Psi_1|,$$

that is, the number of stages at which player 1's action differs from  $a(h_1, \dots, h_{t-1})$  is at most  $\log_2 |\Psi_1|$ . Thus

$$g_T(\sigma, \tau^*) = \frac{1}{T} \sum_{t=1}^T g(h_t) \leq \frac{1}{T} \sum_{t=1}^T \left( (1 - I_t)w + I_t \frac{\|g\|}{2} \right) \leq w + \|g\| \frac{\log_2 |\Psi_1|}{T}.$$

This completes the proof. **Q.E.D.**

**3.2. Slowly Growing Strategy Set.** Next we consider an infinite strategy set  $\Psi_1$ . Recall that  $\Psi_1(t)$  is formed by identifying strategies in  $\Psi_1$  that are equivalent up to the  $t$ -th stage and  $\psi_1(t) = |\Psi_1(t)|$ . The next theorem states that if the growth rate of  $\psi_1(t)$  is subexponential in  $t$ , then player 1 cannot guarantee more than  $w$  in the long run. Note that whether player 1 can actually attain  $w$  or not depends on what strategies are in  $\Psi_1$ . For example, if  $a^* = \operatorname{argmax}_{a \in A_1} \min_{b \in A_2} g(a, b)$ , and a strategy that takes  $a^*$  in every stage is available, then  $w$  can be achieved by using such strategy.

**Theorem 2.** *Suppose that  $\frac{\log_2 \psi_1(t)}{t} \xrightarrow{t \rightarrow \infty} 0$ . Then there is a strategy  $\tau^* \in \Sigma_2$  such that*

$$\overline{\lim}_{T \rightarrow \infty} \max_{\sigma \in \Psi_1} g_T(\sigma, \tau^*) \leq w.$$

*Consequently, the minimax value of the repeated game is at most  $w$ . Also, there is a  $\Psi_1 \subset \Sigma_1$  with  $|\Psi_1| = 1$  and relative to which the maximin value of the repeated game is at least  $w$ .*

**Proof:** Let  $\{t_k\}_{k=1}^\infty$  be an increasing<sup>11</sup> sequence of positive integers satisfying the following properties: (A)  $\frac{t_{k+1} - t_k}{t_k} \xrightarrow{k \rightarrow \infty} 0$ , and (B)  $\frac{\log_2 \psi_1(t_{k+1})}{t_{k+1} - t_k} \xrightarrow{k \rightarrow \infty} 0$ . It is easy to verify that such sequence exists under the condition of the theorem. Call a consecutive stages  $t_k + 1, \dots, t_{k+1}$  of the repeated game the  $k$ -th block.

The construction of player 2's strategy  $\tau^*$  is similar to the one in the proof of Theorem 1. Given a history  $h = (h_1, \dots, h_{t-1})$ , there is a unique  $k$  with  $t_k \leq t < t_{k+1}$ . Let  $\Psi_1^h(t_{k+1})$  be the set of player 1's strategies in  $\Psi_1(t_{k+1})$  that are compatible with  $h$  and, for each  $a \in A_1$ , set

$$\Psi_1^{h,a}(t_{k+1}) = \{\sigma \in \Psi_1^h(t_{k+1}) \mid \sigma(h) = a\}.$$

Let  $a(h) = \operatorname{argmax}_{a \in A_1} |\Psi_1^{h,a}(t_{k+1})|$  and define

$$\tau^*(h) = \operatorname{argmin}_{b \in A_2} g(a(h), b).$$

---

<sup>11</sup>In this paper, we use the terms “nondecreasing” and “increasing” (resp. “nonincreasing” and “decreasing”), rather than “increasing” and “strictly increasing” (resp. “decreasing” and “strictly decreasing”).

In short,  $\tau^*$  plays the strategy constructed in the proof of Theorem 1 against  $\Psi_1^h(t_{k+1})$  during the  $k$ -th block.

Fix  $\sigma \in \Psi_1$ . Let  $h = (h_1, h_2, \dots)$  be the play induced by  $(\sigma, \tau^*)$ . Note that  $\sigma \in \Psi_1^{(h_1, \dots, h_{t-1})}(t_{k+1}) \subset \Psi_1(t_{k+1})$  whenever  $t_k + 1 \leq t \leq t_{k+1}$ . Applying Theorem 1 to the  $k$ -th block we have

$$\sum_{t=t_k+1}^{t_{k+1}} g(h_t) \leq (t_{k+1} - t_k)w + \|g\| \log_2 \psi_1(t_{k+1}).$$

Therefore,

$$\sum_{t=1}^{t_k} g(h_t) \leq t_k w + \|g\| \sum_{\ell < k} \log_2 \psi_1(t_{\ell+1}).$$

As  $(t_k)_k$  is strictly increasing, condition (B) implies that

$$\frac{1}{t_k} \sum_{\ell < k} \log_2 \psi_1(t_{\ell+1}) \xrightarrow[k \rightarrow \infty]{} 0.$$

Fix  $\varepsilon > 0$  and let  $k_0$  be sufficiently large so that for every  $k \geq k_0$  we have

$$\|g\| \frac{1}{t_k} \sum_{\ell < k} \log_2 \psi_1(t_{\ell+1}) < \varepsilon/2.$$

Therefore, for  $k \geq k_0$  we have

$$\frac{1}{t_k} \sum_{t=1}^{t_k} g(h_t) < w + \varepsilon/2.$$

Together with condition (A), we have, for sufficiently large  $T$ ,

$$\frac{1}{T} \sum_{t=1}^T g(h_t) < w + \varepsilon.$$

The second part of the theorem follows immediately from the remark preceding the statement of the theorem. **Q.E.D.**

**3.3. Growth of Strategy Sets and Entropy.** In this section we prove a generalization of Theorem 2 for the case when  $\frac{\log_2 \psi_1(t)}{t}$  converges to an arbitrary positive number. To do this we will use the concept of entropy and its properties which we will now recall.<sup>12</sup>

<sup>12</sup>For more details and examples on entropy and related information theoretic tools, see Cover and Thomas (1991)

Let  $X$  be a random variable that takes values in a finite set  $\Omega$  and let  $p(x)$  denote the probability that  $X = x$  for each  $x \in \Omega$ . Then the entropy of  $X$  is defined as

$$H(X) = - \sum_{x \in \Omega} p(x) \log_2 p(x)$$

where  $0 \log_2 0 \equiv 0$ . The entropy as a function of distribution  $p$  is uniformly continuous (in  $L_1$ -norm), concave, and  $H(X) \leq \log_2 |\Omega|$  where the equality is achieved when  $p(x) = \frac{1}{|\Omega|}$  for all  $x \in \Omega$ .

The conditional entropy of a random variable  $X$  given another random variable  $Y$  is defined as follows. Given the event  $Y = y$ , let  $H(X|y)$  be the entropy of  $X$  with respect to the conditional distribution of  $X$  given  $y$ , that is,

$$H(X|y) = - \sum_x p(x|y) \log_2 p(x|y).$$

Then the conditional entropy of  $X$  given  $Y$  is the expected value of  $H(X|y)$  with respect to the (marginal) distribution of  $Y$ :

$$H(X|Y) = E_Y[H(X|y)] = \sum_y p(y) H(X|y).$$

Conditioning reduces entropy, i.e.,  $H(X) \geq H(X|Y) \geq H(X|Y, Z)$  etc., and  $H(X|Y) = H(X)$  if, and only if,  $X$  and  $Y$  are independent. An important consequence of the definition of the conditional entropy is the ‘‘chain rule’’:

$$H(X_1, \dots, X_T) = H(X_1) + \sum_{t=2}^T H(X_t | X_1, \dots, X_{t-1}).$$

Let  $(\Omega, \mathcal{F}, \mu)$  be a probability space and let  $\mathcal{P}$  be a finite partition of  $\Omega$  into sets in  $\mathcal{F}$ . Then the entropy of the partition  $\mathcal{P}$ , with respect to  $\mu$  is defined by

$$H_\mu(\mathcal{P}) = - \sum_{F \in \mathcal{P}} \mu(F) \log_2 \mu(F).$$

It is easy to see that if  $\mathcal{Q}$  is a refinement of  $\mathcal{P}$ , then  $H_\mu(\mathcal{P}) \leq H_\mu(\mathcal{Q})$ .

Given a feasible strategy set of player 1,  $\Psi_1 \subset \Sigma_1$ , we have defined, for each  $t$ , the set  $\Psi_1(t)$  to be the partition of  $\Psi_1$  induced by an equivalence relation. Specifically, we define an equivalence relation  $\sim_t$  by

$$\sigma \sim_t \sigma' \iff \forall \tau \in \Sigma_2, a_s(\sigma, \tau) = a_s(\sigma', \tau) \quad \text{for } s = 1, \dots, t.$$

Then  $\Psi_1(t) = \Psi_1 / \sim_t$ .



Now fix player 2's strategy  $\tau$ . Define an equivalence relation  $\sim_{t,\tau}$  by

$$\sigma \sim_{t,\tau} \sigma' \iff a_s(\sigma, \tau) = a_s(\sigma', \tau) \quad \text{for } s = 1, \dots, t.$$

and let  $\Psi_1(t, \tau) = \Psi_1 / \sim_{t,\tau}$ . Clearly  $\Psi_1(t, \tau)$  is a finite partition of  $\Psi_1$ , and  $\Psi_1(t)$  is a refinement of  $\Psi_1(t, \tau)$ . Hence, by the property of the entropy of partitions mentioned above,

$$(2) \quad H_\sigma(\Psi_1(t, \tau)) \leq H_\sigma(\Psi_1(t)) \leq \log_2 |\Psi_1(t)| = \log_2 \psi_1(t).$$

By the definition of the equivalence relation defining  $\Psi_1(t, \tau)$ , each equivalence class  $S \in \Psi_1(t, \tau)$  is associated with a history of length  $t$ , say  $h(S) \in H_t$ . More precisely,  $h(S)$  is the history of length  $t$  which results when the strategy profile  $(s, \tau)$  is played, for any  $s \in S$ . Conversely, for any history  $h \in H_t$ , there is an equivalence class  $S \in \Psi_1(t, \tau)$  such that  $h = h(S)$ . So there is a one-to-one map from  $\Psi_1(t, \tau)$  into  $H_t$ . Furthermore, the event “a strategy  $s \in S \subset \Psi_1(t, \tau)$  is selected by  $\sigma$ ” is equivalent to the event “the history  $h(S)$  occurs when  $(\sigma, \tau)$  is played”. Therefore,

$$\sigma(S) = P_{\sigma,\tau}(h(S)).$$

Let us write  $X_1, \dots, X_t$  for the sequence of action profiles up to stage  $t$  when  $(\sigma, \tau)$  is played. So it is a random vector with distribution  $P_{\sigma,\tau}$ . Then the observation in this paragraph implies that

$$\begin{aligned} H_\sigma(\Psi_1(t, \tau)) &= - \sum_{S \in \Psi_1(t, \tau)} \sigma(S) \log_2 \sigma(S) \\ &= - \sum_{h \in H_t} P_{\sigma,\tau}(h) \log_2 P_{\sigma,\tau}(h) \\ &= H(X_1, \dots, X_t). \end{aligned}$$

Combining this equality with (2) we have

**Lemma 1.** *Let  $\sigma \in \Delta(\Psi_1)$  and  $\tau \in \Sigma_2$  and  $(X_1, \dots, X_t)$  be the random play up to stage  $t$  induced by  $(\sigma, \tau)$ . Then, for every  $t$ ,*

$$H(X_1, \dots, X_t) \leq \log_2 \psi_1(t).$$

Next, for each mixed action  $\alpha$  of player 1, let  $H(\alpha)$  be its entropy, i.e.,

$$H(\alpha) = - \sum_{a \in A_1} \alpha(a) \log_2 \alpha(a).$$

Define a function  $U : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  by

$$U(\gamma) = \max_{\substack{\alpha \in \Delta(A_1) \\ H(\alpha) \leq \gamma}} \min_{b \in A_2} g(\alpha, b).$$

Thus  $U(\gamma)$  is what player 1 can secure in the stage game  $G$  using a mixed action of entropy at most  $\gamma$ . Clearly,  $U(0) = w$ , the maximin payoff in pure actions. On the other hand,  $U(\gamma) = v$ , the minimax payoff, if  $\gamma \geq \bar{\gamma}$  where  $\bar{\gamma} = \min \left\{ H(\alpha) : \alpha \in \Delta(A_1), \min_{b \in A_2} g(\alpha, b) = v \right\}$ . Let  $\text{cav } U$  be the concavification of  $U$ , i.e., the smallest concave function which is at least as large as  $U$  at every point in its domain.

