Available online at www.sciencedirect.com
ScienceDirect
Games and Economic Behavior 66 (2009) 404–425
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Growth of strategy sets, entropy, and nonstationary bounded recall

Abraham Neyman [a,1], Daijiro Okada [b,c,*]

[a] *Center for the Study of Rationality and Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel*
[b] *Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540-0631, USA*
[c] *Department of Economics, Rutgers—The State University of New Jersey, 75 Hamilton Street, New Brunswick, NJ 08901-1248, USA*

Received 15 September 2006

Available online 20 April 2008

Abstract

The paper initiates the study of long term interactions where players' bounded rationality varies over time. Time dependent bounded rationality, for player i, is reflected in part in the number $\psi_i(t)$ of distinct strategies available to him in the first t-stages.

We examine how the growth rate of $\psi_i(t)$ affects equilibrium outcomes of repeated games. An upper bound on the individually rational payoff is derived for a class of two-player repeated games, and the derived bound is shown to be tight.

As a special case we study the repeated games with nonstationary bounded recall and show that, a player can guarantee the minimax payoff of the stage game, even against a player with full recall, by remembering a vanishing fraction of the past. A version of the folk theorem is provided for this class of games.

JEL classification: C72; C73

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 their feasible set of strategies. Aside from the assumption that the players can perform computations necessary for such tasks, it is assumed that players can carry out any strategy in the specified strategy set should they choose to play it. While this latter assumption may seem innocuous in a model where few strategies are available to each player,[2] e.g., prisoner's dilemma and the battle of the sexes, it may be criticized as being unrealistically rational in more complex models where the theoretical definition of strategy leads to a strategy set that contains a large number of choices, many of which are impractically complex.

* Corresponding author at: Institute for Advanced Study, Einstein Drive, Princeton, NJ, USA.
E-mail addresses: aneyman@math.huji.ac.il (A. Neyman), okada@econ.rutgers.edu (D. Okada).
1 Neyman's research is partly supported by Israel Science Foundation Grants 382/98 and 263/03.
2 See however Anderlini (1990).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2008.04.005

A case in point is models of dynamic interaction including repeated games in their most basic formulation. In repeated games, a strategy is a set of history-contingent plans of action. 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 require the capability to process arbitrarily complex history for their implementation.

The idea that the assumption of fully, or unboundedly, rational players is unrealistic is not new (Simon, 1955, 1972; Aumann, 1981, 1997). There have been many attempts to model feasible (implementable) sets of strategies that reflect some aspects of the 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.[3]

Existing literature on bounded complexity in repeated games considers models where the complexity of strategies is fixed during the course of a long interaction. In the case of finite automata and bounded recall (e.g., Neyman, 1985; 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, 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. Such changes bring about the transformation of the set of possible decision rules over time.

As argued in the beginning, complexity of repeated games as a model of interactive decision-making stems, in part, from the wealth of strategies from which the theory allows players to choose. The number of theoretically possible strategies is double-exponential in the number of repetitions. Some, in fact most,[4] 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 on various measures of complexity of implementing strategies and their effects on equilibrium outcomes in particular.

Our aim in this paper is to take a first step toward 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 resources available to the players may expand or contract over time. As a consequence, the limitation would vary over time and, in particular, there may not be a finite upper bound on complexity of strategies for the entire horizon of the game. Such considerations of the more general aspects of bounded rationality cannot be captured by a model with a finite set of feasible strategies. Thus we are led to considering a feasible set consisting of infinitely many strategies. The question that arises then is : "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 fewer elements than the fully rational case. As we take aim at a temporal aspect of an infinite strategy set, we shall consider how the number of strategies induced in the first t stages of the game grows. Specifically, we associate to each subset Ψ_i of the full (theoretically possible) strategy set a function ψ_i from the set of positive integers to itself. The value $\psi_i(t)$ represents the number of strategies in Ψ_i that are distinguishable in the first t stages. The feasible strategy set Ψ_i

[3] Variants of complexity measure associated with a Turing machine include the number of bits needed to implement a strategy by Turing machines with a bounded amount of tape (Stearns, 1997), and algorithmic or Kolmogorov complexity (Lacôte, 2005; Neyman, 2003).

[4] 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.

may contain infinitely many strategies, but it can differ from the fully rational case in the way ψ_i grows reflecting a broad implication of bounded rationality that may vary over time.[5] To be more precise, for each t, let $\Psi_i(t)$ be the projection of Ψ_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 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 simple case of two-person repeated games in which player 1 with a feasible set Ψ_1 plays against a fully rational player 2. The payoff in the repeated games is the long-run average of the stage payoffs. In this setup we will show that there is a continuous nondecreasing function $U : \mathbb{R}_+ \to \mathbb{R}$ such that player 1 cannot guarantee more than $(\operatorname{cav} U)(\gamma)$, where $\operatorname{cav} U$ denotes the concavification of U, whenever $\psi_1(t)$ grows at most as fast as $2^{\gamma t}$. Moreover, this bound is tight. The function U will be defined using the concept of entropy and it will be shown that $U(0)$ is the maximin value of the stage game in pure actions and that for sufficiently large γ, $U(\gamma)$ is the usual maximin value of the stage game in mixed actions.

As a concrete case of an infinite feasible strategy set arising from a complexity consideration, we will study the repeated game with nonstationary bounded recall strategies, which is a model of a player whose depth of 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 guarantee no more than 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 if, for all sufficiently large t, the recall at stage t is at least $K_1 \log t$, the minimax payoff of the stage game can be guaranteed. Hence, in order to secure the minimax payoff of the stage game against a player with full recall, one needs to remember a long enough history. However, the length of that history is only a negligible fraction of the entire history.

In order to avoid possible confusion, we point out that, as is standard in the literature, we consider mixed strategies so long as their 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.

In Section 2 we will set the notation used throughout the paper and formalize the idea of the growth of strategy sets. 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[6] 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 ϵ). Thus the set of all histories is $A^* = \bigcup_{t=0}^{\infty} A^t$ where $A = \times_{i \in I} A_i$ and $A^0 = \{\epsilon\}$. 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 player i, σ_i and σ_i', are equivalent up to the tth stage if, for every profile of other players' strategies σ_{-i}, the sequence of action profiles induced by (σ_i, σ_{-i}) and (σ_i', σ_{-i}) are identical up to,

[5] In this paper, the feasible set Ψ_i, and hence the growth of the function ψ_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 ψ_i. This certainly deserves further research. The work reported here provides limits to what can and cannot be achieved by such a choice.

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

and including, stage t. If two strategies are equivalent up to the tth 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[7] $m_i^{m^0} \times \cdots \times m_i^{m^{t-1}} = m_i^{\frac{m^t-1}{m-1}}$. 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 tth stage. 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. Let $\psi_i(t)$ be the number 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 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

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

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

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

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 $\kappa : \mathbb{N} \to \mathbb{N} \cup \{0\}$ with $\kappa(t) \leqslant t-1$. For each $t \in \mathbb{N}$, the value $\kappa(t)$ represents the length of recall at stage t. 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

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

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

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

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

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

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

---

[7] The number of equivalence classes of strategies (reduced strategies) available to player i in the first t stages is $m_i^{(m_{-i}^t - 1)/(m_{-i}-1)}$ where $m_{-i} = \prod_{j \neq i} m_j$.

[8] In fact, this holds for all $t \geqslant k^2$ and $2^{ck \log k} \leqslant |\Psi_i| \leqslant 2^{dk \log k}$ where k is the bound on the number of states of automata and c and d are positive constants.

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, $\sigma_i : A^* \to A_i$ is oblivious if $\sigma_i((a_{i1}, a_{-i1}), \ldots, (a_{it}, a_{-it}))$ is independent of $a_{-i1}, \ldots, a_{-it}$. 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^∞. 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 $\boldsymbol{a} = (a_{i1}, a_{i2}, \ldots) \in A_i^\infty$, 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, $A_1 = A_2 = \{0, 1\}$. The strategies described in these examples are kinds of "trigger strategy" where a certain action is triggered by specific history or set of histories.

Example 3. For each integer $k \geqslant 0$, 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.

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

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

Example 4. A prefix of a history $h = (h_1, \ldots, h_t)$ is any of its initial segment $h' = (h_1, \ldots, h_s)$, $s \leqslant t$. 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 $L(t) = L \cap (H_1 \cup \cdots \cup H_{t-1})$; $L(t)$ is prefix-free and $L(t) \subset L(t+1)$. Define a strategy σ_1^L as follows.

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

This is a generalization of the trigger strategy: σ_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 Ψ_1 to be the set of player 1's strategies σ_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 $\mathcal{L}$, σ_1^L and σ_1^M are equivalent up to the tth stage whenever $L(t) = M(t)$. Then we have[9] $\psi_1(t) \leqslant |\mathcal{M}(t)|$ where $\mathcal{M}(t) = \{M(t)\colon\ M \in \mathcal{M}\}$. Examples of $\mathcal{M}$ can be constructed so that the corresponding function $\psi_1(t)$ is, e.g., $O(t^p)$ for a given $p \geqslant 1$, 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 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 finitely bounded recall.

Various forms of the folk theorem assert that any feasible payoff vector that gives each player at least his individually rational (I.R.) payoff can be an equilibrium outcome of the repeated game. Thus two repeated games with the same set of feasible payoffs may differ in the set of equilibrium payoff as a result of the difference in the I.R. payoffs. In the repeated game with perfect monitoring played by fully rational players, e.g., Aumann and Shapley (1994), the I.R. payoff of the repeated game coincides with that of the stage game. This is because, for every strategy profile of the other players, a player has a strategy that yields him at least his I.R. payoff of the stage game in the long run. In particular, the minimax theorem implies that, in a two-person game, each player has a repeated game strategy that yields at least the stage game I.R. payoff in the long run regardless of the other player's strategy. However, when the set of feasible strategies of a player differs from the fully rational case, the I.R. payoff of the repeated game may be different from that of the stage game, and, accordingly, the set of equilibrium payoffs may differ from that of the standard folk theorem.

