[This Version: June 6, 2005]

GROWING STRATEGY SETS AND NONSTATIONARY BOUNDED RECALL IN REPEATED GAMES

ABRAHAM NEYMAN† AND DAIJIRO OKADA‡

Abstract. In the existing literature on bounded rationality in repeated games, sets of feasible strategies are assumed to be independent of time (i.e. stage). In this paper we consider a time-dependent description of strategy sets, growing strategy sets. A growing strategy set is characterized by the way the set of strategies available to a player at each stage expands, possibly without bound, but not as fast as it would in the case of full rationality. Growing strategy sets are defined without regard to any specific complexity measure such as the number of states of automata or the length of recall. Rather, we focus on the number of distinct strategies available to a player up to stage t and how this number grows as a function of t.

We then study two-person zerosum games where one player's feasible strategy set grows slowly while playing against a fully rational player. We characterize the value of such games as a function of the growth rate of the strategy set. This function is related to the one that gives the minimax value of the stage game under the constraint that the payoff maximizing player chooses a mixed action with a given upper bound on entropy.

As an application of growing strategy set to strategic complexity, we also study repeated games with non-stationary bounded recall strategies where the length of recall is allowed to grow as the game progresses. We will show that a player with bounded recall can guarantee the value 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 value, it suffices to remember only a vanishing fraction of the past.

(*Journal of Economic Literature* Classification Number:)

Keywords: Growing Strategy Sets, Strategic Complexity, Repeated Games

† Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel.

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

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

Email: okada@econ.rutgers.edu.

1. Introduction

Many social (economic, political etc.) interactions have been modeled as formal games. Such a game consists of players, strategies and their preferences. 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). In this paper we take issues in specification of feasible, or implementable, strategies.

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 (implement) 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[1], 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.

The 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 (see, e.g. Aumann (1981), Aumann (1997)). There have been many attempts to model feasible (implementable) set of strategies that reflects 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: number of states of automata,

[1]See however Anderlini (1990).

the length of recall and the number of bits needed to implement a strategy by a particular type of Turing machine. [2]

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 other words, bounded rationality of players is expressed by a stationary (time-independent) description of their feasible strategies. 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. 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 the salient feature of dynamic decision making, i.e. time-dependent description of feasible strategies, that are not captured by the existing approaches.

Practically any economic decision maker, consumers, firms, government, trade and labor unions etc., is characterized 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 this view, the improvements of the sort alluded to above would enlarge the set of feasible decision rules. In economic applications, information relevant to decision making is often some data, signal,

[2]For example, Stearns (1997) employs linear bounded automata, Turing machines with bounded amount of tape.

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 (set of feasible strategies) 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 number of repetitions. (See Section 2.) 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,[3] 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 various measures of complexity of implementing strategies and long-term interaction among players with bounded strategic complexity.

Our aim in this paper is to take a first step into formalizing the idea of gradual relaxation 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 available (implementable) strategies, but computational resource available to the players may

[3]For instance, if a feasible set of strategies contains K distinct strategies, then one needs $\log K$ bits to encode all of them.

expand, e.g., by adding more memory over time or by learning. As a consequence, the limitation may ease over time and there may not be a finite upper bound on complexity of strategies for the entire horizon of the game. We attempt to capture perhaps a broadest implication of such reasoning by characterizing the feasible set of strategies by the way it expands over time, possibly without bound, but not as fast as it would in the case of full rationality.[4] To be more precise, we imagine player i with a feasible strategy set Ψ_i in a repeated game. Nature of the set Ψ_i is arbitrary. In particular, Ψ_i may contain infinite number of strategies some of which are very complex according to some measure. For each t, let $\Psi_i(t)$ be the projection of Ψ_i to the t-stage game and let $\psi_i(t)$ be the number of distinct strategies in $\Psi_i(t)$. If Ψ_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 Ψ_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 simplest model: repeated two-person zero-sum games in which player 1, the maximizer, has a set of feasible strategies growing in the manner mentioned above while all theoretically conceivable strategies are available to player 2. The payoff in the repeated games is the long-run average of the stage payoffs. In this setup we will in fact show that there is a continuous function, written cav U, such that player 1 cannot guarantee more than $(\text{cav } U)(\gamma)$ whenever $\psi_1(t)$ grows at most as fast as $2^{\gamma t}$. Moreover, for any $\gamma > 0$, we will explicitly construct a strategy set Ψ_1 for which $\psi_1(t) \leq 2^{\gamma t}$ and player 1 can guarantee $(\text{cav } U)(\gamma)$ using a mixture of strategies in Ψ_1. It will be seen that the function cav U is defined using the concept of entropy and has the property that $(\text{cav } U)(0)$ is the maximin value of the stage game in pure actions and $(\text{cav } U)(\gamma)$ is the value for sufficiently large γ.

[4]In this paper, we consider players whose strategy sets expand over time at an exogenously given rate. We recognize the importance of studying models where players may invest in order to expand their strategic possibilities thereby endogenizing the growth rate of their strategy sets. This certainly deserves further research.

As a concrete case of growing 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. In addition, we will show that there is a constant $K_1 > 0$ 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. The idea is that there is a population of players whose bounded rationality is expressed in terms of complexity of implementing strategies. Each player employs a pure strategy that is feasible. A mixed strategy is then viewed as a distribution of feasible pure strategies among the population. 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. A behavioral strategies should be viewed in this spirit. Because there is uncertainty about the type of her opponent, she also faces uncertainty about her opponent's action after each history.

We will set the notation used throughout the paper and formalize the idea of growing strategy sets in Section 2. Some examples of growing strategy sets, 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 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 $G = \big(A_i, g_i\big)_{i \in I}$ 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[5] 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. Thus the set of all histories is $A^* = \bigcup_{t=1}^{\infty} A^t$ where $A = \bigtimes_{i \in I} A_i$, and a pure strategy of player i is a mapping

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

Let Σ_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 a player i, σ_i and σ_i', are equivalent up to the t-th stage if, for every $(n-1)$-tuple of other players' strategies σ_{-i}, the profiles of actions induced by (σ_i, σ_{-i}) and (σ_i', σ_{-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 $m = m_1 \times \cdots \times m_n = |A|$. We note first that the number of strategies available to player i in the first t stages of a repeated game is[6]

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

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 of his strategies. For each positive integer t, let $\Psi_i(t)$ be formed by identifying strategies in Ψ_i that are equivalent up to the t-th stage.[7] Let $\psi_i(t)$ be the number

---

[5]In this paper we consider the most basic model of repeated games, i.e., ones with complete information, perfect monitoring and standard signaling.