The function  $U(\gamma)$  is strictly increasing and piecewise convex for  $0 \leq \gamma \leq \bar{\gamma}$ , and then constant,  $v$ , for  $\gamma \geq \bar{\gamma}$ . Thus, for every  $\gamma \leq \bar{\gamma}$ , there is an  $\alpha \in \Delta(A_1)$  such that  $H(\alpha) = \gamma$  and  $\min_{b \in A_2} g(\alpha, b) = U(\gamma)$ . In other words, the entropy constraint defining  $U(\gamma)$  is binding for  $\gamma \leq \bar{\gamma}$ . See Neyman and Okada (2000a) for examples.

The theorem below asserts that, if  $\psi_1(t)$  grows like an exponential function  $2^{\gamma t}$ , then player 1's maximin payoff in the repeated game is at most  $(\text{cav } U)(\gamma)$ .

**Theorem 3.** *Suppose that  $\overline{\lim}_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} \leq \gamma$ . Then, for every  $\sigma \in \Delta(\Psi_1)$ , there is  $\tau \in \Sigma_2$  such that*

$$\overline{\lim}_{T \rightarrow \infty} g_T(\sigma, \tau) \leq (\text{cav } U)(\gamma).$$

**Proof:** Fix player 1's strategy  $\sigma \in \Delta(\Psi_1)$ . Define player 2's strategy as follows. At each stage  $t$ , and at each history  $h \in H_{t-1}$ ,  $\tau(h)$  minimizes player 1's stage payoff, that is,

$$E_{\sigma, \tau}[g(a_t)|h] = \min_{b \in B} E_{\sigma(h)}[g(a, b)].$$

Let  $X_1, X_2, \dots$  be the sequence of random actions induced by  $(\sigma, \tau)$ . Let  $H(X_t|h)$  be the entropy of  $X_t$  given that a history  $h$  is realized. Note that, conditional on the history  $h$ , the entropy of player 1's mixed action at stage  $t$  is  $H(X_t|h)$ . Hence, by the definitions of  $U$ ,  $\text{cav } U$ , and  $\tau$ , we have

$$E_{\sigma, \tau}[g(X_t)|h] \leq U(H(X_t|h)) \leq (\text{cav } U)(H(X_t|h)).$$

Taking the expectation, we have

$$E_{\sigma, \tau}[g(X_t)] \leq E_{\sigma, \tau}[(\text{cav } U)(H(X_t|h))] \leq (\text{cav } U)(E_{\sigma, \tau}[H(X_t|h)])$$

where the second inequality follows from the concavity of  $\text{cav } U$  and Jensen's inequality. Summing over  $t = 1, \dots, T$  we have

$$\begin{aligned}
g_T(\sigma, \tau) &= \frac{1}{T} \sum_{t=1}^T E_{\sigma, \tau}[g(X_t)] \\
&\leq \frac{1}{T} \sum_{t=1}^T (\text{cav } U)(E_{\sigma, \tau}[H(X_t|h)]) \\
&\leq (\text{cav } U) \left( \frac{1}{T} \sum_{t=1}^T E_{\sigma, \tau}[H(X_t|h)] \right) \\
&= (\text{cav } U) \left( \frac{1}{T} \sum_{t=1}^T H(X_t|X_1, \dots, X_{t-1}) \right) \\
&= (\text{cav } U) \left( \frac{1}{T} H(X_1, \dots, X_T) \right) \\
&\leq (\text{cav } U) \left( \frac{\log_2 \psi_1(T)}{T} \right).
\end{aligned}$$

The second inequality follows from the concavity of  $\text{cav } U$  and Jensen's inequality. The second and the third equalities follow from the definition of conditional entropy and the chain rule respectively. The last inequality follows from Lemma 1. Since  $\varliminf_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} \leq \gamma$ , we have the desired result. **Q.E.D.**

As in Theorem 2, whether player 1 can achieve  $(\text{cav } U)(\gamma)$  or not depends on what strategies are available to him. The next theorem states that there is indeed a strategy set with an appropriate growth rate with which  $(\text{cav } U)(\gamma)$  can be achieved. Furthermore, it states that it suffices to consider oblivious strategies. Combined with Theorem 3, it implies that there is a strategy set  $\Psi_1$  of player 1 consisting only of oblivious strategies for which  $\psi_1(t)$  grows like  $2^{\gamma t}$  and, relative to which, the maximin value of the repeated game is precisely  $(\text{cav } U)(\gamma)$ .

**Theorem 4.** *For every  $\gamma \geq 0$  and a function  $f : \mathbb{R}_+ \rightarrow [1, \infty)$  with  $\frac{\log_2 f(t)}{t} \xrightarrow[t \rightarrow \infty]{} \gamma$ , there exists a set of oblivious strategies  $\Psi_1 \subset \Sigma_1$  and a mixed strategy  $\hat{\sigma} \in \Delta(\Psi_1)$  with the following properties.*

- (i)  $\psi_1(t) \leq f(t)$  for every  $t \in \mathbb{N}$
- (ii)  $\varliminf_{T \rightarrow \infty} \left( \inf_{\tau \in \Delta(\Sigma_2)} g_T(\hat{\sigma}, \tau) \right) \geq (\text{cav } U)(\gamma)$

$$(iii) \quad \inf_{\tau \in \Delta(\Sigma_2)} \mathbf{E}_{\hat{\sigma}, \tau} \left[ \liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g(a_t, b_t) \right] \geq (\text{cav } U)(\gamma)$$

**Proof:** CONSTRUCTION OF  $\Psi_1$ : Recall from Example 2 that, for each sequence  $\mathbf{a} = (a_1, a_2, \dots)$  of player 1's pure actions,  $\sigma\langle \mathbf{a} \rangle$  denotes his oblivious strategy that takes action  $a_t$  at stage  $t$  regardless of the past history. We will define a particular class of sequences  $F \subset A_1^\infty$  and then set

$$(3) \quad \Psi_1 = \{\sigma\langle \mathbf{a} \rangle : \mathbf{a} \in F\}.$$

If  $\gamma = 0$ , then  $(\text{cav } U)(\gamma)$  is the maximin payoff in pure actions,  $w$ . In this case the set  $F$  can be taken as a singleton  $\{\mathbf{a} = (a, a, a, \dots)\}$  where  $a$  is any one of player 1's pure actions that guarantees him  $w$ .

Suppose that  $\gamma > 0$ . Recall that  $\bar{\gamma} = \min\{H(\alpha) : \alpha \in \Delta(A_1), \min_{b \in A_2} g(\alpha, b) = v\}$ . As  $U(\gamma) = v$  for all  $\gamma \geq \bar{\gamma}$ , we assume w.l.o.g. that  $\gamma \leq \bar{\gamma}$ . By modifying  $f(t)$  to  $\hat{f}(t) = \inf_{s \geq t} f(s)^{t/s}$  if necessary, we will also assume that  $\frac{\log_2 f(t)}{t}$  is nondecreasing in  $t$ , or, equivalently,

$$(4) \quad \log_2 f(s) - \log_2 f(t) \geq (s-t) \frac{\log_2 f(t)}{t} \text{ whenever } s > t.$$

In particular, this implies that  $f(t)$  is also nondecreasing in  $t$ .

In order to construct the set  $F \subset A_1^\infty$ , we first partition the stages into blocks. Set  $t_0 = 0$ . The  $n$ -th block consists of stages  $t_{n-1} + 1$  to  $t_n$ . We denote the length of the  $n$ -th block by  $d_n$ , i.e.,  $d_n = t_n - t_{n-1}$ . Second, we will define for each  $n$  a set  $F_n$  consisting of finite sequences of player 1's actions of length  $d_n$  with certain properties. Then we set  $F$  to be those sequences  $\mathbf{a} = (a_1, a_2, \dots)$  in  $A_1^\infty$  whose  $n$ -th segment  $\mathbf{a}[n] = (a_{t_{n-1}+1}, \dots, a_{t_n})$  belongs to  $F_n$ :

$$(5) \quad F = \{\mathbf{a} = (a_1, a_2, \dots) \in A_1^\infty : \mathbf{a}[n] = (a_{t_{n-1}+1}, \dots, a_{t_n}) \in F_n\}.$$

Now we describe the detail.

The blocks are chosen so that  $d_n$  is increasing,  $d_n \xrightarrow[n \rightarrow \infty]{} \infty$ ,  $\frac{d_{n-1}}{d_n} \xrightarrow[n \rightarrow \infty]{} 1$  and thus  $\frac{d_n}{t_n} \xrightarrow[n \rightarrow \infty]{} 0$ . For example, take  $d_n = n$ . Next, we construct the sets  $(F_n)_n$  by means of two sequences of nonnegative reals,  $(\gamma_n)_n$  with  $\gamma_n \leq \bar{\gamma}$  and  $(\eta_n)_n$  where  $\eta_n \geq \frac{|A_1|}{d_n}$  and  $\eta_n \xrightarrow[n \rightarrow \infty]{} 0$ . The sequence  $(\gamma_n)_n$  depends on the function  $f$  and will be specified in the last part of the proof. For each  $n$ , choose player 1's mixed

action  $\alpha_n$  so that  $H(\alpha_n) = \gamma_n$  and  $\min_{b \in A_2} g(\alpha_n, b) = U(\gamma_n)$ . (See the remark on the property of  $U$  on p.18.)

If, by our choice later,  $\gamma_n = 0$ , then  $\alpha_n$  is a pure action, say  $a^*$ , that guarantees  $w$ . In this case we set  $F_n$  to be a singleton consisting of  $(a^*, \dots, a^*) \in A^{d_n}$ . If we choose  $\gamma_n > 0$ , then  $F_n$  is the set of all sequences  $(a_1, \dots, a_{d_n}) \in A_1^{d_n}$  whose empirical distribution is within  $\eta_n$  of  $\alpha_n$ . Formally,

$$(6) \quad F_n = \left\{ (a_1, \dots, a_{d_n}) \in A_1^{d_n} : \sum_{a \in A_1} \left| \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(a_k = a) - \alpha_n(a) \right| \leq \eta_n \right\}.$$

The condition  $\eta_n \geq \frac{|A_1|}{d_n}$  ensures that  $F_n \neq \emptyset$  in this case.