[9] 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 ψ_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 σ and τ, 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 player 1's average payoff in the first T stages.

3.1. Slowly growing strategy set

Recall that $\Psi_1(t)$ is formed by identifying strategies in Ψ_1 that are equivalent up to the tth stage and $\psi_1(t) = |\Psi_1(t)|$. Our first theorem states that if the growth rate of $\psi_1(t)$ is subexponential in t, then player 1 cannot guarantee more than the maximin payoff in pure actions, w, in the long run. We first present a lemma[10] which provides a bound on player 1's minimax payoff in the repeated game for an arbitrary feasible set Ψ_1. Set $\|g\| = 2\max\{|g(a,b)|\colon a \in A_1, b \in A_2\}$.

Lemma 1. *For every $\Psi_1 \subset \Sigma_1$ and every nondecreasing*[11] *sequence of positive integers $\{t_k\}_{k=0}^{\infty}$ with $t_0 = 0$, there exists $\tau^* \in \Sigma_2$ such that*

$$g_{t_k}(\sigma,\tau^*) \leqslant w + \|g\| \frac{1}{t_k} \sum_{\ell=1}^{k} \log_2 \psi_1(t_\ell)$$

for all $\sigma \in \Psi_1$ and $k = 1, 2, \ldots$.

Proof. We construct the strategy $\tau^* \in \Sigma_2$ as follows. Fix a stage t and let ℓ be the unique index with $t_{\ell-1} < t \leqslant t_\ell$. If a history $h = (h_1,\ldots,h_{t-1}) = ((a_1,b_1),\ldots,(a_{t-1},b_{t-1}))$ is observed, let $\Psi_1(t_\ell,h)$ be the set of player 1's strategies in $\Psi_1(t_\ell)$ that are compatible with h, i.e., $\sigma \in \Psi_1(t_\ell,h)$ if, and only if, $\sigma \in \Psi(t_\ell)$, $\sigma(\epsilon) = a_1$, and $\sigma(h_1,\ldots,h_{s-1}) = a_s$ for all $s = 2,\ldots,t-1$. For each $a \in A_1$, let $\Psi_1(t_\ell,h,a)$ be the set of strategies in $\Psi_1(t_\ell,h)$ that takes the action a after the history h, i.e., $\Psi_1(t_\ell,h,a) = \{\sigma \in \Psi_1(t_\ell,h)\colon \sigma(h) = a\}$. Choose $a(h) \in A_1$ such that $|\Psi_1(t_\ell,h,a(h))| \geqslant |\Psi_1(t_\ell,h,a)|$ for all $a \in A_1$. The action $a(h)$ may be considered the most likely action taken by player 1 after the history h. Now define τ^* by

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

[10] Lemma 1 and Theorem 1 are slight generalizations of Theorem 3.1 (and the remarks following it) in (Neyman and Okada, 2000b).

[11] 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").

Clearly, $\{\Psi_1(t_\ell, h, a) \mid a \in A_1\}$ forms a partition of $\Psi_1(t_\ell)$. From the definition of $a(h)$ it follows that $|\Psi_1(t_\ell, h, a)| \leqslant \frac{1}{2}|\Psi_1(t_\ell, h)|$ for all $a \neq a(h)$. Thus, if $h' = hh_t = (h_1, \ldots, h_{t-1}, h_t)$ and $h_t = (a_t, b_t)$ with $a_t \neq a(h)$, then

$$\left|\Psi_1(t_\ell, h')\right| \leqslant \frac{1}{2}\left|\Psi_1(t_\ell, h)\right|. \tag{1}$$

Fix $\sigma \in \Psi_1$ and let $(h_1, h_2, \ldots) = ((a_1, b_1), (a_2, b_2), \ldots)$ be the play generated by (σ, τ^*). For each t, let $\mathrm{I}_t = 0$ or 1 according to $a_t = a(h_1, \ldots, h_{t-1})$ or $a_t \neq a(h_1, \ldots, h_{t-1})$. Then (1) implies that

$$\sum_{t=t_{\ell-1}+1}^{t_\ell} \mathrm{I}_t \leqslant \log_2\left|\Psi_1\big(t_\ell, (h_1, \ldots, h_{t_{\ell-1}})\big)\right| \leqslant \log_2 \psi_1(t_\ell) \quad \text{for all } \ell = 1, 2, \ldots.$$

That is, the number of stages t with $t_{\ell-1} + 1 \leqslant t \leqslant t_\ell$ at which player 1's action differs from $a(h_1, \ldots, h_{t-1})$ is at most $\log_2 \psi_1(t_\ell)$. Hence

$$\sum_{t=t_{\ell-1}+1}^{t_\ell} g(h_t) \leqslant \sum_{t=t_{\ell-1}+1}^{t_\ell} \left((1 - \mathrm{I}_t)w + \mathrm{I}_t \frac{\|g\|}{2}\right) \leqslant (t_\ell - t_{\ell-1})w + \|g\| \log_2 \psi_1(t_\ell).$$

Summing over $\ell = 1, \ldots, k$ (where $t_0 = 0$) we have

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

If Ψ_1 is a finite set, e.g., the set of finite automata of a bounded size, then there is a $\hat{t}$ such that $\Psi_1(t) = \Psi_1$ for all $t \geqslant \hat{t}$. In this case, a straightforward modification of the proof of Lemma 1 (in fact, the same proof but with $t_1 = \infty$) shows that[12] there exists $\tau^* \in \Sigma_2$ such that

$$g_T(\sigma, \tau^*) \leqslant w + \|g\| \frac{\log_2 |\Psi_1|}{T}$$

for all $\sigma \in \Psi_1$ and $T = 1, 2, \ldots$.

Theorem 1. *Suppose that* $\frac{\log_2 \psi_1(t)}{t} \underset{t\to\infty}{\longrightarrow} 0$. *Then there is a strategy* $\tau^* \in \Sigma_2$ *such that*

$$\overline{\lim_{T\to\infty}} \max_{\sigma\in\Psi_1} g_T(\sigma, \tau^*) \leqslant w.$$

Proof. Let $\{t_k\}_{k=1}^\infty$ be an increasing sequence of positive integers satisfying the following properties:

$$\text{(A)}\ \frac{t_{k+1} - t_k}{t_k} \underset{k\to\infty}{\longrightarrow} 0, \quad \text{and} \quad \text{(B)}\ \frac{\log_2 \psi_1(t_{k+1})}{t_{k+1} - t_k} \underset{k\to\infty}{\longrightarrow} 0.$$

It is easy to verify that such a sequence exists under the condition of the theorem.

Lemma 1 and (B) imply that there is a $\tau^* \in \Sigma_2$ such that, for every $\varepsilon > 0$, $g_{t_k}(\sigma, \tau^*) \leqslant w + \varepsilon/2$ for all $\sigma \in \Psi_1$ and all sufficiently large k. Hence, (A) implies that $g_T(\sigma, \tau^*) < w + \varepsilon$ for all $\sigma \in \Psi_1$ and all sufficiently large T. $\square$

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 a strategy.

3.2. Growth of strategy sets and entropy

In this section we prove a generalization of Theorem 1 for the case when $\frac{\log_2 \psi_1(t)}{t}$ converges to an arbitrary positive number. To do this we use the concept of entropy and its properties which we now recall.[13]

---

[12] This first appeared in Neyman and Okada (2000b) in a study of repeated games with finite automata.

[13] For more details 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 Ω 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 the distribution p is uniformly continuous (in L_1-norm), concave, and $0 \leqslant H(X) \leqslant \log_2|\Omega|$ where the lower bound 0 is achieved by any one of the degenerate distributions, $p(x) = 1$ for some $x \in \Omega$, and the upper bound is achieved by the uniform distribution, $p(x) = 1/|\Omega|$ 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:

$$H(X|Y) = E_Y\bigl[H(X|y)\bigr] = \sum_{y} p(y)H(X|y).$$

Conditioning reduces entropy, i.e., $H(X) \geqslant H(X|Y) \geqslant H(X|Y,Z)$, 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 Ω into sets in $\mathcal{F}$. Then the entropy of the partition $\mathcal{P}$ with respect to μ is defined by

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

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

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 relation. Specifically, we define an equivalence relation $\underset{t}{\sim}$ by

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

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

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

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

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 a refinement of $\Psi_1(t,\tau)$. Hence, by the property of the entropy of partitions mentioned above,

$$H_\sigma\bigl(\Psi_1(t,\tau)\bigr) \leqslant H_\sigma\bigl(\Psi_1(t)\bigr) \leqslant \log_2\bigl|\Psi_1(t)\bigr| = \log_2 \psi_1(t). \tag{2}$$

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,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 $h = h(S)$. 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 σ" is equivalent to the event "the history $h(S)$ occurs when (σ,τ) is played." Therefore,

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

Let us write $X_1,\dots,X_t$ 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

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

Combining this equality with (2) we have

Lemma 2. *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 (σ,τ). Then, for every t,*

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

Next, for each mixed action α 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 $U:\mathbb{R}_+\to\mathbb{R}_+$ by

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

Thus $U(\gamma)$ is what player 1 can secure in the stage game G using a mixed action of entropy at most γ. Clearly, $U(0)=w$, the maximin payoff in pure actions. On the other hand, $U(\gamma)=v$, the minimax payoff, if $\gamma\geqslant\bar\gamma$ where $\bar\gamma=\min\{H(\alpha)\colon\ \alpha\in\Delta(A_1), \min_{b\in A_2} g(\alpha,b)=v\}$. Let $\operatorname{cav}U$ 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 $0\leqslant\gamma\leqslant\bar\gamma$, and then constant, v, for $\gamma\geqslant\bar\gamma$. Thus, for every $\gamma\leqslant\bar\gamma$, 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 $\gamma\leqslant\bar\gamma$. 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 $(\operatorname{cav}U)(\gamma)$. The proof, though standard (see Theorem 5.1 in Neyman and Okada, 2000a, and also Proposition 14 in Gossner and Vieille, 2002), is provided here for completeness.