[6]The number of reduced strategies available to player i in the first t stages is $m_i^{\frac{m_{-i}^t - 1}{m_{-i} - 1}}$ where $m_{-i} = \times_{j \neq i} m_j$.

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

of elements in $\Psi_i(t)$. Any consideration on strategic complexity gives rise to some strategy set Ψ_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 Ψ_i is a finite set and $\Psi_i(t) = \Psi_i$ for all sufficiently large[8] 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, or time-dependent, 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 $\sigma_i \bar{\wedge} k : A^* \to A_i$ by

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

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 A^k to A_i which is $m_i^{m^k}$.

Now consider a function $\kappa : \mathbb{N} \to \mathbb{N} \cup \{0\}$. For each $t \in \mathbb{N}$, the value $\kappa(t)$ represents the length of recall at stage t. If $\kappa(t) \geq t - 1$, then we have the case of full recall. So we will assume w.l.o.g. $\kappa(t) \leq t - 1$. In particular, $\kappa(1) = 0$. A κ-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 $\sigma_i \wedge \kappa : A^* \to A_i$ by

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

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

[8]In fact, this holds for all $t \geq k^2$ and $|\Psi_i| \approx 2^{ck \log k}$ where k is the bound on the number of states of automata.

in the last k stages. The set of κ-recall strategies is

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

From its definition it is easily seen that the number of equivalence classes of strategies in $\mathbf{B}_i(\kappa)$ is that of $\bigtimes_{k \in \kappa(\mathbb{N})} \bar{\mathbf{B}}(k)$ which is $\prod_{k \in \kappa(\mathbb{N})} m_i^{m^k}$. Here, for any $D \subset \mathbb{N}$, $\kappa(D) = \{\kappa(t) : t \in D\}$. So if $\Psi_i = \mathbf{B}_i(\kappa)$, then

$$\psi_i(t) = \prod_{k \in \kappa(\{1,\cdots,t\})} m_i^{m^k} = \prod_{k=1}^{\kappa(t)} m_i^{m^k} \leq m_i^{O\left(m^{\kappa(t)}\right)}. \tag{1}$$

This fact will be used later in Section 4

Example 2. A strategy of player i is said to be oblivious (Gossner and Hernandez (200?)) if it depends only on the history of his own actions. That is, $\sigma_i : A^* \to A_i$ is oblivious if

$$\sigma_i\big((a_{i1}, a_{-i1}), \cdots, (a_{it}, a_{-it})\big) = \sigma_i\big((b_{i1}, b_{-i1}), \cdots, (b_{it}, b_{-it})\big)$$

whenever $(a_{i1}, \cdots, a_{it}) = (b_{i1}, \cdots, b_{it})$. 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^∞. 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 $\boldsymbol{a} = (a_{i1}, a_{i2}, \cdots) \in A_i^\infty$, we denote by $\sigma_i\langle \boldsymbol{a} \rangle$ 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 positive integer k, 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 $\Psi_1 = \{\sigma_1^{(1)}, \sigma_1^{(2)}, \dots\}$. Then $\Psi_1(t) = \{\sigma_1^{(1)}, \dots, \sigma_1^{(t)}\}$ and $\psi_i(t) = t$.

Example 4. A prefix of a history $h = (h_1, \ldots, h_t)$ is any of its initial segment $h' = (h_1, \ldots, h_s)$, $s \leq t$. A set of histories $L \subset \cup_{t=1}^{\infty} H_t$ is said to be prefix-free if no element of L is a prefix of another. Now, for each positive integer t, let $L(t) \subset H_1 \cup \cdots \cup H_t$ be prefix-free and $L(t) \subset L(t+1)$. Write λ for the sequence $L(1), L(2), \cdots$, and define a strategy σ_1^{λ} as follows.

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

This is a generalization of the trigger strategy: σ_1^{λ} takes action 1 forever as soon as a history in some $L(t)$ occurs. Let $\mathcal{L}$ be the set of all increasing[9] sequences of prefix-free sets of histories. Take a subset $\mathcal{M}$ of $\mathcal{L}$ and define Ψ_1 to be the set of player 1's strategies σ_1^{λ} with $\lambda \in \mathcal{M}$. Let us examine $\Psi_1(t)$ and $\psi_1(t)$.

It is easy to verify that, for any $\lambda = (L(t))_t$ and $\mu = (M(t))_t$ in $\mathcal{L}$, σ_1^{λ} and σ_1^{μ} are equivalent up to the t-th stage if, and only if, $L(t) = M(t)$. We say that λ and μ are equivalent up to the t-th stage if $L(t) = M(t)$. This is an equivalence relation on $\mathcal{L}$, and hence on $\mathcal{M}$. We denote by $\mathcal{M}(t)$ the set of the equivalence classes when this relation is taken on $\mathcal{M}$. For notational simplicity, the elements of $\mathcal{M}(t)$ will be denoted by λ, μ and so on as for the elements of $\mathcal{M}$ themselves. Then we have

$$\Psi_1(t) = \{\sigma_1^{\lambda} \mid \lambda \in \mathcal{M}(t)\} \quad \text{and} \quad \psi_1(t) = |\mathcal{M}(t)|.$$

Examples of $\mathcal{M}$ can be constructed as follows. Let $f : \mathbb{N} \to \mathbb{N}$ be a nondecreasing function and let $\mathcal{M} = \{(L(t))_t \in \mathcal{L} \mid |L(t)| = O(f(t))\}$. It is not difficult to construct examples of $(L(t))_t$ for which $|L(t)| = O(t)$, $O(t^p)$ for each $p > 1$, and $O(2^{\alpha t})$ for $0 < \alpha < 1$.

3. Games Against a Fully Rational Player

We now derive a few consequences of bounded rationality expressed as a growth rate of $\psi_i(t) = |\Psi_i(t)|$, which may be interpreted as the number of strategies available to player i up to the t-th stage. We emphasize that the nature of the feasible strategy set Ψ_i is completely arbitrary. It may include infinitely many strategies and also the strategies that cannot be represented by any finite state machines or

[9]In this paper, we use the terms "nondecreasing" and "increasing", rather than "increasing" and "strictly increasing".

finitely bounded recall. As such, it is quite difficult, if not impossible, to obtain results on optimal strategies or equilibrium payoffs which require examination and construction of specific strategies. In what follows we study what may appear to be an extreme case: a two-person zero-sum repeated game where a player with bounded rationality plays against a fully rational player. Although the repeated game we study in this paper is rather special, 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 $\psi_1(t)$.

As we study two-person games in this section, we will follow the following notational rule. Actions of player 1 and 2 are denoted by a and b, respectively, and their strategies are denoted by σ and τ, respectively, with sub or superscripts and other affixes added as necessary.