The sequence  $(\eta_n)_n$  is chosen to satisfy, in addition, the following property. Let  $\mathbf{x} = (x_1, x_2, \dots)$  be a sequence of independent  $A_1$ -valued random variables where  $x_t$  is distributed according to  $\alpha_n$  whenever  $t$  is in the  $n$ -th block, i.e.  $t_{n-1} + 1 \leq t \leq t_n$ . Then we require<sup>13</sup>

$$(7) \quad \sum_{n=1}^{\infty} \mathbb{P}(\mathbf{x}[n] = (x_{t_{n-1}+1}, \dots, x_{t_n}) \notin F_n) < \infty.$$

We complete this part by defining  $F$  by (15) and then  $\Psi_1$  by (13).

CONSTRUCTION OF  $\hat{\sigma} \in \Delta(\Psi_1)$ : Generate a sequence of independent actions  $\mathbf{a} = (a_1, a_2, \dots)$  of player 1 where actions in the  $n$ -th block  $\mathbf{a}[n] = (a_{t_{n-1}+1}, \dots, a_{t_n})$  are drawn according to the mixed action  $\alpha_n$ . In the  $n$ -th block, player 1 plays according to  $\mathbf{a}[n]$  whenever  $\mathbf{a}[n] \in F_n$ , otherwise he plays  $\bar{\mathbf{a}}[n] = (\bar{a}_{t_{n-1}+1}, \dots, \bar{a}_{t_n})$  where  $\bar{\mathbf{a}} = (\bar{a}_1, \bar{a}_2, \dots) \in F$  is some fixed sequence.

<sup>13</sup>For  $n$  with  $\gamma_n = 0$ , it is obvious that  $\mathbb{P} \mathbf{x}[n] \notin F_n = 0$ . For  $n$  with  $\gamma_n > 0$ , note that  $\mathbf{x}[n] \notin F_n$  implies that  $\frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(x_{t_{n-1}+k} = a) - \alpha_n(a) > \frac{\eta_n}{|A_1|}$  for some  $a \in A_1$ . For each  $a \in A_1$  and  $k = 1, \dots, d_n$ , the random variable  $\mathbf{1}(x_{t_{n-1}+k} = a)$  takes values 0 and 1, and has mean  $\alpha_n(a)$ . Hence by a large deviation inequality due to Hoeffding (1963) we have

$$\mathbb{P} \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(x_{t_{n-1}+k} = a) - \alpha_n(a) > \frac{\eta_n}{|A_1|} \leq 2 \exp -2d_n \frac{\eta_n^2}{|A_1|^2},$$

and so

$$\mathbb{P} \mathbf{x}[n] \notin F_n \leq 2|A_1| \exp -2d_n \frac{\eta_n^2}{|A_1|^2}.$$

Take, for example,  $\eta_n = |A_1|/d_n^{1/4}$  ( $> |A_1|/d_n$ ). Then the exponential term on the right side of the above inequality is  $\exp -2\sqrt{d_n}$  and (7) holds.

To formally define the mixed strategy  $\hat{\sigma}$ , recall that  $\mathbf{x} = (x_1, x_2, \dots)$  is a sequence of independent  $A_1$ -valued random variables where  $x_t$  is distributed according to  $\alpha_n$  whenever  $t_{n-1} + 1 \leq t \leq t_n$ . Define a sequence of  $A_1$ -valued random variables  $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \dots)$  by

$$\hat{\mathbf{x}}[n] = (\hat{x}_{t_{n-1}+1}, \dots, \hat{x}_{t_n}) = \begin{cases} \mathbf{x}[n] & \text{if } \mathbf{x}[n] \in F_n, \\ \bar{\mathbf{a}}[n] & \text{otherwise.} \end{cases}$$

Then  $\hat{\sigma} = \sigma(\hat{\mathbf{x}})$ . Note that  $\hat{\sigma}$  is indeed a mixture of strategies in  $\Psi_1$ .

VERIFICATION OF THE THEOREM: From (15) it is clear that we can identify each sequence in the set  $F$  with an element of  $\bigtimes_{n=1}^{\infty} F_n$  and each strategy in  $\Psi_1(t_N)$  with an element in  $\bigtimes_{n=1}^N F_n$ ,  $N = 1, 2, \dots$ . Hence  $\psi_1(t_N) = |\Psi_1(t_N)| = \prod_{n=1}^N |F_n|$  for each  $N = 1, 2, \dots$ . Since both  $\psi_1(t)$  and  $f(t)$  are nondecreasing, in order to verify that  $\Psi_1$  has the property (i) of the theorem, it is enough to ensure that<sup>14</sup>

$$(8) \quad \sum_{n=1}^N \log_2 |F_n| \leq \log_2 f(t_{N-1}) \quad \text{for each } N > 1.$$

Recall that  $|F_n| = 1$  for  $n$  with  $\gamma_n = 0$ . For  $n$  with  $\gamma_n > 0$  we estimate  $|F_n|$  as follows. Each element  $\mathbf{a}_n = (a_1, \dots, a_{d_n})$  in  $F_n$  has an empirical distribution  $\rho(\mathbf{a}_n)$  whose  $L_1$ -distance from  $\alpha_n$  is at most  $\eta_n$ . Since the entropy  $H(\alpha)$  as a function on  $\Delta(A_1)$  is uniformly continuous (in the  $L_1$ -distance), there is a nondecreasing function  $\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\varepsilon(\eta) \xrightarrow{\eta \rightarrow 0} 0$  such that  $|H(\rho(\mathbf{a}_n)) - H(\alpha_n)| = |H(\rho(\mathbf{a}_n)) - \gamma_n| < \varepsilon(\eta_n)$  for all  $\mathbf{a}_n \in F_n$ . Since the number of distinct empirical distributions arising from sequences in  $A_1^{d_n}$  is at most  $d_n^{|A_1|}$ , and the number of sequences in  $A_1^{d_n}$  with an empirical distribution  $\alpha$  is at most  $2^{d_n H(\alpha)}$ , we deduce that  $|F_n| \leq d_n^{|A_1|} 2^{d_n(\gamma_n + \varepsilon(\eta_n))}$ . Since  $d_n \uparrow_{n \rightarrow \infty} \infty$  and  $\eta_n \xrightarrow{n \rightarrow \infty} 0$ , we can choose a nonincreasing sequence of nonnegative reals  $(\varepsilon_n)_n$  with  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$  such that<sup>15</sup>

$$(9) \quad |F_n| \leq 2^{d_n(\gamma_n + \varepsilon_n)}.$$

<sup>14</sup>Observe that (i) asserts that  $\psi_1(t) \leq f(t)$  for all  $t$  which will be satisfied, as these functions are nondecreasing, if  $\psi_1(t_N) \leq f(t_{N-1})$  for each  $N$ .

<sup>15</sup>For example,  $\varepsilon_n = \sup_{\ell \geq n} \varepsilon(\eta_\ell) + |A_1| \frac{\log_2 d_n}{d_n}$ .

Thus, to ensure that (17) holds, it is enough to choose  $(\gamma_n)_n$  so that

$$(10) \quad \sum_{n=1}^N \mathbf{1}(\gamma_n > 0) d_n (\gamma_n + \varepsilon_n) \leq \log_2 f(t_{N-1}) \quad \text{for each } N > 1.$$

Next we derive a sufficient condition to be verified in order to show that  $\hat{\sigma}$  has the property (ii). It is easy to see that the  $L_1$ -distance between the conditional distributions of  $\mathbf{x}[n]$  and  $\hat{\mathbf{x}}[n]$  given  $\mathbf{x}[1], \dots, \mathbf{x}[n-1]$  is at most  $2\mathbb{P}(\mathbf{x}[n] \notin F_n)$ , that is,

$$\sum_{\mathbf{a}_n \in A_1^{d_n}} \left| \mathbb{P}(\mathbf{x}[n] = \mathbf{a}_n) - \mathbb{P}(\hat{\mathbf{x}}[n] = \mathbf{a}_n) \right| \leq 2\mathbb{P}(\mathbf{x}[n] \notin F_n).$$

It follows that, in the  $n$ -th block, we have

$$\min_{\tau \in \Sigma_2} \mathbf{E}_{\hat{\sigma}, \tau} \left[ \sum_{t=t_{n-1}+1}^{t_n} g(a_t, b_t) \right] \geq d_n U(\gamma_n) - 2\|g\| \mathbb{P}(\mathbf{x}[n] \notin F_n)$$

and hence, for each  $N = 1, 2, \dots$ ,

$$\min_{\tau \in \Sigma_2} \mathbf{E}_{\hat{\sigma}, \tau} \left[ \sum_{t=1}^{t_N} g(a_t, b_t) \right] \geq \sum_{n=1}^N d_n U(\gamma_n) - 2\|g\| \sum_{n=1}^N \mathbb{P}(\mathbf{x}[n] \notin F_n).$$

Thus, by virtue of (7), in order to show part (ii) of the theorem it suffices to choose  $(\gamma_n)_n$  so that

$$(11) \quad \frac{1}{t_N} \sum_{n=1}^N d_n U(\gamma_n) \xrightarrow{N \rightarrow \infty} (\text{cav } U)(\gamma).$$

We now exhibit a choice of  $(\gamma_n)_n$  that satisfy (17) and (21). We distinguish two cases.

CASE 1 -  $(\text{cav } U)(\gamma) = U(\gamma)$ : Since  $\frac{d_{n-1}}{d_n} \xrightarrow{n \rightarrow \infty} 1$ , we have  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} \xrightarrow{n \rightarrow \infty} \gamma$ . Let  $\bar{n}$  be sufficiently large so that  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} > \varepsilon_n$  for all  $n > \bar{n}$ . Then set  $\gamma_n = 0$  for  $n \leq \bar{n}$  and  $\gamma_n = \frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} - \varepsilon_n$  for  $n > \bar{n}$ . With this choice of  $(\gamma_n)_n$  it is easy to verify that (18), and hence (17), is satisfied. As  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ , we have  $\gamma_n \xrightarrow{n \rightarrow \infty} \gamma$ . Thus the condition (21) is satisfied as well.

CASE 2 -  $(\text{cav } U)(\gamma) > U(\gamma)$ : In this case, the definitions of  $U$  and  $\text{cav } U$  imply the existence of  $\gamma_-, \gamma_+$  with  $0 \leq \gamma_- < \gamma < \gamma_+$  and  $\alpha_-, \alpha_+ \in \Delta(A_1)$  together with a  $p \in (0, 1)$  such that

- a)  $\gamma = p\gamma_- + (1-p)\gamma_+$
- b)  $(\text{cav } U)(\gamma) = pU(\gamma_-) + (1-p)U(\gamma_+)$

- c)  $H(\alpha_-) \leq \gamma_-$  and  $H(\alpha_+) \leq \gamma_+$   
d)  $g(\alpha_-, b) \geq U(\gamma_-)$  and  $g(\alpha_+, b) \geq U(\gamma_+)$  for all  $b \in A_2$ .