Theorem 2. *Suppose that $\overline{\lim}_{t\to\infty}\frac{\log_2\psi_1(t)}{t}\leqslant\gamma$. Then, for every $\sigma\in\Delta(\Psi_1)$, there is $\tau\in\Sigma_2$ such that*

$$\overline{\lim_{T\to\infty}}\, g_T(\sigma,\tau)\leqslant(\operatorname{cav}U)(\gamma).$$

Proof. Fix player 1's strategy $\sigma\in\Delta(\Psi_1)$. Let τ be player 2's strategy such that $E_{\sigma,\tau}[g(a)\big|h]=\min_{b\in B}E_{\sigma(h)}[g(a,b)]$ for any history h. Let $X_1,X_2,\dots$ 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 $h\in H_{t-1}$ is realized. Then, by the definitions of U, $\operatorname{cav}U$, and τ, we have $E_{\sigma,\tau}[g(X_t)|h]\leqslant U(H(X_t|h))\leqslant(\operatorname{cav}U)(H(X_t|h))$. Taking the expectation and using Jensen's inequality, we have $E_{\sigma,\tau}[g(X_t)]\leqslant(\operatorname{cav}U)(E_{\sigma,\tau}[H(X_t|h)])=(\operatorname{cav}U)(H(X_t|X_1,\dots,X_{t-1}))$. Summing over $t=1,\dots,T$ and using Jensen's inequality again, we have

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}E_{\sigma,\tau}\big[g(X_t)\big] &\leqslant (\operatorname{cav}U)\left(\frac{1}{T}\sum_{t=1}^{T}H(X_t|X_1,\dots,X_{t-1})\right)\\
&= (\operatorname{cav}U)\left(\frac{1}{T}H(X_1,\dots,X_T)\right)\quad\text{by the chain rule}\\
&\leqslant (\operatorname{cav}U)\left(\frac{\log_2\psi_1(T)}{T}\right)\quad\text{by Lemma 2.}
\end{aligned}
$$

Since $\overline{\lim}_{t\to\infty}\frac{\log_2\psi_1(t)}{t}\leqslant\gamma$, we have the desired result. □

As in Theorem 1, whether player 1 can achieve $(\operatorname{cav} U)(\gamma)$ or not depends on what strategies are available to him. The next two theorems state that there is indeed a strategy set with an appropriate growth rate with which $(\operatorname{cav} U)(\gamma)$ can be achieved. Furthermore, it states that it suffices to consider oblivious strategies. Combined with Theorem 2, it implies that there is a strategy set Ψ_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 $(\operatorname{cav} U)(\gamma)$. We present first the result for finitely repeated games as its proof may aid the reader in grasping the main idea while avoiding a few complications arising in infinitely repeated games.

Theorem 3. *For every $\gamma > 0$ and $\varepsilon > 0$, there is a positive integer T^* with the following properties. For every $T \geqslant T^*$, there is a set of oblivious strategies $\Psi_1(T)$ such that $|\Psi_1(T)| \leqslant 2^{\gamma T}$ and*

$$\min_{\tau \in \Sigma_2} \max_{\sigma \in \Delta(\Psi_1(T))} g_T(\sigma, \tau) \geqslant (\operatorname{cav} U)(\gamma) - \varepsilon.$$

Proof. Recall from Example 2 that, for each sequence $\boldsymbol{a} = (a_1, a_2, \ldots)$ 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.

If $\gamma = 0$, then $(\operatorname{cav} U)(\gamma)$ is the maximin payoff in pure actions, w. In this case the set $\Psi_1(T)$ can be taken as a singleton $\sigma\langle \boldsymbol{a} \rangle$ where $\{\boldsymbol{a} = (a, a, a, \ldots)\}$ with a being any one of player 1's pure actions that guarantees him w.

If $\gamma > 0$, choose $\theta > 0$ sufficiently small so that $(\operatorname{cav} U)(\gamma - \theta) > U(\gamma - \theta)$. Then there are γ_-, γ_+ and a $0 < p < 1$ such that $\gamma_- < \gamma - \theta < \gamma_+$, $(\operatorname{cav} U)(\gamma_\pm) = U(\gamma_\pm)$, $\gamma - \theta = p\gamma_- + (1-p)\gamma_+$ and $(\operatorname{cav} U)(\gamma - \theta) = pU(\gamma_-) + (1-p)U(\gamma_+)$. Let $\alpha_-, \alpha_+ \in \Delta(A_1)$ be such that $H(\alpha_\pm) = \gamma_\pm$ and $\min_{b \in A_2} g(\alpha_\pm, b) = U(\gamma_\pm)$.

Given a sufficiently large positive integer T, define

$$F_- = \left\{ (a_1, \ldots, a_{pT}) \,\Big|\, \sum_{a \in A_1} \left| \frac{1}{pT} \sum_{t=1}^{pT} \mathbf{1}(a_t = a) - \alpha_-(a) \right| \leqslant \frac{|A_1|}{pT} \right\},$$

$$F_+ = \left\{ (a_1, \ldots, a_{(1-p)T}) \,\Big|\, \sum_{a \in A_1} \left| \frac{1}{(1-p)T} \sum_{t=1}^{(1-p)T} \mathbf{1}(a_t = a) - \alpha_+(a) \right| \leqslant \frac{|A_1|}{(1-p)T} \right\},$$

and $F = F_- \times F_+$. (Assume for simplicity and without loss of generality that pT and $(1-p)T$ are integers.) Note that F_- and F_+ are nonempty. Let $\Psi_1 = \{\sigma\langle \boldsymbol{a} \rangle \colon \boldsymbol{a} \in F\}$.

Let $\mathbf{z} = (z_1, z_2, \ldots, z_T)$ be a sequence of A_1-valued random variables such that the first pT elements $\mathbf{z}_- = (z_1, \ldots, z_{pT})$ is drawn uniformly from F_-, and the next $(1-p)T$ elements $\mathbf{z}_+ = (z_{pT+1}, \ldots, z_T)$ are drawn uniformly from F_+ and independently from $\mathbf{z}_-$. Then define $\hat{\sigma} = \sigma\langle \mathbf{z} \rangle$. Observe that $\hat{\sigma}$ is indeed a mixture of strategies in Ψ_1.

Note that $|\Psi_1(T)| = |F_-| \times |F_+|$. We estimate the size of F_- and F_+ as follows. Since the entropy $H(\alpha)$ as a function on $\Delta(A_1)$ is uniformly continuous (in the L_1-distance), there is an $\varepsilon_- > 0$ such that $|H(\rho(\boldsymbol{a}_-)) - H(\alpha_-)| = |H(\rho(\boldsymbol{a}_-)) - \gamma_-| < \varepsilon_-$ for all $\boldsymbol{a}_- \in F_-$ where $\rho(\boldsymbol{a})$ denotes the empirical distribution of the sequence $\boldsymbol{a}$. Observe that ε_- can be made arbitrarily small by choosing T large enough. Since the number of distinct empirical distributions arising from sequences in A_1^{pT} is at most $(pT+1)^{|A_1|}$, and the number of sequences in A_1^{pT} with an empirical distribution α_- is at most $2^{pTH(\alpha_-)} = 2^{pT\gamma_-}$, we deduce that

$$\frac{1}{(pT+1)^{|A_1|}} 2^{pT(\gamma_- - \varepsilon_-)} \leqslant |F_-| \leqslant (pT+1)^{|A_1|} 2^{pT(\gamma_- + \varepsilon_-)}$$

or

$$2^{pT(\gamma_- - \delta_-)} \leqslant |F_-| \leqslant 2^{pT(\gamma_- + \delta_-)}$$

where $\delta_- = \varepsilon_- + |A_1| \log_2(pT+1)/pT$. Note that δ_- can be made arbitrarily small by choosing T large enough. Similarly, there is an $\delta_+ > 0$ such that

$$2^{(1-p)T(\gamma_+ - \delta_+)} \leqslant |F_+| \leqslant 2^{(1-p)T(\gamma_+ + \delta_+)}.$$

From these estimates it is clear that $|F| < 2^{\gamma T}$ for sufficiently large T.

In order to show the second part of the theorem, fix an arbitrary strategy $\tau \in \Sigma_2$. Let $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_{pT}, \mathrm{y}_{pT})$ be the random action pairs induced by $(\hat{\sigma}, \tau)$ in the first pT stages and let $(\bar{\mathrm{x}}, \bar{\mathrm{y}})$ be a $(A_1 \times A_2)$-valued random variable whose distribution is given by

$$\mathrm{P}\big((\bar{\mathrm{x}}, \bar{\mathrm{y}}) = (a, b)\big) := \frac{1}{pT} \sum_{t=1}^{pT} \mathrm{P}\big((\mathrm{x}_t, \mathrm{y}_t) = (a, b)\big).$$

Then, from the definitions of F_- and $\hat{\sigma}$, it easily follows that

$$\sum_{a \in A_1} \big|\mathrm{P}(\bar{\mathrm{x}} = a) - \alpha_-(a)\big| \leqslant \frac{|A_1|}{pT} < \delta_-. \tag{3}$$