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. 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 Ψ_1 is a finite set, then it is obvious that player 2, with entire strategy set Σ_2 available to her, can construct a strategy, say τ_2^*, which eventually identifies the strategy chosen by player 1 and gives him at most w at each stage thereafter.[10] Therefore $g_t(\sigma, \tau^*)$ converges to w for every $\sigma \in \Psi_1$. The first proposition provides its speed of convergence. It has appeared in Neyman and Okada (2000) in a study of nonzero-sum two person finitely repeated games with finite automata. In order to make this paper self-contained, and, since this proposition will be used in the proof of the second proposition, we will give the proof.

Theorem 1. *For every finite subset Ψ_1 of Σ_1 there exists $\tau^* \in \Sigma_2$ such that for all $\sigma \in \Psi_1$*

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

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

[10] Ben-Porath (1993)[Lemma 1, Theorem 1] provides an explicit construction of such strategy when Ψ_1 arises from complexity bound in terms of finite automata.

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 Ψ_1^h be the set of strategies in Ψ_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 Ψ_1^h that takes the action a given the history h, i.e.,

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

Let $a(h) \in A_1$ be such that $\left|\Psi_1^{h,a(h)}\right| \geq \left|\Psi_1^{h,a}\right|$ for all $a \in A_1$. Now define τ^* by

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

Clearly, $\left\{\Psi_1^{h,a} \mid a \in A_1\right\}$ is a partition of Ψ_1^h. If $a \neq a(h)$, then $\left|\Psi_1^{h,a}\right|$ is at most one half of $\left|\Psi_1^h\right|$. This implies that

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

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 (σ, τ^*). If we set

$$I_t = \begin{cases} 1 & \text{if } a_t \neq a(h_1, \dots, h_{t-1}) \\ 0 & \text{otherwise,} \end{cases}$$

then (2) implies that

$$\left|\Psi_1\right| 2^{-\sum_{t=1}^{T} I_t} \geq \left|\Psi_1^{(h_1,\dots,h_t)}\right| \geq 1.$$

Therefore

$$\sum_{t=1}^{T} I_t \leq \log_2 \left|\Psi_1\right|.$$

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

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

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

3.2. **Slowly Growing Strategy Set.** Next we consider an infinite strategy set Ψ_1. Recall that $\Psi_1(t)$ is formed by identifying strategies in Ψ_1 that are equivalent up to the t-th stage and $\psi_1(t) = \left|\Psi_1(t)\right|$. 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 Ψ_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* $\dfrac{\log \psi_1(t)}{t} \xrightarrow{t\to\infty} 0$. *Then there is a strategy* $\tau^* \in \Sigma_2$ *such that, for every* $\sigma \in \Psi_1$,

$$\limsup_{T\to\infty} g_T(\sigma, \tau^*) \leq w.$$

Proof: Let $\{t_k\}_{k=1}^{\infty}$ be a sequence satisfying the following properties

(A) $\dfrac{t_{k+1} - t_k}{t_k} \xrightarrow{k\to\infty} 0$, and

(B) $\dfrac{\log \psi_1(t_{k+1})}{t_{k+1} - t_k} \xrightarrow{k\to\infty} 0$.

It is easy to verify that such sequence exists under the condition of the theorem. Call a consecutive stages $t_k + 1, \ldots, t_{k+1}$ of the repeated game the k-th block.

The construction of player 2's strategy τ^* is similar to the one in the proof of Theorem 1. Given a history $h = (h_1, \ldots, h_{t-1})$, 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 $a(h) = \operatorname{argmax}_{a\in A_1} \left|\Psi_1^{h,a}(t_{k+1})\right|$ and define

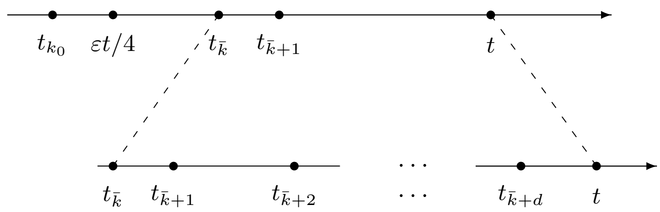
$$\tau^*(h) = \operatorname*{argmin}_{b \in A^2} g(a(h), b).$$

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

Fix $\varepsilon > 0$ and let k_0 be such that, for all $k \geq k_0$,

$$\frac{t_{k+1} - t_k}{t_k} < \frac{\varepsilon}{4}. \tag{3}$$

$$\frac{\log_2 \psi_1(t_{k+1})}{t_{k+1} - t_k} < \frac{\varepsilon}{4}. \tag{4}$$



Take $t > \dfrac{4t_{k_0}}{\varepsilon}$ and let $\bar{k}$ be the smallest index k for which $t_k > \dfrac{\varepsilon t}{4}$. Then, $\bar{k} > k_0$ and $t_{\bar{k}-1} < \dfrac{\varepsilon t}{4} < t_{\bar{k}}$. See the figure above.

Fix $\sigma \in \Psi_1$. Let $h = (h_1, h_2, \dots)$ be the play induced by (σ, τ^*). Note that $\sigma \in \Psi_1^{(h_1,\dots,h_{t-1})}(t_{k+1}) \subset \Psi_1(t_{k+1})$ whenever $t_k + 1 \le t \le t_{k+1}$. The average payoff to player 1 up to stage T is

$$g_T(\sigma, \tau^*) = \frac{1}{T}\left(\sum_{t=1}^{t_{\bar{k}}} g(h_t) + \sum_{t=t_{\bar{k}}+1}^{T} g(h_t)\right).$$

W.l.o.g, assume $\|g\| \le 1$. First, note that, by (3),

$$\frac{1}{T}\sum_{t=1}^{t_{\bar{k}}} g(h_t) \le \frac{t_{\bar{k}}}{T} = \frac{1}{T}\left(t_{\bar{k}} - \frac{\varepsilon T}{4} + \frac{\varepsilon T}{4}\right) \le \frac{t_{\bar{k}} - t_{\bar{k}-1}}{t_{\bar{k}-1}} + \frac{\varepsilon}{4} < \frac{\varepsilon}{2}.$$

Next suppose that there are d blocks between $t_{\bar{k}} + 1$ and T. Then,

$$\frac{1}{T}\sum_{t=t_{\bar{k}}+1}^{T} g(h_t) = \frac{1}{T}\sum_{j=1}^{d}\sum_{t=t_{\bar{k}+j-1}+1}^{t_{\bar{k}+j}} g(h_t) + \frac{1}{T}\sum_{t=t_{\bar{k}+d}+1}^{T} g(h_t).$$