Choose  $\bar{n}$  large enough so that for  $n \geq \bar{n}$  we have  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-2})}{t_{n-2}} - \varepsilon_n > \gamma_-$ .  
Set  $\gamma_n = 0$  for  $n \leq \bar{n}$  and, for  $n > \bar{n}$ , define  $\gamma_n$  by induction as follows:

$$\gamma_n = \begin{cases} \gamma_+ & \text{if } \sum_{\ell=1}^{n-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) + d_n (\gamma_+ + \varepsilon_n) \leq \log_2 f(t_{n-1}), \\ \gamma_- & \text{otherwise.} \end{cases}$$

With the above choices of the sequence  $(\gamma_n)_n$ , we have  $|F_n| = 1$  for  $n \leq \bar{n}$ . So for  $N \leq \bar{n}$ , the inequality (17) trivially holds. For  $N > \bar{n}$ , inequality (18) is proved by induction. Assume that  $\sum_{\ell=1}^{N-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) \leq \log_2 f(t_{N-2})$ . Then,

$$\begin{aligned} \sum_{\ell=1}^{N-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) + d_N (\gamma_- + \varepsilon_N) &\leq \log_2 f(t_{N-2}) + d_{N-1} \frac{\log_2 f(t_{N-2})}{t_{N-2}} \\ &\leq \log_2 f(t_{N-1}) \end{aligned}$$

where the second inequality follows from (14). Therefore, if  $\gamma_N = \gamma_-$  then inequality (18) holds for  $N$ . Obviously, by the definition of  $\gamma_N$ , if  $\gamma_N = \gamma_+$  then inequality (18) again holds for  $N$ . We conclude that inequality (18), hence (17), holds for all  $N > \bar{n}$ .

Since  $\gamma_n = \gamma_-$  or  $\gamma_+$  for all  $n > \bar{n}$ , in order to show (21), it suffices to verify that

$$(12) \quad \frac{1}{t_N} \sum_{n=\bar{n}+1}^N d_n \mathbf{1}(\gamma_n = \gamma_-) \xrightarrow{N \rightarrow \infty} p.$$

For each  $N$  let  $M_N = \max\{n \leq N : \gamma_n = \gamma_-\}$ . Note that  $M_N \rightarrow \infty$  as  $N \rightarrow \infty$  since  $\gamma_+ > \gamma$ . Since  $\frac{\log_2 f(t_{n-1})}{t_n} = \left(1 - \frac{d_n}{t_n}\right) \frac{\log_2 f(t_{n-1})}{t_{n-1}} \xrightarrow{n \rightarrow \infty} \gamma$ , we have, for



every  $\delta > 0$ , an  $N$  such that  $\frac{\log_2 f(t_{n-1})}{t_n} \geq \gamma - \delta$  for all  $n \geq M_N$ . Hence

$$\begin{aligned} & \sum_{n=1}^N \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \varepsilon_n) \\ &= \sum_{n=1}^{M_N} \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \varepsilon_n) + \sum_{n=M_N+1}^N \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \varepsilon_n) \\ &\geq \log_2 f(t_{M_N-1}) - d_{M_N}(\gamma_+ - \gamma_-) + (t_N - t_{M_N})\gamma_+ \\ &\geq t_{M_N}(\gamma - \delta) - d_N(\gamma_+ - \gamma_-) + (t_N - t_{M_N})\gamma \\ &\geq t_N(\gamma - \delta) - d_N(\gamma_+ - \gamma_-). \end{aligned}$$

Since  $\sum_{n=1}^N \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \varepsilon_n) \leq \log f(t_{N-1})$ ,  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ , and  $\frac{d_N}{t_N} \xrightarrow{N \rightarrow \infty} 0$ , we conclude that  $\frac{1}{t_N} \sum_{n=1}^N d_n \gamma_n \xrightarrow{N \rightarrow \infty} \gamma$  which implies (22). **Q.E.D.**

**Proof:** An alternative proof using the lemma

CONSTRUCTION OF  $\Psi_1$ : Recall that, for each sequence  $\mathbf{a} = (a_1, a_2, \dots)$  of player 1's pure actions,  $\sigma(\mathbf{a})$  denotes his oblivious strategy that takes action  $a_t$  at stage  $t$  regardless of the past history. We will define a particular class of sequences  $F \subset A_1^\infty$  and then set

$$(13) \quad \Psi_1 = \{\sigma(\mathbf{a}) : \mathbf{a} \in F\}.$$

If  $\gamma = 0$ , then  $(\text{cav } U)(\gamma)$  is the maximin payoff in pure actions,  $w$ . In this case the set  $F$  can be taken as a singleton  $\{\mathbf{a} = (a, a, a, \dots)\}$  where  $a$  is any one of player 1's pure actions that guarantees him  $w$ .

Suppose that  $\gamma > 0$ . Recall that  $\bar{\gamma} = \min\left\{H(\alpha) : \alpha \in \Delta(A_1), \min_{b \in A_2} g(\alpha, b) = v\right\}$ . As  $U(\gamma) = v$  for all  $\gamma \geq \bar{\gamma}$ , we assume w.l.o.g. that  $\gamma \leq \bar{\gamma}$ . By modifying  $f(t)$  to  $\hat{f}(t) = \inf_{s \geq t} f(s)^{t/s}$  if necessary, we will also assume that  $\frac{\log_2 f(t)}{t}$  is nondecreasing in  $t$ , or, equivalently,

$$(14) \quad \log_2 f(s) - \log_2 f(t) \geq (s-t) \frac{\log_2 f(t)}{t} \text{ whenever } s > t.$$

In particular, this implies that  $f(t)$  is also nondecreasing in  $t$ .

In order to construct the set  $F \subset A_1^\infty$ , we first partition the stages into blocks. Set  $t_0 = 0$ . The  $n$ -th block consists of stages  $t_{n-1} + 1$  to  $t_n$ . We denote the length

of the  $n$ -th block by  $d_n$ , i.e.,  $d_n = t_n - t_{n-1}$ . Second, we will define for each  $n$  a set  $F_n$  consisting of finite sequences of player 1's actions of length  $d_n$  with certain properties. Then we set  $F$  to be those sequences  $\mathbf{a} = (a_1, a_2, \dots)$  in  $A_1^\infty$  whose  $n$ -th segment  $\mathbf{a}[n] = (a_{t_{n-1}+1}, \dots, a_{t_n})$  belongs to  $F_n$ :

$$(15) \quad F = \{ \mathbf{a} = (a_1, a_2, \dots) \in A_1^\infty : \mathbf{a}[n] = (a_{t_{n-1}+1}, \dots, a_{t_n}) \in F_n \}.$$

Now we describe the detail.

The blocks are chosen so that  $d_n$  is increasing,  $d_n \xrightarrow{n \rightarrow \infty} \infty$ ,  $\frac{d_{n-1}}{d_n} \xrightarrow{n \rightarrow \infty} 1$  and thus  $\frac{d_n}{t_n} \xrightarrow{n \rightarrow \infty} 0$ . For example, take  $d_n = n$ . Next, we construct the sets  $(F_n)_n$  by means of a sequence of nonnegative reals,  $(\gamma_n)_n$  with  $\gamma_n \leq \bar{\gamma}$ . The sequence  $(\gamma_n)_n$  depends on the function  $f$  and will be specified in the last part of the proof. For each  $n$ , choose player 1's mixed action  $\alpha_n$  so that  $H(\alpha_n) = \gamma_n$  and  $\min_{b \in A_2} g(\alpha_n, b) = U(\gamma_n)$ . (See the remark on the property of  $U$  on p.18.)

If, by our choice later,  $\gamma_n = 0$ , then  $\alpha_n$  is a pure action, say  $a^*$ , that guarantees  $w$ . In this case we set  $F_n$  to be a singleton consisting of  $(a^*, \dots, a^*) \in A^{d_n}$ . If we choose  $\gamma_n > 0$ , then  $F_n$  is defined to be the set of all sequences  $(a_1, \dots, a_{d_n}) \in A_1^{d_n}$  whose empirical distribution is within  $\frac{|A_1|}{d_n}$  of  $\alpha_n$ . Formally,

$$(16) \quad F_n = \left\{ (a_1, \dots, a_{d_n}) \in A_1^{d_n} : \sum_{a \in A_1} \left| \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(a_k = a) - \alpha_n(a) \right| \leq \frac{|A_1|}{d_n} \right\}.$$

Note that  $F_n \neq \emptyset$ . We complete this part by defining  $F$  by (15) and then  $\Psi_1$  by (13).

CONSTRUCTION OF  $\hat{\sigma} \in \Delta(\Psi_1)$ : Generate a sequence of independent strings of actions  $\mathbf{a} = (\mathbf{a}[1], \mathbf{a}[2], \dots)$  where the  $n$ -th string  $\mathbf{a}[n]$  is selected uniformly from  $F_n$ . Then play according to the oblivious strategy  $\sigma(\mathbf{a})$ . More formally, let  $\mathbf{z} = (z_1, z_2, \dots)$  be a sequence of  $A_1$ -valued random variables such that its  $n$ -th segment  $\mathbf{z}[n] = (z_{t_{n-1}+1}, \dots, z_{t_n})$  is drawn uniformly from  $F_n$ , and independently from  $\mathbf{z}[m]$  for  $m \neq n$ . Then define  $\hat{\sigma} = \sigma(\mathbf{z})$ . Observe that  $\hat{\sigma}$  is indeed a mixture of strategies in  $\Psi_1$ .

VERIFICATION OF THE THEOREM: From (15) it is clear that we can identify each sequence in the set  $F$  with an element of  $\prod_{n=1}^{\infty} F_n$  and each strategy in  $\Psi_1(t_N)$  with

an element in  $\prod_{n=1}^N F_n$ ,  $N = 1, 2, \dots$ . Hence  $\psi_1(t_N) = |\Psi_1(t_N)| = \prod_{n=1}^N |F_n|$  for each  $N = 1, 2, \dots$ . Since both  $\psi_1(t)$  and  $f(t)$  are nondecreasing, in order to verify that  $\Psi_1$  has the property (i) of the theorem, it is enough to ensure that<sup>16</sup>

$$(17) \quad \sum_{n=1}^N \log_2 |F_n| \leq \log_2 f(t_{N-1}) \quad \text{for each } N > 1.$$

Recall that  $|F_n| = 1$  for  $n$  with  $\gamma_n = 0$ . For  $n$  with  $\gamma_n > 0$  we estimate  $|F_n|$  as follows. Each element  $\mathbf{a}_n = (a_1, \dots, a_{d_n})$  in  $F_n$  has an empirical distribution  $\rho(\mathbf{a}_n)$  whose  $L_1$ -distance from  $\alpha_n$  is at most  $\frac{|A_1|}{d_n}$ . Since the entropy  $H(\alpha)$  as a function on  $\Delta(A_1)$  is uniformly continuous (in the  $L_1$ -distance), there is a nonincreasing sequence of positive reals  $(\varepsilon_n)_n$  with  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$  such that  $|H(\rho(\mathbf{a}_n)) - H(\alpha_n)| = |H(\rho(\mathbf{a}_n)) - \gamma_n| < \varepsilon_n$  for all  $\mathbf{a}_n \in F_n$ . Since the number of distinct empirical distributions arising from sequences in  $A_1^{d_n}$  is at most  $d_n^{|A_1|}$ , and the number of sequences in  $A_1^{d_n}$  with an empirical distribution  $\alpha$  is at most  $2^{d_n H(\alpha)}$ , we deduce that

$$2^{d_n(\gamma_n - \varepsilon_n)} \leq |F_n| \leq d_n^{|A_1|} 2^{d_n(\gamma_n + \varepsilon_n)} = 2^{d_n(\gamma_n + \delta_n)}$$

where  $\delta_n = \varepsilon_n + |A_1| \frac{\log_2 d_n}{d_n}$ . Note that the sequence  $(\delta_n)_{n \geq 2}$  is nonincreasing and  $\delta_n \xrightarrow{n \rightarrow \infty} 0$ .