Since the conditional entropy $H(X|Y)$ is concave in the joint distribution of (X, Y) (see, e.g., Gossner et al., 2006, Lemma 1), we have

$$\begin{aligned} H(\bar{\mathrm{x}}|\bar{\mathrm{y}}) &\geqslant \frac{1}{pT} \sum_{t=1}^{pT} H(\mathrm{x}_t|\mathrm{y}_t) \\ &\geqslant \frac{1}{pT} \sum_{t=1}^{pT} H(\mathrm{x}_t|\,\mathrm{y}_t, \mathrm{x}_1, \ldots, \mathrm{x}_{t-1}) \\ &= \frac{1}{pT} \sum_{t=1}^{pT} H(\mathrm{x}_t|\,\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}) \end{aligned}$$

where the last equality holds due to the fact that τ is a pure strategy and hence y_t is a deterministic function $\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}$. Using the chain rule for entropy, and since the string $(\mathrm{z}_1, \ldots, \mathrm{z}_{pT})$ is chosen uniformly from F_-, the last expression above is equal to

$$\frac{H(\mathrm{x}_1, \ldots, \mathrm{x}_{pT})}{pT} = \frac{H(\mathrm{z}_1, \ldots, \mathrm{z}_{pT})}{pT} = \frac{\log_2 |F_-|}{pT} \geqslant \gamma_- - \delta_-.$$

As $\gamma_- = H(\alpha_-)$, we conclude that

$$H(\bar{\mathrm{x}}|\,\bar{\mathrm{y}}) > H(\alpha_-) - \delta_-. \tag{4}$$

The inequalities (3) and (4), together with Lemma 3 in Appendix A (applied to $(\bar{\mathrm{x}}, \bar{\mathrm{y}})$), imply that there is a function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\eta(\delta) \underset{\delta \to 0}{\longrightarrow} 0$ such that

$$\mathbf{E}\left[\frac{1}{pT} \sum_{t=1}^{pT} g(\mathrm{x}_t, \mathrm{y}_t)\right] = \mathbf{E}\big[g(\bar{\mathrm{x}}, \bar{\mathrm{y}})\big] \geqslant U(\gamma_-) - \eta(\delta_-).$$

By a similar operation performed on $(x_{pT+1}, y_{pT+1}), \ldots, (x_T, y_T)$ conditional on a realization $h = ((a_1, b_1), \ldots, (a_{pT}, b_{pT}))$ of the first pT stages, we have

$$\mathbf{E}\left[\frac{1}{(1-p)T} \sum_{t=pT+1}^{T} g(\mathrm{x}_t, \mathrm{y}_t) \,\Big|\, h\right] \geqslant U(\gamma_+) - \eta(\delta_+).$$

Therefore,

$$\mathbf{E}\left[\frac{1}{T} \sum_{t=1}^{T} g(\mathrm{x}_t, \mathrm{y}_t)\right] \geqslant (\mathrm{cav})U(\gamma - \theta) - \big(p\eta(\delta_-) + (1-p)\eta(\delta_+)\big).$$

Given any $\varepsilon > 0$, the last term can be made larger than $(\mathrm{cav})U(\gamma) - \varepsilon$ by choosing θ sufficiently small and then taking a sufficiently large T. □

Theorem 4. *For every $\gamma \geqslant 0$ and a function $f : \mathbb{R}_+ \to [1,\infty)$ with $\frac{\log_2 f(t)}{t} \underset{t\to\infty}{\longrightarrow} \gamma$, 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) $\psi_1(t) \leqslant f(t)$ *for every* $t \in \mathbb{N}$,

(ii) $\underline{\lim}_{T\to\infty} \Big(\inf_{\tau\in\Delta(\Sigma_2)} g_T(\hat{\sigma},\tau)\Big) \geqslant (\operatorname{cav} U)(\gamma)$,

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

Proof.[14] *Construction of* Ψ_1: As in the proof of Theorem 3 we will define a particular class of sequences $F \subset A_1^\infty$ and then set

$$\Psi_1 = \big\{\sigma\langle \boldsymbol{a}\rangle : \boldsymbol{a} \in F\big\}. \tag{5}$$

If $\gamma = 0$, then $(\operatorname{cav} U)(\gamma)$ is the maximin payoff in pure actions, w. In this case the set F can be taken as a singleton $\{\boldsymbol{a} = (a,a,a,\ldots)\}$ 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 $\gamma \geqslant \bar{\gamma}$, we assume w.l.o.g. that $\gamma \leqslant \bar{\gamma}$. By modifying $f(t)$ to $\hat{f}(t) = \inf_{s\geqslant t} f(s)^{t/s}$ if necessary, we also assume that $\frac{\log_2 f(t)}{t}$ is nondecreasing in t, or, equivalently,

$$\log_2 f(s) - \log_2 f(t) \geqslant (s-t)\frac{\log_2 f(t)}{t} \quad \text{whenever } s > t. \tag{6}$$

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 nth block consists of stages $t_{n-1}+1$ to t_n. We denote the length of the nth block by d_n, i.e., $d_n = t_n - t_{n-1}$. Second, we 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 $\boldsymbol{a} = (a_1,a_2,\ldots)$ in A_1^∞ whose nth segment $\boldsymbol{a}[n] = (a_{t_{n-1}+1},\ldots,a_{t_n})$ belongs to F_n:

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

Now we describe the construction of the set F_n in detail.

The blocks are chosen so that d_n is increasing, $d_n \underset{n\to\infty}{\longrightarrow} \infty$, $\frac{d_{n-1}}{d_n} \underset{n\to\infty}{\longrightarrow} 1$, and thus $\frac{d_n}{t_n} \underset{n\to\infty}{\longrightarrow} 0$. 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 $\gamma_n \leqslant \bar{\gamma}$. 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 α_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 $\gamma_n = 0$, then α_n is a pure action, say a^*, that guarantees w. In this case we set F_n to be a singleton consisting of $(a^*,\ldots,a^*) \in A_1^{d_n}$. If $\gamma_n > 0$, then F_n is defined to be the set of all sequences $(a_1,\ldots,a_{d_n}) \in A_1^{d_n}$ whose empirical distribution is within $\frac{|A_1|}{d_n}$ of α_n. Formally,

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

Note that $F_n \neq \emptyset$. We complete this part by defining F by (7) and then Ψ_1 by (5).

Construction of $\hat{\sigma} \in \Delta(\Psi_1)$: Let $\mathbf{z} = (z_1,z_2,\ldots)$ be a sequence of A_1-valued random variables such that its nth segment $\mathbf{z}[n] = (z_{t_{n-1}+1},\ldots,z_{t_n})$ is drawn uniformly from F_n, and independently from $\mathbf{z}[1],\ldots,\mathbf{z}[n-1]$. Then define $\hat{\sigma} = \sigma\langle\mathbf{z}\rangle$. Observe that $\hat{\sigma}$ is indeed a mixture of strategies in Ψ_1.

[14] The reader will see that the proof presented here makes use of a lemma stated and proved in Appendix A. One may wonder, however, whether a different proof not relying on the lemma can be devised. Such a proof, though slightly longer, indeed exists and is presented in Appendix B. We have decided to present the proof using the lemma in the main body of this paper because the lemma is used in a crucial way in the proof of Theorem 6 in the next section, and its inclusion here gives a methodological consistency as well as acquaint the reader with how the lemma is used.

Verification of the theorem: It remains to specify the sequence $(\gamma_n)_n$ so that the strategy set Ψ_1 and the mixed strategy $\hat{\sigma} \in \Delta(\Psi_1)$ satisfy the conditions (i), (ii) and (iii) of the theorem.

From (7) it is clear that we can identify each sequence in the set F with an element of $\times_{n=1}^{\infty} F_n$ and each strategy in $\Psi_1(t_N)$ with an element in $\times_{n=1}^{N} F_n$, $N = 1, 2, \ldots$. Hence $\psi_1(t_N) = |\Psi_1(t_N)| = \prod_{n=1}^{N} |F_n|$ for each $N = 1, 2, \ldots$. Since both $\psi_1(t)$ and $f(t)$ are nondecreasing, in order to verify that Ψ_1 has the property (i) of the theorem, it is enough to ensure that[15]

$$\sum_{n=1}^{N} \log_2 |F_n| \leqslant \log_2 f(t_{N-1}) \quad \text{for each } N > 1. \tag{9}$$

Recall that $|F_n| = 1$ for n with $\gamma_n = 0$. For n with $\gamma_n > 0$ we estimate $|F_n|$ in a similar manner as in the proof of Theorem 3:

$$\frac{1}{(d_n+1)^{|A_1|}} 2^{d_n(\gamma_n - \varepsilon_n)} \leqslant |F_n| \leqslant (d_n+1)^{|A_1|} 2^{d_n(\gamma_n + \varepsilon_n)}$$

where $\varepsilon_n \underset{n\to\infty}{\longrightarrow} 0$. Setting $\delta_n = \varepsilon_n + |A_1| \frac{\log_2(d_n+1)}{d_n}$ we have

$$2^{d_n(\gamma_n - \delta_n)} \leqslant |F_n| \leqslant 2^{d_n(\gamma_n + \delta_n)}.$$

Note that the sequence $(\delta_n)_{n\geqslant 2}$ is decreasing and $\delta_n \underset{n\to\infty}{\longrightarrow} 0$.