The definition of τ^* and Theorem 1, together with (4) above, imply

$$\begin{aligned}\frac{1}{T}\sum_{t=t_{\bar{k}+j-1}+1}^{t_{\bar{k}+j}} g(h_t) &= \frac{t_{\bar{k}+j} - t_{\bar{k}+j-1}}{T} \cdot \frac{1}{t_{\bar{k}+j} - t_{\bar{k}+j-1}} \sum_{t=t_{\bar{k}+j-1}+1}^{t_{\bar{k}+j}} g(h_t) \\ &\le \frac{t_{\bar{k}+j} - t_{\bar{k}+j-1}}{T}\left(w + \frac{\varepsilon}{4}\right),\end{aligned}$$

and (3) implies that

$$\frac{1}{T}\sum_{t=t_{\bar{k}+d}+1}^{T} g(h_t) \le \frac{t_{\bar{k}+d+1} - t_{\bar{k}+d}}{t_{\bar{k}+d}} < \frac{\varepsilon}{4}.$$

Hence

$$\begin{aligned}\frac{1}{T}\sum_{t=t_{\bar{k}}+1}^{T} g(h_t) &< \frac{1}{T}\Big(w+\frac{\varepsilon}{4}\Big)\sum_{j=1}^{d}(t_{\bar{k}+j}-t_{\bar{k}+j-1})+\frac{\varepsilon}{4}\\ &= \Big(w+\frac{\varepsilon}{4}\Big)\frac{t_{\bar{k}+d}-t_{\bar{k}}}{T}+\frac{\varepsilon}{4}\\ &< w+\frac{\varepsilon}{2}.\end{aligned}$$

Therefore, $g_T(\sigma,\tau^*) < w+\varepsilon$. **Q.E.D.**

When Ψ_1 is a Cartesian product, as is essentially the case in Example 1, the construction of τ^* is easy. Let $\Psi_1 = \Psi_{11}\times\Psi_{12}\times\ldots$. So $\Psi_1(t) = \Psi_{11}\times\cdots\times\Psi_{1t}$. Hence $\psi_1(t) = \big|\Psi_{11}\big|\times\cdots\times\big|\Psi_{1t}\big|$. For each t, let $n(t) = \big|\{s\,|\,s\le t,\ \big|\Psi_{1s}\big|\ge 2\}\big|$. Then, $\psi_1(t)\ge 2^{n(t)}$. Thus $\dfrac{\log\psi_i(t)}{t}\xrightarrow{t\to\infty} 0$ implies that $\dfrac{n(t)}{t}\xrightarrow{t\to\infty} 0$. For each history $h=(h_1,\ldots,h_{t-1})$, define $\tau^*(h)$ by

$$\tau^*(h)=\begin{cases}\operatorname{argmin}_{b\in A_2} g(\sigma_{1t}(h),b) & \text{if } \Psi_{1t}=\{\sigma_{1t}\}\\ \textit{arbitrary action} & \text{if } \big|\Psi_{1s}\big|\ge 2\end{cases}$$

Then

$$g_t(\sigma,\tau^*)\le\frac{t-n(t)}{t}w+\frac{n(t)}{t}\|g\|\to w \ \text{ as }\ t\to\infty.$$

3.3. **Growing Strategy Sets and Entropy.** In this section we prove a generalization of Theorem 2 for the case when $\dfrac{\log\psi_1(t)}{t}$ converges to an arbitrary positive number. To do this we will use the concept of entropy and its properties which we will now introduce.

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

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

The entropy of a vector of random variables, $H(X_1,\cdots,X_n)$, is similarly defined.

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,

$$H(X|y) = -\sum_{x} p(x|y)\log p(x|y).$$

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

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

The following "chain rule" for entropy, which we will use in the proof of the next theorem, is easy to verify. See Cover and Thomas (1991).

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

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

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

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 Ψ_1 induced by an equivalence of pure strategies. That is, we define an equivalence relation $\underset{t}{\sim}$ by

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

Then $\Psi_1(t) = \Psi_1/\underset{t}{\sim}$.

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

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

and let $\Psi_1(t,\tau) = \Psi_1/\underset{t,\tau}{\sim}$. Clearly $\Psi_1(t,\tau)$ is a finite partition of Ψ_1 and $\Psi_1(t)$ is its refinement. Hence, by the property of the entropy of partitions mentioned above,

$$H_\sigma(\Psi_1(t,\tau)) \leq H_\sigma(\Psi_1(t)) \leq \log|\Psi_1(t)| = \log\psi(t). \tag{5}$$

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, $h(S) \in H_t$. More precisely, $h(S)$ is the history of length t which results when the strategy profile (s,τ) 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 $h = h(S)$. Clearly, this correspondence between $\Psi_1(t,\tau)$ and H_t is one-to-one. Furthermore, the event "a strategy $s \in S \subset \Psi_1(t,\tau)$ is selected by σ" is equivalent to the event "the history $h(S)$ occurs when (σ, τ) is played". Therefore,

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

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

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

Combining this equality with (5) we have

Lemma 2. *Let $\sigma \in \Delta(\Psi_1)$ and $\tau \in \Sigma_2$ and $(X_1, \ldots, X_t)$ be the random play up to stage t induced by (σ, τ). Then, for every t,*

$$H(X_1, \ldots, X_t) \leq \log \psi_1(t).$$

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

$$H(\alpha) = - \sum_{a \in A} \alpha(a) \log \alpha(a).$$

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

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

Thus $U(\gamma)$ is what player 1 can secure in the stage game G using a mixed action of entropy at most γ. Clearly, $U(0) = w$, the maximin value in pure actions. On the other hand, if $\gamma \geq \min_\alpha H(\alpha)$ where α is taken over player 1's optimal strategies in G, then $U(\gamma) = \mathrm{Val}(G)$, the value of G. Let cav U be the concavification of U, i.e., the smallest concave function which is at least as large as U at every point of its domain.

Theorem 3. *Suppose that $\limsup_{t \to \infty} \dfrac{\log \psi(t)}{t} = \gamma$. Then, for every $\sigma \in \Delta(\Psi_1)$, there is $\tau \in \Sigma_2$ such that*

$$\limsup_{T \to \infty} g_T(\sigma, \tau) \leq (\mathrm{cav}\, U)(\gamma).$$

Proof: Fix player 1's strategy $\sigma \in \Delta(\Psi_1)$. For the purpose of the payoff calculation, identify σ with its equivalent behavioral strategy. 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,

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

Let $X_1, X_2, \cdots$ be the sequence of random actions induced by (σ,τ). 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, cav U, and τ, we have

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

Taking the expectation, we have

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

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

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

The second inequality follows from Jensen's inequality. The second and the third equalities follow from the definition of conditional entropy and Lemma 1, respectively. The last inequality follows from Lemma 2. Since $\limsup_{t\to\infty}\dfrac{\log\psi_1(t)}{t} = \gamma$, we have the desired result. **Q.E.D.**

As in Theorem 2, whether player 1 can achieve $(\text{cav } U)(x)$ or not depends on what strategies are available to him.