Thus, to ensure that (17) holds, it is enough to choose  $(\gamma_n)_n$  so that

$$(18) \quad \sum_{n=1}^N \mathbf{1}(\gamma_n > 0) d_n (\gamma_n + \varepsilon_n) \leq \log_2 f(t_{N-1}) \quad \text{for each } N > 1.$$

Next we derive a sufficient condition to be verified in order to show that  $\hat{\sigma}$  has the property (ii). Fix an arbitrary strategy  $\tau \in \Sigma_2$ . For each  $n = 1, 2, \dots$ , let  $(x_1^{(n)}, y_1^{(n)}), \dots, (x_{d_n}^{(n)}, y_{d_n}^{(n)})$  be the random action pairs induced by  $(\hat{\sigma}, \tau)$  in the  $n$ -th block. Fix a realization  $h_{n-1} \in A^{t_{n-1}}$  of  $(x_1^{(1)}, y_1^{(1)}), \dots, (x_{d_{n-1}}^{(n-1)}, y_{d_{n-1}}^{(n-1)})$  and let  $(\bar{x}^{(n)}, \bar{y}^{(n)})$  be a  $(A_1 \times A_2)$ -valued random variable such that

$$\mathbf{P}((\bar{x}^{(n)}, \bar{y}^{(n)}) = (a, b)) := \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{P}((x_k^{(n)}, y_k^{(n)}) = (a, b) | h_{n-1}).$$

As  $\mathbf{P}(x_k^{(n)} = a) = \mathbf{E}[\mathbf{1}(z_{t_{n-1}+k} = a) | h_{n-1}]$  for each  $a \in A_1$  and  $k = 1, \dots, d_n$ , and since, conditional on  $h_{n-1}$ ,  $\mathbf{z}[n] = (z_{t_{n-1}+1}, \dots, z_{t_n}) \in F_n$  with probability 1,

<sup>16</sup>Observe that (i) asserts that  $\psi_1(t) \leq f(t)$  for all  $t$  which will be satisfied, as these functions are nondecreasing, if  $\psi_1(t_N) \leq f(t_{N-1})$  for each  $N$ .

it follows that

(19)

$$\begin{aligned} \sum_{a \in A_1} |\mathbf{P}(\bar{x}^{(n)} = a) - \alpha_n(a)| &= \sum_{a \in A_1} \left| \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{E}[\mathbf{1}(z_{t_{n-1}+k} = a) | h_{n-1}] - \alpha_n(a) \right| \\ &\leq \mathbf{E} \left[ \sum_{a \in A_1} \left| \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(z_{t_{n-1}+k} = a) - \alpha_n(a) \right| \middle| h_{n-1} \right] \\ &\leq \frac{|A_1|}{d_n} < \delta_n. \end{aligned}$$

Next, since  $\sum_{k=1}^d p_k \log_2 \frac{p_k}{q_k} \geq \left( \sum_{k=1}^d p_k \right) \log_2 \frac{\sum_{k=1}^d p_k}{\sum_{k=1}^d q_k}$  for any nonnegative numbers  $p_1, \dots, p_d$  and  $q_1, \dots, q_d$  -this is called the log-sum inequality- and since conditioning reduces entropy, we have

$$\begin{aligned} H(\bar{x}^{(n)} | \bar{y}^{(n)}) &\geq \frac{1}{d_n} \sum_{k=1}^{d_n} H(x_k^{(n)} | y_k^{(n)}, h_{n-1}) \\ &\geq \frac{1}{d_n} \sum_{k=1}^{d_n} H(x_k^{(n)} | y_k^{(n)}, h_{n-1}, x_1^{(n)}, \dots, x_{k-1}^{(n)}) \\ &= \frac{1}{d_n} \sum_{k=1}^{d_n} H(x_k^{(n)} | h_{n-1}, x_1^{(n)}, \dots, x_{k-1}^{(n)}) \end{aligned}$$

where the last equality holds due to the fact that  $\tau$  is a pure strategy and hence  $y_k^{(n)}$  is a deterministic function of  $h_{n-1}$  and  $x_1^{(n)}, \dots, x_{k-1}^{(n)}$ . Using the chain rule for entropy, and since, conditional on  $h_{n-1}$ , the string  $(z_{t_{n-1}+1}, \dots, z_{t_n})$  is chosen uniformly from  $F_n$ , the last expression above is equal to

$$\frac{H(x_1^{(n)}, \dots, x_{d_n}^{(n)} | h_{n-1})}{d_n} = \frac{H(z_{t_{n-1}+1}, \dots, z_{t_n} | h_{n-1})}{d_n} = \frac{\log_2 |F_n|}{d_n} \geq \gamma_n - \varepsilon_n.$$

As  $\gamma_n = H(\alpha_n)$  and  $\delta_n > \varepsilon_n$ , we conclude that

$$(20) \quad H(\bar{x}^{(n)} | \bar{y}^{(n)}) > H(\alpha_n) - \delta_n.$$

It can be shown (see Appendix) that there is a function  $\eta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , which depends on the stage game, with  $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$  such that if  $(x, y)$  is a  $(A_1 \times A_2)$ -valued random variable satisfying (i)  $H(x|y) \geq H(\alpha^*) - \delta$  and (ii)  $\sum_{a \in A_1} |\mathbf{P}(x = a) - \alpha^*(a)| <$

$\delta$ , then  $\mathbf{E}[g(x, y)] \geq \min_{b \in A_2} g(\alpha^*, b) - \eta(\delta)$ . Thus, the inequalities (19) and (20) imply that there is a function  $\eta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$  such that

$$\mathbf{E} \left[ \frac{1}{d_n} \sum_{k=1}^{d_n} g(x_k^{(n)}, y_k^{(n)}) \middle| h_{n-1} \right] = \mathbf{E} [g(\bar{x}^{(n)}, \bar{y}^{(n)}) | h_{n-1}] \geq U(\gamma_n) - \eta(\delta_n).$$

Recall that  $\min_{b \in A_2} g(\alpha_n, b) = U(\gamma_n)$ . As this holds for any  $\tau$ ,  $n$ , and  $h_{n-1}$ , it follows that, for any  $N = 1, 2, \dots$ ,

$$\min_{\tau \in \Sigma_2} g_{t_N}(\hat{\sigma}, \tau) = \frac{1}{t_N} \sum_{n=1}^N \mathbf{E} \left[ \sum_{k=1}^{d_n} g(x_k^{(n)}, y_k^{(n)}) \right] \geq \frac{1}{t_N} \sum_{n=1}^N d_n U(\gamma_n) - \frac{1}{t_N} \sum_{n=1}^N d_n \eta(\delta_n).$$

Since  $t_N = \sum_{n=1}^N d_n$  and  $\eta(\delta_n) \xrightarrow{n \rightarrow \infty} 0$ , the second term on the right side converges to 0 as  $N \rightarrow \infty$ . In addition, recall that  $\frac{d_N}{t_N} \xrightarrow{N \rightarrow \infty} 0$ . Hence, in order to show part (ii) of the theorem, it suffices to choose  $(\gamma_n)_n$  so that

$$(21) \quad \frac{1}{t_N} \sum_{n=1}^N d_n U(\gamma_n) \xrightarrow{N \rightarrow \infty} (\text{cav } U)(\gamma).$$

We now exhibit a choice of  $(\gamma_n)_n$  that satisfy (17) and (21). We distinguish two cases.

CASE 1 -  $(\text{cav } U)(\gamma) = U(\gamma)$ : Since  $\frac{d_{n-1}}{d_n} \xrightarrow{n \rightarrow \infty} 1$ , we have  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} \xrightarrow{n \rightarrow \infty} \gamma > 0$ . Let  $\bar{n}$  be sufficiently large so that  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} > \varepsilon_n$  for all  $n > \bar{n}$ . Then set  $\gamma_n = 0$  for  $n \leq \bar{n}$  and  $\gamma_n = \frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-1})}{t_{n-1}} - \varepsilon_n$  for  $n > \bar{n}$ . With this choice of  $(\gamma_n)_n$  it is easy to verify that (18), and hence (17), is satisfied. As  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ , we have  $\gamma_n \xrightarrow{n \rightarrow \infty} \gamma$ . Thus the condition (21) is satisfied as well.

CASE 2 -  $(\text{cav } U)(\gamma) > U(\gamma)$ : In this case, the definitions of  $U$  and  $\text{cav } U$  imply the existence of  $\gamma_-, \gamma_+$  with  $0 \leq \gamma_- < \gamma < \gamma_+$  and  $\alpha_-, \alpha_+ \in \Delta(A_1)$  together with a  $p \in (0, 1)$  such that

- a)  $\gamma = p\gamma_- + (1-p)\gamma_+$
- b)  $(\text{cav } U)(\gamma) = pU(\gamma_-) + (1-p)U(\gamma_+)$
- c)  $H(\alpha_-) \leq \gamma_-$  and  $H(\alpha_+) \leq \gamma_+$
- d)  $g(\alpha_-, b) \geq U(\gamma_-)$  and  $g(\alpha_+, b) \geq U(\gamma_+)$  for all  $b \in A_2$ .

Choose  $\bar{n}$  large enough so that for  $n \geq \bar{n}$  we have  $\frac{d_{n-1}}{d_n} \frac{\log_2 f(t_{n-2})}{t_{n-2}} - \varepsilon_n > \gamma_-$ . Set  $\gamma_n = 0$  for  $n \leq \bar{n}$  and, for  $n > \bar{n}$ , define  $\gamma_n$  by induction as follows:

$$\gamma_n = \begin{cases} \gamma_+ & \text{if } \sum_{\ell=1}^{n-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) + d_n (\gamma_+ + \varepsilon_n) \leq \log_2 f(t_{n-1}), \\ \gamma_- & \text{otherwise.} \end{cases}$$

With the above choices of the sequence  $(\gamma_n)_n$ , we have  $|F_n| = 1$  for  $n \leq \bar{n}$ . So for  $N \leq \bar{n}$ , the inequality (17) trivially holds. For  $N > \bar{n}$ , inequality (18) is proved by induction. Assume that  $\sum_{\ell=1}^{N-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) \leq \log_2 f(t_{N-2})$ . Then,

$$\begin{aligned} \sum_{\ell=1}^{N-1} \mathbf{1}(\gamma_\ell > 0) d_\ell (\gamma_\ell + \varepsilon_\ell) + d_N (\gamma_- + \varepsilon_N) &\leq \log_2 f(t_{N-2}) + d_{N-1} \frac{\log_2 f(t_{N-2})}{t_{N-2}} \\ &\leq \log_2 f(t_{N-1}) \end{aligned}$$

where the first inequality holds by the induction hypothesis and since  $N > \bar{n}$ , while the second inequality follows from (14). Therefore, if  $\gamma_N = \gamma_-$  then inequality (18) holds for  $N$ . Obviously, by the definition of  $\gamma_N$ , if  $\gamma_N = \gamma_+$  then inequality (18) again holds for  $N$ . We conclude that inequality (18), hence (17), holds for all  $N > \bar{n}$ .