Thus, to ensure that (9) holds, it is enough to choose $(\gamma_n)_n$ so that

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

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 $n = 1, 2, \ldots$, let $(\mathrm{x}_1^{(n)}, \mathrm{y}_1^{(n)}), \ldots, (\mathrm{x}_{d_n}^{(n)}, \mathrm{y}_{d_n}^{(n)})$ be the random action pairs induced by $(\hat{\sigma}, \tau)$ in the nth block. Fix a realization $h_{n-1} \in A^{t_{n-1}}$ of $(\mathrm{x}_1^{(1)}, \mathrm{y}_1^{(1)}), \ldots, (\mathrm{x}_{d_{n-1}}^{(n-1)}, \mathrm{y}_{d_{n-1}}^{(n-1)})$, i.e., until the end of the $(n-1)$-th block, and let $(\bar{\mathrm{x}}^{(n)}, \bar{\mathrm{y}}^{(n)})$ be a $(A_1 \times A_2)$-valued random variable whose distribution is given by

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

Then, from (8) and the definition of $\hat{\sigma}$, it easily follows that

$$\sum_{a\in A_1} \big|\mathrm{P}\big(\bar{\mathrm{x}}^{(n)} = a\big) - \alpha_n(a)\big| \leqslant \frac{|A_1|}{d_n} < \delta_n. \tag{11}$$

In addition, by the same argument as in the proof of Theorem 3, it follows that

$$H\big(\bar{\mathrm{x}}^{(n)} | \bar{\mathrm{y}}^{(n)}\big) > H(\alpha_n) - \delta_n. \tag{12}$$

Thus by Lemma 3 there is a function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\eta(\delta) \underset{\delta\to 0}{\longrightarrow} 0$ such that

$$\mathbf{E}\left[\frac{1}{d_n} \sum_{k=1}^{d_n} g\big(\mathrm{x}_k^{(n)}, \mathrm{y}_k^{(n)}\big) \,\Big|\, h_{n-1}\right] = \mathbf{E}\big[g\big(\bar{\mathrm{x}}^{(n)}, \bar{\mathrm{y}}^{(n)}\big) \,\big|\, h_{n-1}\big] \geqslant U(\gamma_n) - \eta(\delta_n).$$

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

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

[15] Observe that (i) asserts that $\psi_1(t) \leqslant f(t)$ for all t which is satisfied, as these functions are nondecreasing, if $\psi_1(t_N) \leqslant f(t_{N-1})$ for each N.

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

$$\frac{1}{t_N}\sum_{n=1}^N d_n U(\gamma_n) \underset{N\to\infty}{\longrightarrow} (\operatorname{cav} U)(\gamma). \tag{13}$$

We now exhibit a choice of $(\gamma_n)_n$ that satisfies (10) and (13). We distinguish two cases.

CASE 1: $(\operatorname{cav} U)(\gamma) = U(\gamma)$. From the assumption $\frac{d_{n-1}}{d_n} \underset{n\to\infty}{\longrightarrow} 1$, it follows that $\frac{d_{n-1}}{d_n}\frac{\log_2 f(t_{n-1})}{t_{n-1}} \underset{n\to\infty}{\longrightarrow} \gamma > 0$. As $\delta_n \underset{n\to\infty}{\longrightarrow} 0$, there is an $\bar n$ such that $\frac{d_{n-1}}{d_n}\frac{\log_2 f(t_{n-1})}{t_{n-1}} > \delta_n$ for all $n > \bar n$. For $n \leqslant \bar n$, set $\gamma_n = 0$, and, for $n > \bar n$, let $\gamma_n = \frac{d_{n-1}}{d_n}\frac{\log_2 f(t_{n-1})}{t_{n-1}} - \delta_n$. With this choice of $(\gamma_n)_n$ it is easy to verify that (10), and hence (9), is satisfied. Since $\gamma_n \underset{n\to\infty}{\longrightarrow} \gamma$, the condition (13) is satisfied as well.

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

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

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

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

With the above choice of the sequence $(\gamma_n)_n$, the inequality (10) trivially holds for $N \leqslant \bar n$. For $N > \bar n$, inequality (10) is proved by induction. Assume that $\sum_{\ell=1}^{N-1} 1(\gamma_\ell > 0) d_\ell(\gamma_\ell + \delta_\ell) \leqslant \log_2 f(t_{N-2})$. Then,

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

where the first inequality holds by the induction hypothesis and since $N > \bar n$, while the second inequality follows from (6). Therefore, if $\gamma_N = \gamma_-$ then inequality (10) holds for N. Obviously, by the definition of γ_N, if $\gamma_N = \gamma_+$ then inequality (10) again holds for N. We conclude that inequality (10), hence (9), holds for all $N > \bar n$.

From (b) above and since $\gamma_n = \gamma_-$ or γ_+ for all $n > \bar n$, in order to show (13), it suffices to verify that

$$\frac{1}{t_N}\sum_{n=\bar n+1}^N d_n \mathbf{1}(\gamma_n = \gamma_-) \underset{N\to\infty}{\longrightarrow} p. \tag{14}$$

For each N let $M_N = \max\{n \leqslant N\colon \gamma_n = \gamma_-\}$. Note that $M_N \to \infty$ as $N \to \infty$ since $\gamma_+ > \gamma$. Since $\frac{\log_2 f(t_{n-1})}{t_n} = (1 - \frac{d_n}{t_n})\frac{\log_2 f(t_{n-1})}{t_{n-1}} \underset{n\to\infty}{\longrightarrow} \gamma$, we have, for every $\delta > 0$, an N such that $\frac{\log_2 f(t_{n-1})}{t_n} \geqslant \gamma - \delta$ for all $n \geqslant M_N$. Hence

$$\begin{aligned}&\sum_{n=1}^N 1(\gamma_n > 0) d_n(\gamma_n + \delta_n) \\ &\quad= \sum_{n=1}^{M_N} \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \delta_n) + \sum_{n=M_N+1}^{N} \mathbf{1}(\gamma_n > 0) d_n(\gamma_n + \delta_n)\end{aligned}$$

$$\geqslant \log_2 f(t_{M_N-1}) - d_{M_N}(\gamma_+ - \gamma_-) + (t_N - t_{M_N})\gamma_+$$
$$\geqslant t_{M_N}(\gamma - \delta) - d_N(\gamma_+ - \gamma_-) + (t_N - t_{M_N})\gamma$$
$$\geqslant t_N(\gamma - \delta) - d_N(\gamma_+ - \gamma_-).$$

Since $\sum_{n=1}^N 1(\gamma_n > 0)d_n(\gamma_n + \delta_n) \leqslant \log f(t_{N-1})$, $\delta_n \underset{n\to\infty}{\longrightarrow} 0$, and $\frac{d_N}{t_N} \underset{N\to\infty}{\longrightarrow} 0$, we conclude that $\frac{1}{t_N}\sum_{n=1}^N d_n\gamma_n \underset{N\to\infty}{\longrightarrow} \gamma$, which, together with a), implies (14).

Finally we verify that $\hat\sigma$ has the property (iii). By the same argument as above, one can show that, for every $\tau \in \Sigma_2$,

$$\sum_{n=1}^N \mathbf{E}_{\hat\sigma,\tau}\left[\sum_{k=1}^{d_n} g\big(\mathrm{x}_k^{(m+n)}, \mathrm{y}_k^{(m+n)}\big)\,\Big|\, h_m\right] \geqslant \sum_{n=1}^N d_{m+n}U(\gamma_{m+n}) - \sum_{n=1}^N d_{m+n}\eta(\delta_{m+n})$$

holds for every m and $h_m \in A^{t_m}$. The analysis of Case 1 and Case 2 above (performed conditional on h_m) together with a classical result in probability theory implies that, for any $\tau \in \Sigma_2$,

$$\varliminf_{N\to\infty} \frac{1}{t_N}\sum_{n=1}^N\sum_{k=1}^{d_n} g\big(\mathrm{x}_k^{(n)}, \mathrm{y}_k^{(n)}\big) \geqslant (\operatorname{cav} U)(\gamma) \quad \text{almost surely}$$

from which (iii) readily follows. □

Remark 3.1. An additional property of ψ_1 constructed in the above proof is that $\lim_{t\to\infty}\frac{\log_2\psi_1(t)}{t} = \gamma$. 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 $\varlimsup_{t\to\infty}\frac{\log_2\psi_1(t)}{t} \leqslant \gamma$. On the other hand, the property (ii) and our previously published result on strategic entropy (Neyman and Okada, 2000a) imply that $\varliminf_{t\to\infty}\frac{\log_2\psi_1(t)}{t} \geqslant \gamma$. To see this, let us recall that the t-strategic entropy of player 1's strategy σ is defined by $H^t(\sigma) = \max_{\tau\in\Sigma_2} H(X_1,\ldots,X_t)$ where $X_1,\ldots,X_t$ is the random sequence of action profiles induced by (σ,τ). Lemma 2 in Section 3.2 then implies that $H^T(\sigma) \leqslant \log_2\psi_1(T)$ for all $\sigma \in \Delta(\Psi_1)$ and T. This, together with Theorem 5.1 of Neyman and Okada (2000a), implies that

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

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

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 $\mathbf{B}_1(\kappa) = \{\sigma \wedge \kappa\colon \sigma \in \Sigma_1\}$, the set of κ-recall strategies. Player 2 is assumed to have full recall. Let $G^*(\kappa)$ be the repeated game under consideration and let $\mathrm{V}(\kappa)$ be player 1's minimax payoff in $G^*(\kappa)$. The results in this section characterize $V(\kappa)$ in terms of asymptotic behavior of the recall function κ. We will also discuss a folk theorem in repeated games with nonstationary bounded recall strategies at the end of the section.

Recall that, with $\Psi_1 = \mathbf{B}_1(\kappa)$, we have $\log_2\psi_1(t) \leqslant c\,m^{\bar\kappa(t)}$ for a constant c (e.g., $c = m/(m-1)$) where $\bar\kappa(t) = \max_{s\leqslant t}\kappa(s)$. 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[16] $\frac{\log_2\psi_1(t)}{t} \underset{t\to\infty}{\longrightarrow} 0$. Hence, by Theorem 1 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.

[16] For the conclusion it suffices to have $\kappa(t) < \frac{\log_2 t}{\log_2 m} - c(t)$ where $c(t) \underset{t\to\infty}{\longrightarrow} \infty$.

Theorem 5.

$$\textit{If} \quad \overline{\lim_{t\to\infty}} \frac{\kappa(t)}{\log_2 t} < \frac{1}{\log_2 m}, \quad \textit{then } \mathrm{V}(\kappa) = w.$$

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 constant times $\log_2 t$ stages back. It is thus natural to ask,"How fast should κ 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, $K_1 \geqslant 1/\log_2 m$.) In particular, 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 α 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* $\underline{\lim}_{t\to\infty} \frac{\kappa(t)}{\log_2 t} > K_1(G)$, *then there is a* $\hat{\sigma} \in \Delta(\mathbf{B}_1(\kappa))$ *with the following properties*:

(i) $\underline{\lim}_{T\to\infty} \left(\min_{\tau\in\Sigma_2} g_T(\hat{\sigma}, \tau)\right) \geqslant v,$

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

In particular, if G is zero-sum, then the repeated game $G^*(\kappa)$ *has the value*[17] *that coincides with the value of G. Moreover, player 1 has an optimal strategy* $\hat{\sigma} \in \Delta(\mathbf{B}_1(\kappa))$.