Theorem 4. *For every $\gamma \geq 0$ and a function $f : \mathbb{N} \to \mathbb{N}$ with* $\dfrac{\log f(t)}{t} \xrightarrow{t\to\infty} \gamma$, *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) (a) $\psi_1(t) \leq f(t)$ *for every* $t \in \mathbb{N}$

(b) $\dfrac{\log \psi_1(t)}{t} \xrightarrow{t\to\infty} \gamma$

(ii) (a) $\underline{\lim}_{T\to\infty} \left(\inf_{\tau\in\Delta(\Sigma_2)} g_T(\hat{\sigma}, \tau) \right) \geq (\mathrm{cav}\ U)(\gamma)$

(b) $\inf_{\tau\in\Delta(\Sigma_2)} \mathbf{E}_{\hat{\sigma},\tau} \left[\underline{\lim}_{T\to\infty} \dfrac{1}{T} \sum_{t=1}^{T} g(a_t, b_t) \right] \geq (\mathrm{cav}\ U)(\gamma)$.

Proof: CONSTRUCTION OF Ψ_1: Recall from Example 2 that, for each sequence $\boldsymbol{a} = (a_1, a_2, \cdots)$ of player 1's pure actions, $\sigma\langle \boldsymbol{a} \rangle$ 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

$$\Psi_1 = \{\sigma\langle \boldsymbol{a} \rangle : \boldsymbol{a} \in F\}. \tag{6}$$

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

Suppose that $\gamma > 0$. By modifying $f(t)$ to $\hat{f}(t) = \inf_{s\geq t} f(s)^{t/s}$ if necessary, we will assume that $\dfrac{\log f(t)}{t}$ is nondecreasing in t and, in particular, $f(t)$ is nondecreasing in t.

In order to construct the set $F \subset A_1^\infty$ in this case, we first partition the stages into blocks. Set $t_0 = 0$. The n-th block ($n = 1, 2, \cdots$) 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 finally we set F to be those sequences $\boldsymbol{a} = (a_1, a_2, \cdots)$ in A_1^∞ whose n-th segment $\boldsymbol{a}[n] = (a_{t_{n-1}+1}, \cdots, a_{t_n})$ belongs to F_n:

$$F = \left\{ \boldsymbol{a} = (a_1, a_2, \cdots) \in A_1^\infty : \boldsymbol{a}[n] = (a_{t_{n-1}+1}, \cdots, a_{t_n}) \in F_n \right\}. \tag{7}$$

Now we describe the detail.

The blocks of stages are chosen so that[11] d_n is increasing, and $\frac{d_n}{t_n} \xrightarrow{n\to\infty} 0$. Next, we construct the sets $F_1, F_2, \cdots, F_n, \cdots$ by means of two sequences of nonnegative reals $\gamma_1, \gamma_2, \cdots, \gamma_n, \cdots$ and $\eta_1, \eta_2, \cdots, \eta_n, \cdots$. For each n, choose player 1's mixed action α_n so that $H(\alpha_n) \le \gamma_n$ and $\min_{b\in A_2} g(\alpha_n, b) \ge U(\gamma_n)$. Let F_n be the set of all sequences $(a_1, \ldots, a_{d_n}) \in A_1^{d_n}$ whose empirical distribution is within η_n of α_n. Formally,

$$F_n = \left\{ (a_1, \ldots, a_{d_n}) \in A_1^{d_n} : \sum_{a\in A_1} \left| \frac{1}{d_n} \sum_{k=1}^{d_n} \mathbf{1}(a_k = a) - \alpha_n(a) \right| \le \eta_n \right\}. \tag{8}$$

The sequence $(\eta_n)_n$ is chosen to satisfy, $\eta_n \xrightarrow{n\to\infty} 0$ and, in addition, the following property. Let $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \cdots)$ be a sequence of independent A_1-valued random variables where x_t is distributed according to α_n whenever t is in the n-th block, i.e. $t_{n-1} + 1 \le t \le t_n$. Then we require[12]

$$\sum_{n=1}^{\infty} \mathrm{P}\big(\mathbf{x}[n] = (\mathrm{x}_{t_{n-1}+1}, \cdots, \mathrm{x}_{t_n}) \notin F_n\big) < \infty. \tag{9}$$

The sequence $(\gamma_n)_n$ depends on the function f and will be specified later. We complete this part by defining F by (7) and then Ψ_1 by (6).

<u>Construction of $\hat\sigma \in \Delta(\Psi_1)$</u>: Fix a sequence $\bar{\boldsymbol{a}} = (\bar a_1, \bar a_2, \cdots) \in F$. First, generate a sequence of independent actions $\boldsymbol{a} = (a_1, a_2, \cdots)$ of player 1 where actions in the n-th block $\boldsymbol{a}[n] = (a_{t_{n-1}+1}, \cdots, a_{t_n})$ are drawn according to the identical mixed action α_n. In the n-th block, player 1 plays according to $\boldsymbol{a}[n]$ whenever $\boldsymbol{a}[n] \in F_n$ while he plays $\bar{\boldsymbol{a}}[n] = (\bar a_{t_{n-1}+1}, \cdots, \bar a_{t_n})$ otherwise.

To formally define the mixed strategy $\hat\sigma$, recall that $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \cdots)$ is a sequence of independent A_1-valued random variables where x_t is distributed according to α_n

---

[11] For example, $t_1 = 1, t_2 = 3, t_3 = 6, \cdots, t_n = \frac{n(n+1)}{2}, \cdots$. In this case $d_n = n$.

[12] For example, take $\eta_n = \min\left\{ d_n^{-3}, 1 - \max_{a_1\in A_1} \alpha_n(a) \right\}$. Indeed, with this choice of η_n, we have $\eta_n \xrightarrow{n\to\infty} 0$ and, in addition, an application of a large deviation inequality due to Hoeffding (1963),

$$\mathrm{P}\big(\mathbf{x}[n] = (\mathrm{x}_{t_{n-1}+1}, \cdots, \mathrm{x}_{t_n}) \notin F_n\big) \le 2|A_1| \exp\left(\frac{2 d_n \eta_n^2}{|A_1|^2} \right),$$

ensures that (9) holds.

whenever $t_{n-1}+1 \le t \le t_n$. Define a sequence of A_1-valued random variables $\hat{\mathbf{x}} = (\hat{\mathrm{x}}_1, \hat{\mathrm{x}}_2, \cdots)$ by

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

Then $\hat{\sigma} = \sigma\langle\hat{\mathbf{x}}\rangle$. Note that $\hat{\sigma}$ is indeed a mixture of strategies in Ψ_1.