Since  $\gamma_n = \gamma_-$  or  $\gamma_+$  for all  $n > \bar{n}$ , in order to show (21), it suffices to verify that

$$(22) \quad \frac{1}{t_N} \sum_{n=\bar{n}+1}^N d_n \mathbf{1}(\gamma_n = \gamma_-) \xrightarrow{N \rightarrow \infty} p.$$

For each  $N$  let  $M_N = \max\{n \leq N : \gamma_n = \gamma_-\}$ . Note that  $M_N \rightarrow \infty$  as  $N \rightarrow \infty$  since  $\gamma_+ > \gamma$ . Since  $\frac{\log_2 f(t_{n-1})}{t_n} = \left(1 - \frac{d_n}{t_n}\right) \frac{\log_2 f(t_{n-1})}{t_{n-1}} \xrightarrow{n \rightarrow \infty} \gamma$ , we have, for every  $\delta > 0$ , an  $N$  such that  $\frac{\log_2 f(t_{n-1})}{t_n} \geq \gamma - \delta$  for all  $n \geq M_N$ . Hence

$$\begin{aligned} &\sum_{n=1}^N \mathbf{1}(\gamma_n > 0) d_n (\gamma_n + \varepsilon_n) \\ &= \sum_{n=1}^{M_N} \mathbf{1}(\gamma_n > 0) d_n (\gamma_n + \varepsilon_n) + \sum_{n=M_N+1}^N \mathbf{1}(\gamma_n > 0) d_n (\gamma_n + \varepsilon_n) \\ &\geq \log_2 f(t_{M_N-1}) - d_{M_N} (\gamma_+ - \gamma_-) + (t_N - t_{M_N}) \gamma_+ \\ &\geq t_{M_N} (\gamma - \delta) - d_N (\gamma_+ - \gamma_-) + (t_N - t_{M_N}) \gamma \\ &\geq t_N (\gamma - \delta) - d_N (\gamma_+ - \gamma_-). \end{aligned}$$

Since  $\sum_{n=1}^N 1(\gamma_n > 0)d_n(\gamma_n + \varepsilon_n) \leq \log f(t_{N-1})$ ,  $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ , and  $\frac{d_N}{t_N} \xrightarrow{N \rightarrow \infty} 0$ , we conclude that  $\frac{1}{t_N} \sum_{n=1}^N d_n \gamma_n \xrightarrow{N \rightarrow \infty} \gamma$  which implies (22). **Q.E.D.**

*Remark 3.1.* An additional property of  $\psi_1$  constructed in the above proof is that  $\lim_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} = \gamma$ . Although this can be verified by examining the details of the proof, an alternative derivation of it illuminates a connection between the growth of strategy set and entropy.

On the one hand, the property (i) in Theorem 4 implies that  $\overline{\lim}_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} \leq \gamma$ . On the other hand, the property (ii) and our previously published result on strategic entropy (Neyman and Okada (2000a)) imply that  $\underline{\lim}_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} \geq \gamma$ . To see this, let us recall that the  $t$ -strategic entropy of player 1's strategy  $\sigma$  is defined by  $H^t(\sigma) = \max_{\tau \in \Sigma_2} H(X_1, \dots, X_t)$  where  $X_1, \dots, X_t$  is the random sequence of action profiles induced by  $(\sigma, \tau)$ . Lemma 1 in Section 3.3 then implies that  $H^T(\sigma) \leq \log_2 \psi_1(T)$  for all  $\sigma \in \Delta(\Psi_1)$  and  $T$ . This, together with Theorem 5.1 of Neyman and Okada (2000a), implies that

$$\inf_{\tau \in \Sigma_2} g_T(\sigma, \tau) \leq (\text{cav } U) \left( \frac{\log_2 \psi_1(T)}{T} \right).$$

for all  $\sigma \in \Delta(\Psi_1)$  and  $T$ . Thus, if  $\underline{\lim}_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} < \gamma$ , then  $\underline{\lim}_{T \rightarrow \infty} \left( \inf_{\tau \in \Sigma_2} g_T(\sigma, \tau) \right) < (\text{cav } U)(\gamma)$  contradicting property (ii) of Theorem 4. Hence we conclude that  $\lim_{t \rightarrow \infty} \frac{\log_2 \psi_1(t)}{t} = \gamma$ .

#### 4. NONSTATIONARY BOUNDED RECALL STRATEGIES

In this section, we study a concrete case of the game examined in the last section. Specifically, player 1's feasible strategy set is taken to be  $\mathbf{B}_1(\kappa) = \{\sigma \wedge \kappa : \sigma \in \Sigma_1\}$ , the set of  $\kappa$ -recall strategies. Player 2 is assumed to have full recall. Let  $G^*(\kappa)$  be the repeated game under consideration and let  $V(\kappa)$  be player 1's minimax payoff in  $G^*(\kappa)$ .

Recall that, with  $\Psi_1 = \mathbf{B}_1(\kappa)$ , we have  $\log_2 \psi_1(t) \leq c m^{\bar{\kappa}(t)}$  for a constant  $c$  (e.g.  $c = m/(m-1)$ ) where  $\bar{\kappa}(t) = \max_{s \leq t} \kappa(s)$ . Suppose that, for every  $\varepsilon > 0$ , we have  $\kappa(t) < (1-\varepsilon) \frac{\log_2 t}{\log_2 m}$  for sufficiently large  $t$ . Then it follows that<sup>17</sup>  $\frac{\log_2 \psi_1(t)}{t} \xrightarrow{t \rightarrow \infty}$

<sup>17</sup>For the conclusion it suffices to have  $\kappa(t) < \frac{\log_2 t}{\log_2 m} - c(t)$  where  $c(t) \xrightarrow{t \rightarrow \infty} \infty$ .

0. Hence, by Theorem 2 together with the fact that player 1 can always guarantee  $w = \max_{a \in A_1} \min_{a_2 \in A_2} g(a, a_2)$  with a stationary bounded recall strategy of size 0, we obtain the following result.

**Theorem 5.** *If  $\overline{\lim}_{t \rightarrow \infty} \frac{\kappa(t)}{\log_2 t} < \frac{1}{\log_2 m}$ , then  $V(\kappa) = w$ .*

This suggests that, in order to gain any benefit from recalling the past (to get a payoff above  $w$ ) against a player with perfect recollection, one must remember at least some sufficiently large constant times  $\log_2 t$  stages back. It is thus natural to ask, “how fast should  $\kappa$  grow (asymptotically) in order to guarantee the minimax payoff  $v$  against a player with full recall?”. This question will be answered in the next theorem. It asserts that, in order to secure  $v$  in the long run, it suffices that player 1’s recall  $\kappa(t)$  grows at least as fast as  $K_1 \log_2 t$  for some  $K_1 > 0$ . (Of course,  $K_1 \geq 1/\log_2 m$ ) Hence, player 1 can guarantee the minimax payoff in the long run by recalling a vanishing fraction of the history even against a player with full recall.

In order to exhibit the constant  $K_1$  explicitly, let  $\zeta(G) = \max_{\alpha} \max_{a \in A_1} \alpha(a)$  where  $\alpha$  is taken over all mixed actions of player 1 in the stage game  $G$  with  $\min_{b \in A_2} g(\alpha, b) = v$ . For example,  $\zeta(G) = 1$  if the minimax payoff  $v$  can be secured by a pure action. Define

$$K_1(G) = \begin{cases} 0 & \text{if } \zeta(G) = 1, \\ \frac{2}{|\log_2 \zeta(G)|} & \text{if } \zeta(G) < 1. \end{cases}$$

For instance, in matching pennies,  $\zeta(G) = 1/2$  and so  $K_1(G) = 2$ .

**Theorem 6.** *If  $\underline{\lim}_{t \rightarrow \infty} \frac{\kappa(t)}{\log_2 t} > K_1(G)$ , then there is a  $\hat{\sigma} \in \Delta(\mathbf{B}_1(\kappa))$  with the following properties.*

- (i)  $\underline{\lim}_{T \rightarrow \infty} \left( \min_{\tau \in \Sigma_2} g_T(\hat{\sigma}, \tau) \right) \geq v$ ,
- (ii)  $\inf_{\tau \in \Sigma_2} \mathbf{E}_{\hat{\sigma}, \tau} \left[ \underline{\lim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g(a_t, b_t) \right] \geq v$ .

**Proof:** Let  $\alpha^* \in \Delta(A_1)$  be such that  $\min_{b \in A_2} g(\alpha^*, b) = v$  and  $\max_{a \in A_1} \alpha^*(a) = \zeta(G)$ . If  $\alpha^*$  is a pure action, then the theorem is trivially true. So suppose that  $\alpha^*$  is not a pure action. Then there are two distinct actions  $a^0, a^1 \in A_1$  with  $\alpha^*(a^0) > 0$  and  $\alpha^*(a^1) > 0$ .



In order to define the strategy  $\hat{\sigma}$ , we introduce some notation. First, our condition on  $\kappa$  implies that there are infinitely many stages  $t$  at which  $\kappa(t) \neq \kappa(s)$  for all  $s < t$ . We enumerate these stages as  $t_1 < t_2 < \dots < t_n < \dots$ . For each sequence of player 1's actions  $\mathbf{a} = (a_1, a_2, \dots) \in A_1^\infty$  and each positive integer  $n$ , define a sequence  $\mathbf{a}^n = (a_1^n, a_2^n, \dots) \in A_1^\infty$  by

$$a_t^n = \begin{cases} a^0 & \text{if } t < t_n, \\ a^1 & \text{if } t = t_n, \\ a_t & \text{if } t > t_n. \end{cases}$$

Thus, if player 1 is to play according to the oblivious strategy  $\sigma\langle \mathbf{a}^n \rangle$ , he would take the action  $a^0$  in the first  $t_n - 1$  stages, then  $a^1$  at stage  $t_n$ , and thereafter the actions appearing in the original sequence  $\mathbf{a}$ .