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 α^* is a pure action, then the theorem is trivially true. So suppose that α^* is not a pure action.

First, we recall the argument from Lehrer (1988) which yields a result about the value of the finitely repeated game where player 1 has bounded recall. Let $\mathbf{z} = (z_1, z_2, \ldots,)$ be a sequence of A_1-valued i.i.d. r.v.'s with $z_t \sim \alpha^*$. Note that for every $k < s < s'$ the probability that $(z_{s-k}, \ldots, z_{s-1}) = (z_{s'-k}, \ldots, z_{s'-1})$ is at most $\zeta(G)^k$. Therefore, the probability that $(z_{s-k}, \ldots, z_{s-1}) = (z_{s'-k}, \ldots, z_{s'-1})$ for some $k < s < s' \leqslant T$ (where $k < T-1$) is at most $\zeta(G)^k(T-k)^2/2$ which can be made arbitrarily close to 0 by choosing a large enough T and $k > K_1(G)\log_2 T$. Now, if $z_1, \ldots, z_T$ is a sequence in A_1^T such that $(z_{s-k}, \ldots, z_{s-1}) \neq (z_{s'-k}, \ldots, z_{s'-1})$ for every $k < s < s' \leqslant T$, then the oblivious strategy that plays this sequence is in $\bar{\mathbf{B}}_i(k)$, the set of stationary bounded recall strategies of size k. Therefore, if we set $\kappa(t) = \min\{t-1, k\}$, then the value of $G^T(\kappa(t))$ is close to the value of the stage game G. Note that this argument relies on two properties of $\kappa(t)$: $\kappa(t) \geqslant K_1(G)\log T$ for $t > k$, and $\kappa(t) = t-1$ for $t \leqslant k$. In particular, if $\kappa(t) > K_1(G)\log T$ for every t, the two properties above are satisfied with $k = [K_1 \log T] + 1$.

With the above discussion in mind we proceed to the main part of the proof. Let $a^0, a^1 \in A_1$ be two distinct actions with $\alpha^*(a^0) > 0$ and $\alpha^*(a^1) > 0$. (Recall that α^* is not a pure action.)

In order to define the strategy $\hat{\sigma}$, we introduce some notation. First, our condition on κ implies that there are infinitely many stages t at which $\kappa(t) \neq \kappa(s)$ for all $s < t$. We enumerate these stages as $t_1 < t_2 < \ldots < t_n < \cdots$.

[17] The zero-sum game $G^*(\kappa)$ has a value v if for every $\varepsilon > 0$ there are strategies σ^*_ε and τ^*_ε (ε-optimal strategies) such that for all sufficiently large T and for all strategies σ and τ we have $g_T(\sigma^*_\varepsilon, \tau) \geqslant v - \varepsilon$ and $g_T(\sigma, \tau^*_\varepsilon) \leqslant v + \varepsilon$. A strategy is optimal if it is ε-optimal for all $\varepsilon > 0$.

For each sequence of player 1's actions $\boldsymbol{a} = (a_1, a_2, \ldots) \in A_1^\infty$ and each positive integer n, define a sequence $\boldsymbol{a}^n = (a_1^n, a_2^n, \ldots) \in A_1^\infty$ 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 $t_n - 1$ stages, then a^1 at stage t_n, and thereafter the actions appearing in the original sequence $\boldsymbol{a}$.

Note that the sequence of actions $\boldsymbol{a}^n$ is induced by the κ-recall (oblivious) strategy $\sigma\langle \boldsymbol{a}^n \rangle \wedge \kappa$ if, and only if

$$\forall t, t' \quad \text{s.t.} \quad a_t^n \neq a_{t'}^n\text{: } (a_{t'-\kappa(t')}^n, \ldots, a_{t'-1}^n) \neq (a_{t-\kappa(t)}^n, \ldots, a_{t-1}^n). \tag{15}$$

For t and t' with $t_n \leqslant t < t'$ and $t' - \kappa(t') < t_n$, we have $(a_{t'-\kappa(t')}^n, \ldots, a_{t'-1}^n) \neq (a_{t-\kappa(t)}^n, \ldots, a_{t-1}^n)$. In addition, since $\kappa(t_n) \neq \kappa(t)$ for all $t < t_n$, the condition (15) is implied by

$$\forall t' \quad \text{s.t.} \quad t' - \kappa(t') > t_n, \quad \forall t < t'\text{: } (a_{t'-\kappa(t')}, \ldots, a_{t'-1}) \neq (a_{t-\kappa(t)}^n, \ldots, a_{t-1}^n). \tag{16}$$

Observe that, for any n, t, and k, $(a_{t-k}^n, \ldots, a_{t-1}^n)$ is one of at most $k+2$ strings of length k: $(a^0, \ldots, a^0)$, $(a^0, \ldots, a^0, a^1)$, $(a^0, \ldots, a^0, a^1, a_{t-\ell}, \ldots, a_{t-1})$ where $1 \leqslant \ell \leqslant k-2$, $(a^1, a_{t-k+1}, \ldots, a_{t-1})$, and $(a_{t-k}, \ldots, a_{t-1})$. Let us denote the subset of A_1^k consisting of these strings by $Z(a_{t-k}, \ldots, a_{t-1})$ so that (16), hence (15), is further implied by

$$\forall t' \quad \text{s.t.} \quad t' - \kappa(t') > t_n, \quad \forall t < t'\text{: } (a_{t'-\kappa(t')}, \ldots, a_{t'-1}) \notin Z(a_{t-\kappa(t)}, \ldots, a_{t-1}). \tag{17}$$

To formally define the strategy $\hat{\sigma}$, let $\mathbf{z} = (z_1, z_2, \ldots,)$ 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 $\mathbf{z}^n = (z_1^n, z_2^n, \ldots)$ 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 an $\mathbb{N}$-valued r.v. $\boldsymbol{\nu}$ by $\boldsymbol{\nu} = n$ if n is the smallest positive integer with the property

$$\forall t' \quad \text{s.t.} \quad t' - \kappa(t') > t_n, \quad \forall t < t'\text{: } (z_{t'-\kappa(t')}, \ldots, z_{t'-1}) \notin Z(z_{t-\kappa(t)}, \ldots, z_{t-1}). \tag{18}$$

Then define $\hat{\sigma} = \sigma\langle \mathbf{z}^{\boldsymbol{\nu}} \rangle \wedge \kappa$. Below we 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 $\{\sigma\langle \boldsymbol{a}^n \rangle \wedge \kappa\text{: } \boldsymbol{a} \in A_1^\infty,\ n = 1, 2, \ldots\} \subset \mathbf{B}_1(\kappa)$.

To see this, observe that

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

For the second-to-the-last inequality, we used the fact that $|Z(z_{t-k}, \ldots, z_{t-1})| \leqslant k+2$, as observed above. The last inequality follows from the inequality $|B_k|^2 < 2\sum_{t \in B_k} t$. Our condition on κ implies that there is an $\varepsilon > 0$ and a $\hat{t}$ such that $\kappa(t) \geqslant (K_1(G) + 2\varepsilon)\log_2 t$ for all $t \geqslant \hat{t}$. Therefore there is a $0 < \theta < 1$ and a $\tilde{t} \geqslant \hat{t}$ such that $\theta\kappa(t) \geqslant (K_1(G) + \varepsilon)\log_2 t$ and $(\kappa(t)+2)\zeta(G)^{\kappa(t)} < \zeta(G)^{\theta\kappa(t)}$ for all $t \geqslant \tilde{t}$. As $\zeta(G) < 1$, it follows that $t(\kappa(t)+2)\zeta(G)^{\kappa(t)} \leqslant$

$t^{-(1+\varepsilon|\log_2 \zeta(G)|)}$ for all $t \geqslant \tilde{t}$. Hence $\sum_{t=1}^{\infty} t(\kappa(t)+2)\zeta(G)^{\kappa(t)} < \infty$. Therefore, with probability 1, there are only finitely many pairs (t, t') with $t < t'$ and $(\mathrm{z}_{t'-\kappa(t')}, \ldots, \mathrm{z}_{t'-1}) \in Z(\mathrm{z}_{t-\kappa(t)}, \ldots, \mathrm{z}_{t-1})$. Thus $\boldsymbol{\nu} < \infty$ 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 $(\mathrm{x}_1, \mathrm{y}_1), \ldots, (\mathrm{x}_T, \mathrm{y}_T)$ be the random action pairs induced by $(\hat{\sigma}, \tau)$. Let $(\bar{\mathrm{x}}_T, \bar{\mathrm{y}}_T)$ be a $(A_1 \times A_2)$-valued random variable such that $\mathrm{P}((\bar{\mathrm{x}}_T, \bar{\mathrm{y}}_T) = (a, b)) = \frac{1}{T}\sum_{t=1}^{T} \mathrm{P}((\mathrm{x}_t, \mathrm{y}_t) = (a, b))$. Then, as in the proof of Theorem 4, we have

$$\begin{aligned} H(\bar{\mathrm{x}}_T|\bar{\mathrm{y}}_T) &\geqslant \frac{1}{T}\sum_{t=1}^{T} H(\mathrm{x}_t|\mathrm{y}_t) \geqslant \frac{1}{T}\sum_{t=1}^{T} H(\mathrm{x}_t|\mathrm{y}_t, \mathrm{x}_1, \ldots, \mathrm{x}_{t-1}) \\ &= \frac{1}{T}\sum_{t=1}^{T} H(\mathrm{x}_t|\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}) \\ &\geqslant \frac{1}{T}\sum_{t=1}^{T} H\big(\mathrm{x}_t|\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}, t_\nu \wedge (T+1)\big) \end{aligned} \tag{19}$$

where the equality follows from y_t being a deterministic function of $\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}$.