<u>Verification of the theorem</u>: From (7) 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 $\bigtimes_{n=1}^{N} F_n$, $N = 1, 2, \cdots$. Hence $\psi_1(t_N) = \left|\Psi_1(t_N)\right| = \prod_{n=1}^{N} \left|F_n\right|$ for each $N = 1, 2, \cdots$. Since both $\psi_1(t)$ and $f(t)$ are nondecreasing, in order to verify that Ψ_1 has the property (i)-(a) of the theorem, it is enough to show that

$$\sum_{n=1}^{N} \log\left|F_n\right| \le \log f(t_N) \quad \text{for each } N = 1, 2, \cdots \tag{10}$$

We estimate $\left|F_n\right|$ as follows. By its definition (8), each element $\boldsymbol{a}_n = (a_1, \cdots, a_{d_n})$ in F_n has an empirical distribution $\rho(\boldsymbol{a}_n)$ whose L_1-distance from α_n is at most $2\eta_n$. As the entropy $H(\alpha)$ as a function on $\Delta(A_1)$ is uniformly continuous (in the L_1-distance), there is a nondecreasing function $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ with $\varepsilon(\eta) \xrightarrow{\eta \to 0} 0$ such that $\left|H\big(\rho(\boldsymbol{a}_n)\big) - H(\alpha_n)\right| < \varepsilon(\eta_n)$ for all $\boldsymbol{a}_n \in A_1^{d_n}$. Since the number of distinct empirical distributions arising from sequences in $A_1^{d_n}$ is at most $d_n^{|A_1|}$, we deduce that $\left|F_n\right| \le d_n^{|A_1|} 2^{d_n(H(\alpha_n)+\varepsilon(\eta_n))}$. As $H(\alpha_n) \le \gamma_n$, we have $\left|F_n\right| \le d_n^{|A_1|} 2^{d_n(\gamma_n+\varepsilon(\eta_n))}$. Furthermore, since $d_n \uparrow_{n\to\infty} \infty$ and $\eta_n \xrightarrow{n\to\infty} 0$, we can choose a sequence of nonincreasing nonnegative reals $(\varepsilon_n)_n$ such that[13] $\varepsilon_n \xrightarrow{n\to\infty} 0$ and $\left|F_n\right| \le 2^{d_n(\gamma_n+\varepsilon_n)}$. Thus, to show (10), it is enough to choose $(\gamma_n)_n$ so that

$$\sum_{n=1}^{N} d_n\big(\gamma_n + \varepsilon_n\big) \le \log f(t_N) \quad \text{for each } N = 1, 2, \cdots. \tag{11}$$

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

---

[13] For example, $\varepsilon_n = \sup_{\ell \ge n}\left(\varepsilon(\eta_\ell) + |A_1| \dfrac{\log d_\ell}{d_\ell}\right)$.

that is,

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

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

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

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

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

Thus, by virtue of (9), in order to show part (iii)-(a) of the theorem it suffices to choose $(\gamma_n)_n$ so that

$$\frac{1}{t_N} \sum_{n=1}^{N} d_n U(\gamma_n) \xrightarrow{N \to \infty} (\mathrm{cav}\ U)(\gamma). \tag{12}$$

We now exhibit a choice of $(\gamma_n)_n$ that satisfy (11) and (12). It will be seen that (i)-(b) is also satisfied with our choice of $(\gamma_n)_n$ below. We distinguish two cases.

Case 1 - $(\mathrm{cav}\ U)(\gamma) = U(\gamma)$: In this case, any sequence $(\gamma_n)_n$ with $\gamma_n \xrightarrow{n \to \infty} \gamma$ and $\gamma_n + \varepsilon_n \leq \dfrac{\log f(t_n)}{t_n}$ satisfies (11) and (12) as well as (i)-(b).

Case 2 - $(\mathrm{cav}\ U)(\gamma) > U(\gamma)$: In this case, the definitions of U and cav U imply the existence of γ_-, γ_+ 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) $(\mathrm{cav}\ U)(\gamma) = pU(\gamma_-) + (1-p)U(\gamma_+)$
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 $\dfrac{\log f(t_{\bar{n}})}{t_{\bar{n}}} > \dfrac{\gamma_- + \gamma}{2}$ and $\varepsilon_{\bar{n}} < \dfrac{\gamma - \gamma_-}{2}$. For $n < \bar{n}$, set $\gamma_n = 0$. Then set $\gamma_{\bar{n}} = \gamma_-$ and, for $n > \bar{n}$, define γ_n by induction as follows:

$$\gamma_n = \begin{cases} \gamma_+ & \text{if } \sum_{\ell=\bar{n}}^{n-1} d_\ell(\gamma_\ell + \varepsilon_\ell) + d_n(\gamma_+ + \varepsilon_n) \leq \log f(t_n), \\ \gamma_- & \text{otherwise.} \end{cases}$$

With the above choice of the sequence $(\gamma_n)_n$, we have $|F_1| = \cdots = |F_{\bar{n}-1}| = 1$. So for $N < \bar{n}$, the inequality (10) trivially holds. For $N \geq \bar{n}$, the inequality (11) can be easily verified by induction using the fact that $\dfrac{\log f(t_n)}{t_n}$ is nondecreasing and ε_n is nonincreasing.

We conclude by showing that

$$\frac{1}{t_N}\sum_{n=1}^{N} d_n\gamma_n \xrightarrow{N\to\infty} \gamma \tag{13}$$

which implies (12). Indeed, since $\gamma_k = \gamma_-$ or γ_+, this implies that

$$\frac{1}{t_N}\sum_{n=1}^{N} d_n \mathbf{1}(\gamma_n = \gamma_-) \xrightarrow{N\to\infty} p$$

from which (12) follows immediately.

Let $S_N = \sum_{n=1}^{N} d_\ell(\gamma_n + \varepsilon_n)$. Since $\frac{1}{t_N}\sum_{n=1}^{N} d_n\varepsilon_n \xrightarrow{N\to\infty} 0$, to show (**??**) it is enough to show that $\frac{S_N}{t_N} \xrightarrow{N\to\infty} \gamma$. Note that this also implies (i)-(b).