Suppose that the sequence  $\mathbf{a}$  satisfies the following property for some  $n$ .

$$(23) \quad \forall t, t' \text{ s.t. } a_t^n \neq a_{t'}^n : (a_{t'-\kappa(t')}^n, \dots, a_{t'-1}^n) \neq (a_{t-\kappa(t)}^n, \dots, a_{t-1}^n).$$

Then the sequence of actions  $\mathbf{a}^n$  is induced by the  $\kappa$ -recall (oblivious) strategy  $\sigma\langle \mathbf{a}^n \rangle \wedge \kappa$ . From the definition of the modified sequence  $\mathbf{a}^n$  and since  $\kappa(t_n) \neq \kappa(t)$  for all  $t < t_n$ , the condition (23) is implied by

$$(24) \quad \forall t' \text{ s.t. } t' - \kappa(t') > t_n \forall t < t' : (a_{t'-\kappa(t')}^n, \dots, a_{t'-1}^n) \neq (a_{t-\kappa(t)}^n, \dots, a_{t-1}^n).$$

Observe that, for any  $n$ ,  $t$  and  $k$ ,  $(a_{t-k}^n, \dots, a_{t-1}^n)$  is one of at most  $k + 2$  strings of length  $k$ :  $(a^0, \dots, a^0)$ ,  $(a^0, \dots, a^0, a^1)$ ,  $(a^0, \dots, a^0, a^1, a_{t-\ell}, \dots, a_{t-1})$  where  $1 \leq \ell \leq k - 2$ ,  $(a^1, a_{t-k+1}, \dots, a_{t-1})$ , and  $(a_{t-k}, \dots, a_{t-1})$ . Let us denote the subset of  $A_1^k$  consisting of these strings by  $Z(a_{t-k}, \dots, a_{t-1})$  so that (24), hence (23), is further implied by

$$(25) \quad \forall t' \text{ s.t. } t' - \kappa(t') > t_n \forall t < t' : (a_{t'-\kappa(t')}^n, \dots, a_{t'-1}^n) \notin Z(a_{t-\kappa(t)}, \dots, a_{t-1}).$$

Our  $\hat{\sigma}$  of the theorem will be a mixture over  $\{\sigma\langle \mathbf{a}^n \rangle \wedge \kappa : \mathbf{a} \in A_1^\infty, n = 1, 2, \dots\}$  where actions in the sequence  $\mathbf{a} = (a_1, a_2, \dots)$  are chosen independently according to  $\alpha^*$  and then choosing an  $n$  with the property (25). Of course, we must make sure that such an  $n < \infty$  exists for almost all realizations of the sequence. Formally, let  $\mathbf{z} = (z_1, z_2, \dots)$  be a sequence of  $A_1$ -valued i.i.d. r.v.'s with  $z_t \sim \alpha^*$ , and, for each

$n$ , define a sequence of  $A_1$ -valued r.v.'s  $\mathbf{z}^n = (z_1^n, z_2^n, \dots)$  by

$$z_t^n = \begin{cases} a^0 & \text{if } t < t_n, \\ a^1 & \text{if } t = t_n, \\ z_t & \text{if } t > t_n. \end{cases}$$

Next define a  $\mathbb{N}$ -valued r.v.  $\nu$  by  $\nu = n$  if  $n$  is the smallest positive integer with the property

$$(26) \quad \forall t' \text{ s.t. } t' - \kappa(t') > t_n, \quad \forall t < t' : (z_{t'-\kappa(t')}, \dots, z_{t'-1}) \notin Z(z_{t-\kappa(t)}, \dots, z_{t-1}).$$

Then define  $\hat{\sigma} = \sigma(\mathbf{z}^\nu) \wedge \kappa$ . Below we will show that  $\nu < \infty$  almost surely under the condition on  $\kappa(t)$  stated in the theorem, and hence  $\hat{\sigma}$  is well-defined as a mixture of strategies in  $\{\sigma(\mathbf{a}^n) \wedge \kappa : \mathbf{a} \in A_1^\infty, n = 1, 2, \dots\} \subset \mathbf{B}_1(\kappa)$ .

To see this, observe that

$$\begin{aligned} & \mathbf{E} \left[ \left| \left\{ (t, t') : t < t', (z_{t'-\kappa(t')}, \dots, z_{t'-1}) \in Z(z_{t-\kappa(t)}, \dots, z_{t-1}) \right\} \right| \right] \\ &= \mathbf{E} \left[ \sum_{k=1}^{\infty} \sum_{t < t'} \mathbf{1}(\kappa(t) = \kappa(t') = k) \mathbf{1}((z_{t'-k}, \dots, z_{t'-1}) \in Z(z_{t-k}, \dots, z_{t-1})) \right] \\ &= \sum_{k=1}^{\infty} \sum_{t < t'} \mathbf{1}(\kappa(t) = \kappa(t') = k) \mathbf{P}((z_{t'-k}, \dots, z_{t'-1}) \in Z(z_{t-k}, \dots, z_{t-1})) \\ &\leq \sum_{k=1}^{\infty} (k+2) |B_k|^2 \zeta(G)^k \quad (\text{where } B_k = \{t : \kappa(t) = k\}) \\ &\leq 2 \sum_{t=1}^{\infty} t(\kappa(t) + 2) \zeta(G)^{\kappa(t)}. \end{aligned}$$

For the second-to-the-last inequality, we used the fact that  $|Z(z_{t-k}, \dots, z_{t-1})| \leq k+2$ , as observed above. The last inequality follows from the inequality  $|B_k|^2 < 2 \sum_{t \in B_k} t$ . Our condition on  $\kappa$  implies that there is an  $\varepsilon > 0$  and a  $\hat{t}$  such that  $\kappa(t) \geq (K_1(G) + 2\varepsilon) \log_2 t$  for all  $t \geq \hat{t}$ . Therefore there is a  $0 < \theta < 1$  and a  $\tilde{t} \geq \hat{t}$  such that  $\theta \kappa(t) \geq (K_1(G) + \varepsilon) \log_2 t$  and  $(\kappa(t) + 2) \zeta(G)^{(1-\theta)\kappa(t)} < 1$  for all  $t \geq \tilde{t}$ . As  $\zeta(G) < 1$ , it follows that  $t(\kappa(t) + 2) \zeta(G)^{\kappa(t)} \leq t^{-(1+\varepsilon|\log_2 \zeta(G)|)}$  for all  $t \geq \tilde{t}$ . Hence the series on the last line is convergent. Therefore, with probability 1, there are only finitely many pairs  $(t, t')$  with  $t < t'$  and  $(z_{t'-\kappa(t')}, \dots, z_{t'-1}) \in Z(z_{t-\kappa(t)}, \dots, z_{t-1})$ . Thus  $\nu < \infty$  almost surely.

Next we verify that  $\hat{\sigma}$  has the desired properties. Fix an arbitrary pure strategy  $\tau \in \Sigma_2$  and a stage  $T$ . Let  $(x_1, y_1), \dots, (x_T, y_T)$  be the random action pairs induced by  $(\hat{\sigma}, \tau)$ . Let  $(\bar{x}_T, \bar{y}_T)$  be a  $(A_1 \times A_2)$ -valued random variable such that  $\mathbb{P}((\bar{x}_T, \bar{y}_T) = (a, b)) = \frac{1}{T} \sum_{t=1}^T \mathbb{P}((x_t, y_t) = (a, b))$ . Then, as in the proof of Theorem 4, we have

$$\begin{aligned}
 H(\bar{x}_T | \bar{y}_T) &\geq \frac{1}{T} \sum_{t=1}^T H(x_t | y_t) \geq \frac{1}{T} \sum_{t=1}^T H(x_t | y_t, x_1, \dots, x_{t-1}) \\
 (27) \qquad &= \frac{1}{T} \sum_{t=1}^T H(x_t | x_1, \dots, x_{t-1}) \\
 &\geq \frac{1}{T} \sum_{t=1}^T H(x_t | x_1, \dots, x_{t-1}, t_\nu \wedge (T+1)).
 \end{aligned}$$

where the equality follows from  $y_t$  being a deterministic function of  $x_1, \dots, x_{t-1}$ .

Observe that  $z_t \in A_1$  and thus a conditional entropy of  $z_t$  is bounded by  $\log_2 |A_1|$ , and, conditional on  $t_\nu < t$  we have  $z_t = z_t' = x_t$ . Therefore,  $H(z_t | z_1, \dots, z_{t-1}, t_\nu \wedge (T+1)) \leq H(x_t | x_1, \dots, x_{t-1}, t_\nu \wedge (T+1)) + \log_2 |A_2| \mathbb{P}(t \leq t_\nu)$ . By rearranging the terms we have

$$H(x_t | x_1, \dots, x_{t-1}, t_\nu \wedge (T+1)) \geq H(z_t | z_1, \dots, z_{t-1}, t_\nu \wedge (T+1)) - \log_2 |A_1| \mathbb{P}(t \leq t_\nu).$$

From the chain of inequalities beginning (27), and further applying the chain rule for entropy and using the fact that  $t_\nu \wedge (T+1)$  takes at most  $T+1$  distinct values, we obtain

$$\begin{aligned}
 H(\bar{x}_T | \bar{y}_T) &\geq \frac{1}{T} \sum_{t=1}^T H(z_t | z_1, \dots, z_{t-1}, t_\nu \wedge (T+1)) - \frac{1}{T} \sum_{t=1}^T \log_2 |A_1| \mathbb{P}(t \leq t_\nu) \\
 &= \frac{1}{T} H(z_1, \dots, z_T | t_\nu \wedge (T+1)) - \frac{1}{T} \sum_{t=1}^T \log_2 |A_1| \mathbb{P}(t \leq t_\nu) \\
 &\geq \frac{1}{T} H(z_1, \dots, z_T) - \frac{1}{T} \log_2(T+1) - \frac{1}{T} \sum_{t=1}^T \log_2 |A_1| \mathbb{P}(t \leq t_\nu) \\
 &= H(\alpha^*) - o(1) \xrightarrow{T \rightarrow \infty} H(\alpha^*).
 \end{aligned}$$

Note that the  $o(1)$  function in the last line is independent of  $\tau$ . To summarize, for every  $\delta > 0$  there is a  $T_0$  such that, for every  $T \geq T_0$  and  $\tau \in \Sigma_2$ , the average empirical distribution of action pairs in the first  $T$  stages of the game (i.e. the

distribution of  $(\bar{x}_T, \bar{y}_T)$  obeys

$$(28) \quad H(\bar{x}_T | \bar{y}_T) \geq H(\alpha^*) - \delta.$$

Next we demonstrate that the distribution of  $\bar{x}_T$  is close to  $\alpha^*$ . Since  $x_t = z_t$  whenever  $t_\nu < t$ , we have, for each  $a \in A_1$ ,

$$P(x_t = a) \geq P(z_t = a) - P(t_\nu \geq t) = \alpha^*(a) - P(t_\nu \geq t)$$

and so  $\sum_{a \in A} |P(x_t = a) - \alpha^*(a)| \leq 2|A_1|P(t_\nu \geq t)$ . Hence

$$\begin{aligned} \sum_{a \in A_1} |P(\bar{x}_T = a) - \alpha^*(a)| &= \sum_{a \in A_1} \left| \frac{1}{T} \sum_{t=1}^T P(x_t = a) - \alpha^*(a) \right| \\ &\leq 2|A_1| \frac{1}{T} \sum_{t=1}^T P(t_\nu \geq t) \xrightarrow{T \rightarrow \infty} 0. \end{aligned}$$

Thus, for every  $\delta > 0$  there is a  $T_0$  such that for all  $T \geq T_0$  we have

$$(29) \quad \sum_{a \in A_1} |P(\bar{x}_T = a) - \alpha^*(a)| < \delta.$$

Hence, by the same lemma (see Appendix) as we used in the proof of Theorem 4, the inequalities (28) and (29) imply that, for every  $\varepsilon > 0$  there is a  $T_0$  such that for all  $T \geq T_0$ ,

$$g_T(\hat{\sigma}, \tau) = \frac{1}{T} \sum_{t=1}^T \mathbf{E}_{\hat{\sigma}, \tau} [g(x_t, y_t)] = \mathbf{E}_{\hat{\sigma}, \tau} [g(\bar{x}_T, \bar{y}_T)] \geq \min_{b \in A_2} g(\alpha^*, b) - \varepsilon = v - \varepsilon.$$

This completes the first part of the theorem.

