Limits of Correlation with Bounded Complexity

Gilad Bavly Ron Peretz*

Preliminary version, February 19, 2016

Abstract

While Peretz (2013) showed that, perhaps surprisingly, players whose recall is bounded can correlate in a long repeated game against a player of greater recall capacity, we show that correlation is already impossible against an opponent whose recall capacity is only linearly larger. This result closes a gap in the characterisation of minmax levels, and hence also equilibrium payoffs, of repeated games with bounded recall.

Keywords: repeated games, bounded complexity, bounded recall, concealed correlation.
JEL classification: C72, C73.

1 Introduction

This paper concerns repeated games in which the complexity of the players is bounded. In such games the characterisation of the equilibrium payoffs boils down to identifying the individually rational levels of the players (minmax) (Lehrer, 1988, p. 137). That is, in a sufficiently long game, any payoff profile that is feasible and above each player's minmax in the *repeated* game is close to an equilibrium payoff.

Our presentation here focuses on one model of bounded complexity, that of bounded recall (Aumann (1981), Lehrer (1988)). In this model each player i has a recall capacity k_i, where i's strategy can rely only on the previous k_i stages. In Section 5 we explain how our result extends to a few variations

*Bar Ilan University, Israel.

of the model. In particular, we present a counterpart of our result for finite automata, another common model of bounded complexity.

The case of two players is well understood (Lehrer (1988); Ben-Porath (1993)). When there are more than two players, the difficulty in evaluating the minmax lies in the possibility of correlation between a group of players. Note that in a repeated game without any complexity restrictions, i.e., when any strategy can be employed, effective correlation is not possible: the minmax is the same as in the one-shot game. However, the complexity limitations of a player may be exploited to effectively correlate against her.

Thus, we may not be surprised if the minmax of a player is low, compared to her one-shot minmax, when facing stronger players (although the quantification of this phenomena may be complicated). But surprisingly, effective correlation against a stronger player is also possible. Peretz (2013) showed that a pair of players can completely conceal their correlation from a stronger player, i.e., their sequence of actions appear to that third player as if they were i.i.d. playing a correlated distribution.[1] Moreover, for any constant C, there exists a game in which this can be achieved by a pair of players with k-recall against an opponent with Ck-recall.

We prove a converse result to Peretz (2013). For every one-stage game there is a constant C such that the minmax of a (stronger) player with Ck-recall against players with k-recall is close to that player's minmax in the one-stage game. Note that our result (Theorem 2.1) applies to any duration of the repeated game, including infinite repetition.[2]

Thus far, it was only known that a player who is a lot stronger than the opposition, i.e., exponentially stronger, can "see through" their correlation (see (Lehrer, 1988, Theorem 3) and (Bavly and Neyman, 2014, Theorem 2.3)). Therefore, our result closes a significant gap in the characterization of the minmax levels, and consequently the equilibrium payoffs.

The following observation plays an important role in the proof. Consider two players, each choosing a mixed k-recall strategy (in particular, their randomisation is of course independent). Although their continuation strategies, at some stage t of the game, need not be independent,[3] it turns out they cannot be too far from independent, due to the bounded complexity.

Supposing the third player uses an m-recall strategy, perhaps she can exploit this observation during the following m stages or so. However, this

[1] Bavly and Neyman (2014) consider a different setting, in which there are weak players attempting to conceal their correlation against a medium-strength opponent, with the help of strong players.

[2] The case of very short duration is uninteresting, because then the limitation of bounded recall has no bite.

[3] For more on this see Section 4.2.

cannot be done directly, since her strategy cannot depend on the time t. Therefore, we first define an auxiliary zero-sum game related to this scenario, and show that the value of this game is close to the one-shot minmax. The point of the auxiliary zero-sum game is that, by a minmax theorem, there is a mixed *optimal* strategy for the third player, good for any block of m stages.

With an m-recall strategy, the third player cannot play this optimal mixed strategy infinitely many times *independently*. To conclude the proof, we show that it suffices to play a long cycle consisting of independent such plays.