Then it is easy to see that for every $N \geq \bar{n}$ there is an $n > N$ such that $\frac{S_n}{t_n} \geq \frac{\log f(t_N)}{t_N}$. Suppose that there is an $N \geq \bar{n}$ such that $\frac{S_n}{t_n} < \frac{\log f(t_N)}{t_N}$ for infinitely many n's. Then choose an M large enough so that $\frac{\log f(t_N)}{t_N} < \frac{\log f(t_M)}{t_M}$ and $n > M$ large enough so that $\frac{S_n}{t_n} < \frac{\log f(t_N)}{t_N} < \frac{\log f(t_M)}{t_M} \leq \frac{S_{n-1}}{t_{n-1}}$, and

$$\frac{t_{n-1}}{t_n}\frac{\log f(t_M)}{t_M} + \frac{d_n}{t_n}\gamma_- \geq \frac{\log f(t_N)}{t_N}.$$

Such an n exists by the supposition and **(s-2)**. But then,

$$\frac{S_n}{t_n} = \frac{t_{n-1}}{t_n}\frac{S_{n-1}}{t_{n-1}} + \frac{d_n}{t_n}(\gamma_- + \varepsilon_n) \geq \frac{\log f(t_N)}{t_N},$$

a contradiction. We have thus shown that, for every N, $\frac{S_n}{t_n} \geq \frac{\log f(t_N)}{t_N}$ for all but finitely many n's and so Note that by **(s-7)**, we have and therefore

$$\frac{1}{t_n}\sum_{k=1}^{n} d_k\gamma_k \xrightarrow{n\to\infty} \gamma.$$

Q.E.D.

`End of Theorem 2 with revised proof.`

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 $\mathbf{B}_1(\kappa)$, the set of κ-recall strategies. Player 2 is assumed to have full recall. Recall that

$$\mathbf{B}_1(\kappa) = \{\sigma \wedge \kappa : \sigma \in \Sigma_1\}$$

where $(\sigma \wedge \kappa)(a_1, \cdots, a_t) = \sigma(a_{t-\kappa(t)+1}, \cdots, a_t)$.

As mentioned in Example 1, $\sigma \wedge \kappa$ acts like a fixed stationary bounded recall strategy of size k, indeed it is exactly $\sigma \bar{\wedge} k$, whenever $\kappa(t) = k$. The significance of requiring stationarity in this manner can be illustrated as follows. Suppose that for each t and t' with $\kappa(t) = \kappa(t') = k$, we allow player 1 to use a different stationary bounded recall strategies $\sigma \bar{\wedge} k$ and $\sigma' \bar{\wedge} k$. Then for any sequence $\boldsymbol{a} = (a_1, a_2, \cdots) \in A_1^\infty$, the oblivious strategy $\sigma\langle \boldsymbol{a} \rangle$ can be implemented simply by using a series of 0-recall strategies $\sigma^t \bar{\wedge} 0 \equiv a_t$. Thus a mixed strategy $\sigma\langle \mathbf{z} \rangle$, where $\mathbf{z} = (\mathrm{z}_1, \mathrm{z}_2, \cdots)$ is a sequence of random actions i.i.d. according to an optimal strategy of player 1 in G, trivially guarantees $\mathrm{Val}(G)$ in the long run.

Let $G^*(\kappa)$ be the repeated game under consideration and denote its value by $\mathrm{V}(\kappa)$. If we set $\Psi_1 = \mathbf{B}_1(\kappa)$, then from (1) we have

$$\log \psi_1(t) \leq c\, m^{\kappa(t)} \log m_1. \tag{14}$$

for some constant $c > 0$ (in fact, $c = m/(m-1)$). If $\kappa(t) < \dfrac{\log t}{\log m}$ for sufficiently large t, then (14) implies that $\dfrac{\log \psi_1(t)}{t} \xrightarrow{t \to \infty} 0$. 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* $\limsup_{t \to \infty} \dfrac{\kappa(t)}{\log t} < \dfrac{1}{\log m}$, *then* $\mathrm{V}(\kappa) = w$.

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 $\log t$ stages back. It is thus natural to ask, "if a player has growing recall $\kappa(t) > K_1 \log t$ for some $K_1 > \dfrac{1}{\log m}$, is it enough to guarantee $\mathrm{Val}(G)$ against a player with full recall?". This question will be answered affirmatively in the next theorem. In order to exhibit the constant K_1

explicitly, let $\zeta(G) = \max_{\alpha} \max_{a \in A_1} \alpha(a)$ where α is taken over all optimal strategies of player 1 in the stage game G. For example, $\zeta(G) = 1$ if there is a pure optimal strategy. Define

$$K_1(G) = \begin{cases} 0 & \text{if } \zeta(G) = 1, \\ \dfrac{3}{|\log \zeta(G)|} & \text{if } \zeta(G) < 1. \end{cases}$$

For instance, in matching pennies, $K_1 = 3$.

Theorem 6. *If* $\liminf_{t \to \infty} \dfrac{\kappa(t)}{\log t} > K_1(G)$*, then there is a* $\hat{\sigma} \in \Delta\big(\mathbf{B}_1(\kappa)\big)$ *with the following properties.*

(i) $\liminf_{T \to \infty} \Big(\min_{\tau \in \Sigma_2} g_T(\hat{\sigma}, \tau) \Big) \geq \mathrm{Val}(G)$,

(ii) $\inf_{\tau \in \Sigma_2} \mathbf{E}_{\hat{\sigma}, \tau} \left[\liminf_{T \to \infty} \dfrac{1}{T} \sum_{t=1}^{T} g(a_t, b_t) \right] \geq \mathrm{Val}(G)$.

Proof: By taking the lower envelope of κ if necessary, we assume without loss of generality that κ is nondecreasing. Let $t_1 = 1 < t_2 < \cdots < t_n < \cdots$ be stages at which player 1's recall grows (points of jump of κ). Set $k_n = \kappa(t_n)$ and call $B_n = \{t_n, t_n + 1, \cdots, t_{n+1} - 1\}$ the n-th block.

Let $\alpha^* \in \Delta(A_1)$ be an optimal strategy of player 1 in the stage game that achieves $\zeta(G)$, i.e., $\min_{b \in A_2} g(\alpha^*, b) = \mathrm{Val}(G)$ and $\max_{a \in A_1} \alpha^*(a) = \zeta(G)$. If α^* is a pure action, then the theorem is trivially true. So suppose that α^* 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. For each sequence of player 1's actions $\boldsymbol{a} = (a_1, a_2, \cdots) \in A_1^\infty$ and each positive integer n, define a sequence $\boldsymbol{a}^n = (a_1^n, a_2^n, \cdots) \in A_1^\infty$ 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 $\sigma\langle \boldsymbol{a}^n \rangle$, he would take the action a^0 in the first $n - 1$ blocks, then a^1 at the beginning of the n-th block, and thereafter the actions appearing in the original sequence $\boldsymbol{a}$.