In order to deduce the second part, observe that, by performing the same line of argument as above but conditional on a history  $(x_1, y_1, \dots, x_s, y_s)$ , we can show that for every  $\varepsilon > 0$  there is a  $T_0$  such that for every strategy  $\tau \in \Sigma_2$ , every positive integer  $s$  and every  $T \geq T_0$ , we have

$$\mathbf{E}_{\hat{\sigma}, \tau} \left[ \frac{1}{T} \sum_{t=s+1}^{s+T} g(x_t, y_t) \middle| x_1, y_1, \dots, x_s, y_s \right] \geq v - \varepsilon$$

which, by the classical results in probability, implies that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T g(x_t, y_t) \geq v - \varepsilon \quad \text{almost surely.}$$

As this holds for every  $\varepsilon > 0$  the second part of the theorem follows. **Q.E.D.**

*Remark 4.1.* The strategy  $\hat{\sigma}$  constructed in the above proof relies on the random variable  $\nu$  which depends on the vales of the entire sequence  $(z_t)_t$ . (In particular,  $\nu$  is not a stopping time.) A slightly weaker result can be derived as follows.

Since  $\nu < \infty$  almost surely, we have  $P(\nu \leq n) \xrightarrow[n \rightarrow \infty]{} 1$ . Hence if we choose an  $n$  sufficiently large, the condition (26) holds with a probability close to 1. Therefore, for every  $\varepsilon > 0$ , there is an  $n$  and a  $T_0$  such that for every  $\tau \in \Sigma_2$  and  $T \geq T_0$ , we have  $g_T(\sigma\langle \mathbf{z}^n \rangle, \tau) > v - \varepsilon$ . In other words, for every  $\varepsilon > 0$ , the strategy  $\sigma\langle \mathbf{z}^n \rangle$  is  $\varepsilon$ -optimal for a sufficiently large  $n$ . In the case  $G$  is a zerosum game, this shows that, under our condition on  $\kappa$ , the repeated game  $G^*(\kappa)$  has the value equal to  $v$ . Observe in contrast that  $\hat{\sigma}$  is optimal, that is,  $\varepsilon$ -optimal for every  $\varepsilon > 0$ .

Consider the repeated game  $G^*(\kappa_1, \kappa_2)$  where player  $i$ 's set of feasible strategies is  $\mathbf{B}_i(\kappa_i)$ ,  $i = 1, 2$ . Define  $K_2(G)$  analogously to  $K_1(G)$  above. If  $\underline{\lim}_{t \rightarrow \infty} \frac{\kappa_i(t)}{\log_2 t} > K_i(G)$  for  $i = 1, 2$ , then Theorem 6, or the strategy constructed in its proof, provides the players threats to one another which discourage them from deviating from a path that yields  $x_i \geq v_i$ ,  $i = 1, 2$ .

Let  $E^*(\kappa_1, \kappa_2)$  be the set of equilibrium payoff vectors of  $G^*(\kappa_1, \kappa_2)$  and let  $F^* = \{(x_1, x_2) \in \text{co}(g(A)) : x_1 \geq v^1, x_2 \geq v^2\}$ .

**Theorem 7.** *There is a constant  $K^*(G) > 0$ , that depends on the stage game, such that if  $\underline{\lim}_{t \rightarrow \infty} \min_{i=1,2} \frac{\kappa_i(t)}{\log_2 t} > K^*(G)$ , then  $E^*(\kappa_1, \kappa_2) = F^*$ .*

## APPENDIX

Here we will prove a statement used in the proof of Theorem 4 and Theorem 6. Recall that the  $L_1$ -distance between two probabilities  $P$  and  $Q$  on a finite set  $\Omega$  is  $\|P - Q\|_1 = \sum_{x \in \Omega} |P(x) - Q(x)|$ . For a  $(A_1 \times A_2)$ -valued random variable  $(x, y)$ , we write  $P(a, b)$  for  $P(x = a, y = b)$ , and  $P_1(a)$  (resp.  $P_2(b)$ ) for  $P(x = a)$  (resp.  $P(y = b)$ ). Also,  $P_1 \otimes P_2$  is a probability on  $A_1 \times A_2$  with  $(P_1 \otimes P_2)(a, b) = P_1(a)P_2(b)$  for each  $(a, b) \in A_1 \times A_2$ .

**Lemma 2.** *There exists a function  $\eta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , which depends on the stage game, with the following properties.*

- (1)  $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$ .
- (2) For any  $\alpha \in \Delta(A_1)$  and for any  $(A_1 \times A_2)$ -valued random variable  $(x, y)$  satisfying (i)  $H(x|y) \geq H(\alpha) - \delta$  and (ii)  $\|P_1 - \alpha\|_1 < \delta$ , we have  $\mathbf{E}[g(x, y)] \geq \min_{b \in A_2} g(\alpha, b) - \eta(\delta)$ .

**Proof:** We will show that, for small  $\delta > 0$ , (i) and (ii) imply that  $x$  and  $y$  are nearly independent, or, more precisely,  $P$  is close to  $P_1 \otimes P_2$  in the  $L_1$ -distance. As the expected payoff is continuous with respect to the  $L_1$ -distance on  $\Delta(A_1 \times A_2)$ , the conclusion of the lemma follows.

So suppose that (i) and (ii) are satisfied for a  $\delta > 0$ . Then, since conditioning reduces entropy, (i) implies that  $H(x) \geq H(x|y) \geq H(\alpha) - \delta$ . Next, since  $H$ , as a function on  $\Delta(A_1)$ , is uniformly continuous with respect to the  $L_1$ -norm, (ii) implies that  $H(x) \leq H(\alpha) + \theta(\delta)$  where<sup>18</sup>  $\theta(\delta) > 0$  and  $\theta(\delta) \xrightarrow{\delta \rightarrow 0} 0$ . Thus

$$(30) \quad H(x) - H(x|y) \leq \theta(\delta) + \delta.$$

Let us recall that the relative entropy between two probabilities  $P$  and  $Q$  on  $A_1 \times A_2$  is defined by  $D(P\|Q) = \sum_{a,b} P(a, b) \log_2 \frac{P(a, b)}{Q(a, b)}$  where, for any  $p, q > 0$ , we set  $0 \log_2 \frac{0}{q} \equiv 0$  and  $p \log_2 \frac{p}{0} \equiv \infty$ . The relative entropy is always nonnegative and equal to 0 if, and only if,  $P = Q$ . From this definition, it is easy to verify that  $D(P\|P_1 \otimes P_2) = H(x) - H(x|y)$ . Observe that  $D(P\|P_1 \otimes P_2) = 0$  if, and only if,  $x$  and  $y$  are independent. It can be shown (Cover and Thomas (1991), p.300) that

<sup>18</sup>In fact, one can take  $\theta(\delta) = -\delta \log_2 \frac{\delta}{|A_1|}$  for  $\delta \leq \frac{1}{2}$ . See Cover and Thomas (1991), p 488.

$D(\mathbf{P}||\mathbf{Q}) \geq \frac{1}{2 \ln 2} \|\mathbf{P} - \mathbf{Q}\|_1^2$ , and hence  $H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \geq \frac{1}{2 \ln 2} \|\mathbf{P} - \mathbf{P}_1 \otimes \mathbf{P}_2\|_1^2$ . It follows from (30) that

$$(31) \quad \|\mathbf{P} - \mathbf{P}_1 \otimes \mathbf{P}_2\|_1 = \sum_{a,b} |\mathbf{P}(a,b) - \mathbf{P}_1(a)\mathbf{P}_2(b)| \leq \sqrt{2 \ln 2 (\theta(\delta) + \delta)}.$$

Thus, setting  $\xi(\delta) = \sqrt{2(\theta(\delta) + \delta) \ln 2}$ , we have

$$\begin{aligned} \mathbf{E}[g(\mathbf{x}, \mathbf{y})] &= \sum_{a,b} \mathbf{P}(a,b)g(a,b) \\ &\geq \sum_{a,b} \mathbf{P}_1(a)\mathbf{P}_2(b)g(a,b) - |A_1 \times A_2| \|g\| \xi(\delta) && \text{by (31)} \\ &\geq \sum_{a,b} \alpha(a)\mathbf{P}_2(b)g(a,b) - |A_1| \|g\| \delta - |A_1 \times A_2| \|g\| \xi(\delta) && \text{by (ii)} \\ &\geq \min_b g(\alpha, b) - |A_1| \|g\| (\delta + |A_2| \xi(\delta)). \end{aligned}$$

This completes the proof.

**Q.E.D.**

## REFERENCES

- ANDERLINI, L. (1990): "Some Notes on Church's Thesis and the Theory of Games," *Theory and Decision*, 29(1), 19–52.
- AUMANN, R. J. (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, pp. 11–42. Bibliographisches Institut, Mannheim.
- (1997): "Rationality and Bounded Rationality," *Games and Economic Behavior*, 21(1-2), 2–14.
- AUMANN, R. J., AND L. S. SHAPLEY (1994): "Long-Term Competition - A Game Theoretic Analysis," in *Essays in Game Theory in Honor of Michael Maschler*, ed. by N. Megiddo, pp. 1 – 5. Springer-Verlag, New York.
- BEN-PORATH, E. (1993): "Repeated Games with Finite Automata," *Journal of Economic Theory*, 59(1), 17–32.
- COVER, T. M., AND J. A. THOMAS (1991): *Elements of Information Theory*. John Wiley & Sons, Inc., New York.
- FUDENBERG, D., AND E. MASKIN (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54(3), 533 – 554.
- HOEFFDING, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58(301), 13–30.
- LEHRER, E. (1988): "Repeated Games with Stationary Bounded Recall Strategies," *Journal of Economic Theory*, 46(1), 130–144.
- NEYMAN, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Economics Letters*, 19(3), 227–229.
- (1997): "Cooperation, Repetition, and Automata," in *Cooperation: Game-Theoretic Approaches*, ed. by S. H. Mas-Colell, and A., vol. 155 of *NATO ASI-Series F*, pp. 233–255. Springer-Verlag, New York.
- NEYMAN, A., AND D. OKADA (2000a): "Repeated Games with Bounded Entropy," *Games and Economic Behavior*, 30(2), 228 – 247.
- (2000b): "Two-Person Repeated Games with Finite Automata," *International Journal of Game Theory*, 29(3), 309 – 325.
- O'CONNELL, T. C., AND R. E. STEARNS (1999): "On Finite Strategy Sets for Finitely Repeated Zero-sum Games," Technical report, Department of Computer Science, SUNY at Albany.
- SIMON, H. A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69(1), 99–118.
- (1972): "Theories of Bounded Rationality," in *Decision and Organization*, ed. by C. McGuire, and R. Radner. North Holland, Amsterdam.
- STEARNS, R. E. (1997): "Memory-bounded Game-playing Computing Devices," Discussion Paper mimeo., Department of Computer Science, SUNY at Albany.