Observe that $\mathrm{z}_t \in A_1$ and thus a conditional entropy of z_t is bounded by $\log_2|A_1|$, and, conditional on $t_\nu < t$, we have $\mathrm{z}_t = \mathrm{z}_t^\nu = \mathrm{x}_t$. Therefore, $H(\mathrm{z}_t|\mathrm{z}_1, \ldots, \mathrm{z}_{t-1}, t_\nu \wedge (T+1)) \leqslant H(\mathrm{x}_t|\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}, t_\nu \wedge (T+1)) + \log_2|A_2|\mathrm{P}(t \leqslant t_\nu)$. By rearranging the terms we have

$$H\big(\mathrm{x}_t|\mathrm{x}_1, \ldots, \mathrm{x}_{t-1}, t_\nu \wedge (T+1)\big) \geqslant H\big(\mathrm{z}_t|\mathrm{z}_1, \ldots, \mathrm{z}_{t-1}, t_\nu \wedge (T+1)\big) - \log_2|A_1|\mathrm{P}(t \leqslant t_\nu).$$

From the chain of inequalities beginning (19), 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

$$\begin{aligned} H(\bar{\mathrm{x}}_T|\bar{\mathrm{y}}_T) &\geqslant \frac{1}{T}\sum_{t=1}^{T} H\big(\mathrm{z}_t|\mathrm{z}_1, \ldots, \mathrm{z}_{t-1}, t_\nu \wedge (T+1)\big) - \frac{1}{T}\sum_{t=1}^{T}\log_2|A_1|\mathrm{P}(t \leqslant t_\nu) \\ &= \frac{1}{T} H\big(\mathrm{z}_1, \ldots, \mathrm{z}_T|t_\nu \wedge (T+1)\big) - \frac{1}{T}\sum_{t=1}^{T}\log_2|A_1|\mathrm{P}(t \leqslant t_\nu) \\ &\geqslant \frac{1}{T} H(\mathrm{z}_1, \ldots, \mathrm{z}_T\big) - \frac{1}{T}\log_2(T+1) - \frac{1}{T}\sum_{t=1}^{T}\log_2|A_1|\mathrm{P}(t \leqslant t_\nu) \\ &= H(\alpha^*) - o(1) \underset{T\to\infty}{\longrightarrow} H(\alpha^*). \end{aligned}$$

Note that the $o(1)$ function in the last line is independent of τ. To summarize, for every $\delta > 0$, there is a T_0 such that, for every $T \geqslant T_0$ 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 $(\bar{\mathrm{x}}_T, \bar{\mathrm{y}}_T)$) obeys

$$H(\bar{\mathrm{x}}_T|\bar{\mathrm{y}}_T) \geqslant H(\alpha^*) - \delta. \tag{20}$$

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

$$\mathrm{P}(\mathrm{x}_t = a) \geqslant \mathrm{P}(\mathrm{z}_t = a) - P(t_\nu \geqslant t) = \alpha^*(a) - P(t_\nu \geqslant t)$$

and so $\sum_{a\in A}|\mathrm{P}(\mathrm{x}_t = a) - \alpha^*(a)| \leqslant 2|A_1|\mathrm{P}(t_\nu \geqslant t)$. Hence

$$\begin{aligned} \sum_{a\in A_1}\big|\mathrm{P}(\bar{\mathrm{x}}_T = a) - \alpha^*(a)\big| &= \sum_{a\in A_1}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{P}(\mathrm{x}_t = a) - \alpha^*(a)\right| \\ &\leqslant 2|A_1|\frac{1}{T}\sum_{t=1}^{T}\mathrm{P}(t_\nu \geqslant t) \underset{T\to\infty}{\longrightarrow} 0. \end{aligned}$$

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

$$\sum_{a\in A_1} \left|\mathrm{P}(\bar{\mathrm{x}}_T = a) - \alpha^*(a)\right| < \delta. \tag{21}$$

Hence, by Lemma 3, the inequalities (20) and (21) imply that, for every $\varepsilon > 0$ there is a T_0 such that for all $T \geqslant T_0$,

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

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 $(\mathrm{x}_1, \mathrm{y}_1, \ldots, \mathrm{x}_s, \mathrm{y}_s)$, 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 $T \geqslant T_0$, we have

$$\mathbf{E}_{\hat{\sigma},\tau}\left[\frac{1}{T}\sum_{t=s+1}^{s+T} g(\mathrm{x}_t, \mathrm{y}_t) \,\Big|\, \mathrm{x}_1, \mathrm{y}_1, \ldots, \mathrm{x}_s, \mathrm{y}_s\right] \geqslant v - \varepsilon,$$

which, by the classical results in probability, implies that

$$\varliminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} g(\mathrm{x}_t, \mathrm{y}_t) \geqslant v - \varepsilon \quad \text{almost surely.}$$

As this holds for every $\varepsilon > 0$ the second part of the theorem follows. □

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

Since $\nu < \infty$ almost surely, we have $\mathrm{P}(\nu \leqslant n) \underset{n\to\infty}{\longrightarrow} 1$. Hence if we choose an n sufficiently large, the condition (18) 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 $T \geqslant T_0$, we have $g_T(\sigma\langle \mathbf{z}^n\rangle, \tau) > v - \varepsilon$. In the case where G is a zero-sum game, this shows that, for every $\varepsilon > 0$, the strategy $\sigma\langle \mathbf{z}^n\rangle$ is ε-optimal for a sufficiently large n.

To conclude this section we discuss an immediate implication of Theorem 6 for the set of equilibrium payoff vectors. 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 $\varliminf_{t\to\infty} \frac{\kappa_i(t)}{\log_2 t} > K_i(G)$ for $i = 1, 2$, then Theorem 6, or the strategy constructed in its proof, provides the players with threats to one another which discourage them from deviating from a path[18] that yields $x_i \geqslant v_i$, $i = 1, 2$. In order to state a version of the folk theorem in this context, let $E^*(\kappa_1, \kappa_2)$ be the set of equilibrium payoff vectors of $G^*(\kappa_1, \kappa_2)$ and let $F^* = \{(x_1, x_2) \in \mathrm{co}(g(A))\colon\ x_1 \geqslant v^1, x_2 \geqslant v^2\}$.

Theorem 7. *There is a constant $K^*(G) > 0$ (that depends on the stage game) such that if*

$$\varliminf_{t\to\infty} \min_{i=1,2} \frac{\kappa_i(t)}{\log_2 t} > K^*(G), \quad \text{then } E^*(\kappa_1, \kappa_2) = F^*.$$

5. Remarks on finite automata with time-varying complexity

In the last section we examined a generalization of bounded recall strategy to accommodate time-varying memory. Finite automaton is another useful representation of strategies in the literature on strategic complexity in repeated

[18] Specifically, let ξ be a sequence of action pairs that yields an average payoff vector as close to (x_1, x_2) as one desires. Then we can let the players play a constant action pair $(\bar{a}, \bar{b})$ at each stage until their recall is long enough, and then start the cyclic play of ξ. Thus it can be ensured that any deviation by one player from cooperative phase will be in the memory of the other player long enough to initiate and continue the punishment strategy.

games. The number of states of automaton is usually taken as a measure of complexity. Here the corresponding notion would be called automata with "growing (or, more generally, time-varying) number of states." A correct definition of this concept, however, appears more delicate. One must take an appropriate care to preserve certain consistency of transition among states (or subset of states). There may possibly be more than one reasonable definitions. This topic deserves further investigation.

It is reasonable to anticipate that whatever the definition of an automaton with a growing number of states (namely the number of states at stage t is $m(t)$ with $m(t) \leqslant m(t+1)$) is, the number $\psi_i(t)$ of strategies induced in the first t stages will be bounded by $m(t)^{Cm(t)}$ where C is a positive constant dependent on the stage game. Therefore, some of the results, e.g., Theorem 1 and Theorem 2, that bound from above the value of the repeated game as a function of the function $\psi_i(t)$ will have implications to finite automata with a growing number of states independent of the exact definition. An analog of Theorem 5 for automata would be "if $m(t)\log m(t) = o(t)$ then $\mathrm{V}(m) = w$" where $\mathrm{V}(m)$ is the minimax payoff to player 1 of the two-player repeated game in which player 1 uses growing automata whose complexity grows according to $m : \mathbb{N} \to \mathbb{N}$ and player 2 is fully rational. Similarly, if we allow for the more general time-varying number of states, the same comment applies where $m(t)$ is replaced by $\max_{s \leqslant t} m(s)$.

Acknowledgments

Okada gratefully acknowledges support from The Research Council at Rutgers University and Richard B. Fisher Membership at Institute for Advanced Study. The authors thank anonymous referees for helpful comments.

Appendix A

Here we 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 Ω is $\|\mathrm{P}-\mathrm{Q}\|_1 = \sum_{x\in\Omega} |\mathrm{P}(x) - \mathrm{Q}(x)|$. For a $(A_1 \times A_2)$-valued random variable (x,y), we write $\mathrm{P}(a,b)$ for $\mathrm{P}(\mathrm{x}=a, \mathrm{y}=b)$, and $\mathrm{P}_1(a)$ (resp. $\mathrm{P}_2(b)$) for $\mathrm{P}(\mathrm{x}=a)$ (resp. $\mathrm{P}(\mathrm{y}=b)$). Also, $\mathrm{P}_1 \otimes \mathrm{P}_2$ is a probability on $A_1 \times A_2$ with $(\mathrm{P}_1 \otimes \mathrm{P}_2)(a,b) = \mathrm{P}_1(a)\mathrm{P}_2(b)$ for each $(a,b) \in A_1 \times A_2$.