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

$$\forall \ell \geq n, \forall r, s \in B_\ell : (a_{r-k_\ell}, \cdots, a_{r-1}) \neq (a_{s-k_\ell}, \cdots, a_{s-1}). \tag{15}$$

Then for any block B_ℓ after B_n, no two sequences of length k_ℓ ending in B_ℓ are identical. In this case it is easy to see that the κ-recall (oblivious) strategy $\sigma\langle \boldsymbol{a}^n\rangle \wedge \kappa$ induces the sequence of actions $\boldsymbol{a}^n$.

Our $\hat{\sigma}$ of the theorem will be a mixture over $\{\sigma\langle \boldsymbol{a}^n\rangle \wedge \kappa : \boldsymbol{a} \in A_1^\infty,\, n = 1, 2, \cdots\}$ where actions in the sequence $\boldsymbol{a} = (a_1, a_2, \cdots)$ are chosen independently according to α^* and then choosing an n with the property (15). Of course, we must make sure that such an $n < \infty$ exists for almost all realizations of the sequence. Formally, let $\mathbf{z} = (\mathrm{z}_1, \mathrm{z}_2, \cdots,)$ be a sequence of A_1-valued i.i.d. r.v.'s with $\mathrm{z}_t \sim \alpha^*$, and, for each n, define a sequence of A_1-valued r.v.'s $\mathbf{z}^n = (\mathrm{z}_1^n, \mathrm{z}_2^n, \cdots)$ by

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

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

$$\forall \ell \geq n, \forall r, s \in B_\ell : (\mathrm{z}_{r-k_\ell}, \cdots, \mathrm{z}_{r-1}) \neq (\mathrm{z}_{s-k_\ell}, \cdots, \mathrm{z}_{s-1}). \tag{16}$$

Then define $\hat{\sigma} = \sigma\langle \mathbf{z}^{\boldsymbol{\nu}}\rangle \wedge \kappa$. Below we will show that $\boldsymbol{\nu} < \infty$ 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 $\{\sigma\langle \boldsymbol{a}^n\rangle \wedge \kappa : \boldsymbol{a} \in A_1^\infty,\, n = 1, 2, \cdots\} \subset \mathbf{B}_1(\kappa)$. In fact, we will show that, if $\dfrac{\kappa(t)}{\log t} > K_1(G)$ for all sufficiently large t, then with probability one there are only finitely many n's with the property[14]that, for some $\ell \geq n$ and $r, s \in B_\ell$, $(\mathrm{z}_{r-k_\ell}^n, \cdots, \mathrm{z}_{r-1}^n) = (\mathrm{z}_{s-k_\ell}^n, \cdots, \mathrm{x}_{s-1}^n)$.

To see this fix an n and let $\ell, r, s \in \mathbb{N}$ be such that $\ell \geq n$ and $r, s \in B_\ell$. Then it is easy to verify that

$$\mathrm{Prob}\Big((\mathrm{z}_{r-k_\ell}^n, \cdots, \mathrm{z}_{r-1}^n) = (\mathrm{z}_{s-k_\ell}^n, \cdots, \mathrm{z}_{s-1}^n)\Big) \leq \zeta^{k_\ell}. \tag{17}$$

[14]Of course, this is true for any i.i.d. sequence $X_1, X_2, \cdots \sim X$ where X has a discrete support D and $\sup_{x \in D} \mathrm{Prob}(X = x) < 1$.

Denote the event on the left side of this inequality by $C_n(\ell, s, r)$ and set $C_n = \bigcup_{\ell \geq n} \bigcup_{\substack{s,r \in B_\ell \\ s<r}} C_n(\ell, s, r)$. Then, for every n,

$$\text{Prob}\big(C_n\big) \leq \sum_{\ell \geq n} \zeta^{k_\ell} d_\ell^2$$

where $d_\ell = t_{\ell+1} - t_\ell$ is the length of the block B_ℓ. Recall that $\kappa(t)$ is constant for $t \in B_\ell$. Therefore,

$$\sum_{\ell=1}^{\infty} \zeta^{k_\ell} d_\ell^2 \leq \sum_{t=1}^{\infty} \zeta^{\kappa(t)} t^2.$$

If $\dfrac{\kappa(t)}{\log t} > K_1(G) = \dfrac{3}{|\log \zeta(G)|}$ for sufficiently large t, then for some $\varepsilon > 0$ each summand on the right side is at most $t^{-(1+\varepsilon)}$ and so the series on the right side is convergent. Hence by the first Borel-Cantelli lemma

$$\text{Prob}\Big(\limsup_{n\to\infty} C_n\Big) = 0$$

which is what wanted to show.

Next we verify that $\hat{\sigma}$ has the desired properties. Fix an arbitrary strategy $\tau \in \Sigma_2$ and a stage T. Then

$$\begin{aligned}
&\mathbf{E}_{\hat{\sigma},\tau}\left[\frac{1}{T}\sum_{t=1}^{T} g(a_t, b_t)\right] \\
&\qquad \geq P(t_\nu > T)\big(-\|g\|\big) \\
&\qquad\qquad + P\big(t_\nu \leq T\big)\mathbf{E}_{\hat{\sigma},\tau}\left[\left(1 - \frac{t_\nu}{T}\right)\text{Val}(G) - \frac{t_\nu}{T}\|g\| \middle| t_\nu \leq T\right] \\
&\qquad = P(t_\nu > T)\big(-\|g\|\big) \\
&\qquad\qquad + P(t_\nu \leq T)\left(\text{Val}(G) - \frac{\text{Val}(G) + \|g\|}{T}\mathbf{E}_{\hat{\sigma},\tau}\big[t_\nu \big| t_\nu \leq T\big]\right).
\end{aligned} \tag{18}$$

Since $P(t_\nu \geq t) \xrightarrow{t\to\infty} 0$,

$$\frac{1}{T}\mathbf{E}_{\hat{\sigma},\tau}\big[t_\nu \big| t_\nu \leq T\big] \leq \frac{1}{T}\left(\sum_{t=1}^{T} P(t \leq t_\nu \leq T) + 1\right) \xrightarrow{t\to\infty} 0.$$

Hence the right side of (18) converges to $\text{Val}(G)$. This proves part (i). **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 Instituit, Mannheim.

——— (1997): "Rationality and Bounded Rationality," *Games and Economic Behavior*, 21(1-2), 2–14.

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.

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-Seies F*, pp. 233–255. Springer-Verlag, New York.

Neyman, A., and D. Okada (2000): "Two-Person Repeated Games with Finite Automata," *International Journal of Game Theory*, 29(3), 309 – 325.

Stearns, R. E. (1997): "Memory-bounded Game-playing Computing Devices," Discussion Paper mimeo., Department of Computer Science, SUNY at Albany.