As mentioned above, the linear scale in our result is the best we could hope for, since Peretz (2013) shows that being linearly stronger may not be enough to defend the one-shot minmax.

The result is tight in another sense as well: the stronger player, unless she is extremely strong, cannot hope for more than her one-shot minmax. By Lehrer (1988), the minmax of a player who isn't exponentially stronger than the other players is at most her one-shot minmax. Therefore, we can now tell that asymptotically, the minmax of a "moderately stronger" player is the same as her minmax in the one-stage game.

2 Model and Results

Throughout, a finite three-person game in strategic form is a pair $G = \langle A = A_1 \times A_2 \times A_3, g : A \to [0,1]^3 \rangle$. I.e., it is assumed that the payoffs are scaled[4] between 0 and 1, namely, $0 \le g^i(a) \le 1$, for all $a \in A$ and $i \in \{1,2,3\}$. The minmax value of player $i \in \{1,2,3\}$ is defined as

$$\min\max\nolimits_i G := \min_{x^j \in \Delta(A_j): j \neq i} \max_{a^i \in A_i} g^i(x^{-i}, a^i).$$

where $\Delta(X)$ denotes the set of probability distributions over a finite set X. The correlated minmax value of player $i \in \{1,2,3\}$ is defined as

$$\operatorname{cor}\min\max\nolimits_i G := \min_{x^{-i} \in \Delta(A_{-i}): j \neq i} \max_{a^i \in A_i} g^i(x^{-i}, a^i).$$

For $T \in \mathbb{N} \cup \{\infty\}$, a strategy for player $i \in \{1,2,3\}$ in the T-fold repeated game is a function $s^i : A^{<T} \to A_i$, where $A^{<T} = \bigcup_{0 \le t < T} A^t$. A random variable whose values are strategies is called a *random strategy*. A probability distribution over strategies is called a *mixed strategy*. The set of all

[4]This is merely a normalization: if one considered games with a larger range of payoff, some of our derived constants should simply be multiplied by that range.

strategies for player i is denoted Σ^i_T. For a strategy s^i and a history of play $h_t = (a_1, \ldots, a_t)$, the *continuation strategy* given h_t, denoted $s^i_{|h_t}$, is the strategy induced by s_i and h_t in the remaining stages of the game, i.e., $s^i_{|h_t}(a'_{t+1}, \ldots, a'_{t+r}) = s^i(a_1, \ldots, a_t, a'_{t+1}, \ldots, a'_{t+r})$ for $(a'_{t+1}, \ldots, a'_{t+r}) \in A^r$.

A *k-recall* strategy for player i is a strategy $s^i \in \Sigma^i_\infty$ that depends only on the last k actions. Namely, for any two histories of any length $\bar{a} = (a_1, \ldots, a_{m-1})$ and $\bar{b} = (b_1, \ldots, b_{n-1})$, if $(a_{m-k}, \ldots, a_{m-1}) = (b_{n-k}, \ldots, b_{n-1})$ then $s^i(\bar{a}) = s^i(\bar{b})$.

For a k-recall strategy s^i we can also define the continuation strategy given a k-length suffix of history $h \in A^k$, instead of a complete history. This is of course well-defined, since k-recall implies that for any complete history that ends with h the continuation strategy is the same. This includes, in particular, the case where the complete history is h itself. Hence we can use the above notation, $s^i_{|h}$, also for a continuation strategy of a k-recall strategy given a suffix.

The (finite) set of k-recall strategies for player i is denoted $\Sigma^i(k)$. For natural numbers k_1, k_2, k_3, the un-discounted T-fold repeated version of G where each player i is restricted to k_i-recall strategies is denoted $G^T[k_1, k_2, k_3]$. The payoff for a finite T is the average per-stage payoff, and the limiting average for $T = \infty$. Throughout, we always arrange the players' order such that $k_1 \leq k_2 \leq k_3$.