Lemma 3. *There exists a function $\eta : \mathbb{R}_+ \to \mathbb{R}_+$, which depends on the stage game, with the following properties*:

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

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 $\mathrm{P}_1 \otimes \mathrm{P}_2$ 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 $H(\mathrm{x}) \geqslant H(\mathrm{x}|\mathrm{y}) \geqslant H(\alpha) - \delta$. Next, since H, as a function on $\Delta(A_1)$, is uniformly continuous with respect to the L_1-norm, (ii) implies that $H(\mathrm{x}) \leqslant H(\alpha) + \theta(\delta)$ where[19] $\theta(\delta) > 0$ and $\theta(\delta) \underset{\delta\to 0}{\longrightarrow} 0$. Thus

$$H(\mathrm{x}) - H(\mathrm{x}|\mathrm{y}) \leqslant \theta(\delta) + \delta. \tag{22}$$

Let us recall that the relative entropy between two probabilities P and Q on $A_1 \times A_2$ is defined by $D(\mathrm{P}\|\mathrm{Q}) = \sum_{a,b} \mathrm{P}(a,b) \log_2 \frac{\mathrm{P}(a,b)}{\mathrm{Q}(a,b)}$ 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, $\mathrm{P} = \mathrm{Q}$. From this definition, it is easy to verify that $D(\mathrm{P}\|\mathrm{P}_1 \otimes \mathrm{P}_2) = H(\mathrm{x}) - H(\mathrm{x}|\mathrm{y})$. Observe that $D(\mathrm{P}\|\mathrm{P}_1 \otimes \mathrm{P}_2) = 0$ if, and only if, x and y are independent. It can be shown (Cover and Thomas, 1991, p. 300) that $D(\mathrm{P}\|\mathrm{Q}) \geqslant \frac{1}{2\ln 2}\|\mathrm{P}-\mathrm{Q}\|_1^2$, and hence $H(\mathrm{x}) - H(\mathrm{x}|\mathrm{y}) \geqslant \frac{1}{2\ln 2}\|\mathrm{P} - \mathrm{P}_1 \otimes \mathrm{P}_2\|_1^2$. It follows from (22) that

$$\|\mathrm{P} - \mathrm{P}_1 \otimes \mathrm{P}_2\|_1 = \sum_{a,b} \bigl|\mathrm{P}(a,b) - \mathrm{P}_1(a)\mathrm{P}_2(b)\bigr| \leqslant \sqrt{2(\theta(\delta)+\delta)\ln 2}. \tag{23}$$

[19] In fact, one can take $\theta(\delta) = -\delta \log_2 \frac{\delta}{|A_1|}$ for $\delta \leqslant \frac{1}{2}$. See Cover and Thomas (1991, p. 488).

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

$$\begin{aligned}\mathbf{E}\big[g(\mathrm{x},\mathrm{y})\big] &= \sum_{a,b} \mathrm{P}(a,b)g(a,b)\\ &\geqslant \sum_{a,b} \mathrm{P}_1(a)\mathrm{P}_2(b)g(a,b) - \|g\|\xi(\delta) \quad \text{by (23)}\\ &\geqslant \sum_{a,b} \alpha(a)\mathrm{P}_2(b)g(a,b) - \|g\|\delta - \|g\|\xi(\delta) \quad \text{by (ii)}\\ &\geqslant \min_b g(\alpha,b) - \|g\|\big(\delta+\xi(\delta)\big).\end{aligned}$$

This completes the proof. □

Appendix B

We present an alternative proof of Theorem 4 that does not make use of Lemma 3. The construction of the strategy set Ψ_1 is similar to that presented in the main text and we avoid duplicating detailed descriptions of the notations used.

Construction of Ψ_1: The set F_n is now defined by means of two sequences of nonnegative reals, $(\gamma_n)_n$ with $\gamma_n \leqslant \bar{\gamma}$ and $(\eta_n)_n$ where $\eta_n \geqslant \frac{|A_1|}{d_n}$ and $\eta_n \underset{n\to\infty}{\longrightarrow} 0$. As before, if $\gamma_n = 0$, then we set F_n to be a singleton. For n with $\gamma_n > 0$, we let

$$F_n = \left\{(a_1,\ldots,a_{d_n}) \in A_1^{d_n}\colon \sum_{a\in A_1}\left|\frac{1}{d_n}\sum_{k=1}^{d_n}\mathbf{1}(a_k = a) - \alpha_n(a)\right| \leqslant \eta_n\right\}. \tag{24}$$

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

The sequence $(\eta_n)_n$ is chosen to satisfy, in addition, the following property. Let $\mathbf{x} = (\mathrm{x}_1,\mathrm{x}_2,\ldots)$ be a sequence of independent A_1-valued random variables where x_t is distributed according to α_n whenever t is in the nth block, i.e. $t_{n-1}+1 \leqslant t \leqslant t_n$. Then we require[20]

$$\sum_{n=1}^{\infty} \mathrm{P}(\mathbf{x}[n] = (\mathrm{x}_{t_{n-1}+1},\ldots,\mathrm{x}_{t_n}) \notin F_n) < \infty. \tag{25}$$

As before, we define $F = \{\boldsymbol{a} = (a_1,a_2,\ldots) \in A_1^{\infty}\colon \boldsymbol{a}[n] = (a_{t_{n-1}+1},\ldots,a_{t_n}) \in F_n\}$ and then $\Psi_1 = \{\sigma\langle\boldsymbol{a}\rangle\colon \boldsymbol{a}\in F\}$.

Construction of $\hat{\sigma} \in \Delta(\Psi_1)$: Define a sequence of A_1-valued random variables $\hat{\mathbf{x}} = (\hat{\mathrm{x}}_1,\hat{\mathrm{x}}_2,\ldots)$ by

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

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

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

$$\mathrm{P}\left(\left|\frac{1}{d_n}\sum_{k=1}^{d_n}\mathbf{1}(\mathrm{x}_{t_{n-1}+k} = a) - \alpha_n(a)\right| > \frac{\eta_n}{|A_1|}\right) \leqslant 2\exp\left(-2d_n\frac{\eta_n^2}{|A_1|^2}\right),$$

and so

$$\mathrm{P}\big(\mathbf{x}[n]\notin F_n\big) \leqslant 2|A_1|\exp\left(-2d_n\frac{\eta_n^2}{|A_1|^2}\right).$$

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 (25) holds.

Verification of the theorem: As before, in order to verify that Ψ_1 has the property (i) of the theorem, it is enough to ensure (9). By estimating $|F_n|$ in a manner analogous to the proof in the main text, one sees that it suffices to ensure (10).

To derive a sufficient condition to be verified in order to show that $\hat{\sigma}$ has the property (ii), observe that the L_1-distance between the conditional distributions of $\mathbf{x}[n]$ and $\hat{\mathbf{x}}[n]$ given $\mathbf{x}[1], \ldots, \mathbf{x}[n-1]$ is at most $2\mathrm{P}(\mathbf{x}[n] \notin F_n)$, that is,

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

It follows that, in the nth block, we have

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

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

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

Thus, by virtue of (25), in order to show part (ii) of the theorem it suffices to choose $(\gamma_n)_n$ so that (13) holds.

The part of the proof that exhibits a choice of $(\gamma_n)_n$ that satisfy (10) and (13) as well as the verification of (iii) is identical to the one in the main text and thus is omitted.

References

Anderlini, L., 1990. Some notes on Church's thesis and the theory of games. Theory 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. Bibliographisches Instituit, Mannheim, pp. 11–42.

Aumann, R.J., 1997. Rationality and bounded rationality. Games Econ. Behav. 21 (1–2), 2–14.

Aumann, R.J., Shapley, L.S., 1994. Long-term competition—A game theoretic analysis. In: Megiddo, N. (Ed.), Essays in Game Theory in Honor of Michael Maschler. Springer-Verlag, New York, pp. 1–5.

Ben-Porath, E., 1993. Repeated games with finite automata. J. Econ. Theory 59 (1), 17–32.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley, New York.

Gossner, O., Vieille, N., 2002. How to play with a biased coin? Games Econ. Behav. 41 (2), 206–226.

Gossner, O., Hernandez, P., Neyman, A., 2006. Optimal use of communication resources. Econometrica 74 (6), 1603–1636.

Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. 58 (301), 13–30.

Lacôte, G., 2005. Boundedly complex Turing strategies play the repeated Prisoner's Dilemma: Some results. Draft.

Lehrer, E., 1988. Repeated games with stationary bounded recall strategies. J. Econ. Theory 46 (1), 130–144.

Neyman, A., 1985. Bounded complexity justifies cooperation in the finitely repeated Prisoner's Dilemma. Econ. Letters 19 (3), 227–229.

Neyman, A., 1997. Cooperation, repetition, and automata. In: Hart, S., Mas-Colell, A. (Eds.), Cooperation: Game-Theoretic Approaches. In: NATO ASI-Series F, vol. 155. Springer-Verlag, New York, pp. 233–255.

Neyman, A., 2003. Algorithmic strategic complexity. Discussion paper. Presentation in a PDF file available at: http://ratio.huji.ac.il/dp/neyman/algorslidesnwu.pdf.

Neyman, A., Okada, D., 2000a. Repeated games with bounded entropy. Games Econ. Behav. 30 (2), 228–247.

Neyman, A., Okada, D., 2000b. Two-person repeated games with finite automata. Int. J. Game Theory 29 (3), 309–325.

O'Connell, T.C., Stearns R.E., 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. Quart. J. Econ. 69 (1), 99–118.

Simon, H.A., 1972. Theories of bounded rationality. In: McGuire, C., Radner, R. (Eds.), Decision and Organization. North-Holland, Amsterdam.

Stearns, R.E., 1997. Memory-bounded game-playing computing devices. Discussion paper. Mimeo, Department of Computer Science, SUNY at Albany.