Our main result is the following theorem:

Theorem 2.1. *There exists a constant $C > 0$ such that for every finite three-person game $G = \langle A, g \rangle$ and every $k_3 \geq k_2 \geq k_1 \geq 0$ and $T \in \mathbb{N} \cup \{\infty\}$,*

$$\min\max{}_3 G^T[k_1, k_2, k_3] \geq \min\max{}_3 G - C\sqrt{\frac{k_2 \ln|A|}{k_3}} \ .$$

3 Preliminaries

This section presents some information-theoretic notions that are used in the proof.

Shannon's entropy[5] of a discrete random variable x is the following non-negative quantity

$$H(x) = -\sum_{\xi} \mathbf{P}(x = \xi) \ln(\mathbf{P}(x = \xi)).$$

[5]In the literature, a similar definition using $\log_2$ instead of ln is also commonly referred to as "Shannon's entropy."

The distribution of x is denoted $p(x)$. We have

$$H(x) \leq \ln(|support(p(x))|).$$

If y is another random variable, the entropy of x given y, defined by the chain rule of entropy $H(x|y) = H(x,y) - H(y)$, satisfies

$$H(x) \geq H(x|y)$$

with equality if and only if x and y are independent. The difference $I(x;y) = H(x) - H(x|y)$ is called the mutual information of x and y. The following identity holds:

$$I(x;y) = I(y;x) = H(x,y) - H(x|y) - H(y|x).$$

If z is yet another random variable, then the mutual information of x and y given z is defined by the chain rule of mutual information:

$$I(x;y|z) = I(x,z;y) - I(z;y).$$

3.1 Neyman-Okada Lemma

Let $x_1, \ldots, x_m, y_1, \ldots, y_m$ be finite random variables, and let t be a random variable that distributes uniformly in $[m] := \{1, \ldots, m\}$ independently of $x_1, \ldots, x_m, y_1, \ldots, y_m$. Suppose that y_0 is a random variable such that each y_i is a function of $y_0, x_1, \ldots, x_{i-1}$. Then,

$$I(x_t; y_t) \leq H(x_t) - \frac{1}{m} H(x_1, \ldots, x_m) + \frac{1}{m} I(y_0; x_1, \ldots, x_m).$$

The interpretation is that $x_1, \ldots, x_m$ is a sequence of actions played by an oblivious player[6], y_0 is a random strategy of another player, and $y_1, \ldots, y_m$ are the actions played by the other player.

A case of special interest is when the oblivious player repeats the same mixed action independently, namely, $x_1, \ldots, x_m$ are i.i.d. . In this case we have

$$I(x_t; y_t) \leq \frac{1}{m} I(y_0; x_1, \ldots, x_m). \tag{3.1}$$

[6] An oblivious player is one who ignores the actions of the other players.

4 Proof of Theorem 2.1

In the proof, for any given pair of mixed strategies σ^1, σ^2 of players 1 & 2, we describe a strategy σ^3 of player 3 that guarantees the required payoff. We divide the stages of the game into blocks, and describe σ^3 for each block. At the beginning of a block player 3 should consider the continuation strategies of 1 and 2. An important point is that these continuation strategies are random variables which are a function of the initial strategies employed by 1 & 2 and of their memories at that point. Generally, the continuation strategies of 1 & 2 need not be independent, not even conditional[7] on the memories of 1 & 2.

This leads us to define and analyze the below auxiliary game. Afterwards, we will use this analysis to describe σ^3.

4.1 An auxiliary two-person zero-sum game

For natural numbers k and m, and mixed strategies $\sigma^i \in \Delta(\Sigma^i_{m+k})$ $(i = 1, 2)$, we define a two-person zero-sum game $\Gamma_{\sigma^1,\sigma^2,k,m}$ between Alice who is the minimiser, and Bob, the maximiser (Alice is related to players 1 & 2 in the original game, and Bob is related to 3). The strategy space of Alice is the set

$$X_A = \left\{\rho \in \Delta(\Sigma^1_{m+k} \times \Sigma^2_{m+k} \times A^k) : \ \rho\text{'s marginal on } \Sigma^1_{m+k} \times \Sigma^2_{m+k} \text{ is } \sigma^1 \otimes \sigma^2\right\}.$$

The strategy space of Bob is $\Delta(\Sigma^3_m)$.

The strategies of Alice can also be described as follows. A pair of strategies $s^1 \in \Sigma^1_{m+k}$ and $s^2 \in \Sigma^2_{m+k}$ is randomly chosen by nature, according to the distribution $\sigma^1 \otimes \sigma^2$. *After* seeing s^1 and s^2, Alice chooses a "memory" $h \in A^k$ (or, more generally, a distribution over A^k).

A pair of strategy realisations $r = (s^1, s^2, h) \in \Sigma^1_{m+k} \times \Sigma^2_{m+k} \times A^k$ and $z \in \Sigma^3_m$ induces a play $a_1, \ldots, a_m$ of players $1, 2, 3$, defined by

$$a^i_t = \begin{cases} s^i(h_1, \ldots, h_k, a_1, \ldots, a_{t-1}) & \text{for } i = 1, 2, \\ z(a_1, \ldots, a_{t-1}) & \text{for } i = 3 \end{cases} \tag{4.1}$$

for any $1 \le t \le m$. That is, we look at an m-fold repeated game, in which Player 3 simply employs the strategy z, and Players 1 & 2 act as if the actual play was preceded by the history h (in other words, they employ $s^i_{|h}$).

Hence, a pair of strategies $\rho \in X_A$ and $\zeta \in X_B$ induces a probability measure over plays of length m. The payoff that Alice pays Bob is defined

[7] We further elaborate on this point in Section 4.2.

by

$$\Gamma_{\sigma^1,\sigma^2,k,m}(\rho,\zeta) = \mathbb{E}_{\rho,\zeta}\left[\frac{1}{m}\sum_{j=1}^{m} g^3(a_j)\right]. \tag{4.2}$$

Since the action spaces are convex and compact, the game $\Gamma_{\sigma^1,\sigma^2,k,m}$ admits a value.

Lemma 4.1. *For every three-person game G, natural numbers k and m, and mixed strategies $\sigma^1 \in \Delta(\Sigma^1_{k+m})$ and $\sigma^2 \in \Delta(\Sigma^2_{k+m})$,*

$$Val(\Gamma_{\sigma^1,\sigma^2,k,m}) \geq \min\max{}_3 G - 2\sqrt{\frac{k \ln |A|}{m}}\ .$$

The rest of this section is devoted to proving Lemma 4.1.

Lemma 4.2. *Let x and y be random variables that assume values in A_1 and A_2 respectively. There exists $a^3 \in A_3$ such that*

$$\mathbb{E}[g^3(x,y,a^3)] \geq g^3(p(x),p(y),a^3) - \sqrt{2I(x;y)} \geq \min\max{}_3 G - \sqrt{2I(x;y)}.$$

Proof. By Pinsker Inequality, $\|p(x,y) - p(x)\otimes p(y)\|_1 \leq \sqrt{2I(x;y)}$. The function g^3 is 1-Lipschitz on $\Delta(A)$ w.r.t. the L_1 norm (because all payoffs are between 0 and 1). Therefore, for any $a^3 \in A_3$, $\mathbb{E}[g^3(x,y,a^3)] \geq g^3(p(x),p(y),a^3) - \sqrt{2I(x;y)}$, and so an action that guarantees $\min\max_3 G$ against $p(x)\otimes p(y)$ guarantees $\min\max_3 G - \sqrt{2I(x;y)}$ against $p(x,y)$. □

We next generalise Lemma 4.2 to repeated games.

Lemma 4.3. *Let s^1 and s^2 be random strategies that assume values in Σ^1_m and Σ^2_m, respectively. There exists a pure strategy $s^3 \in \Sigma^3_m$ such that the play $a_1,\ldots,a_m$ induced by (s^1,s^2,s^3) satisfies*

$$\mathbb{E}\left[\frac{1}{m}\sum_{t=1}^{m} g^3(a_t)\right] \geq \min\max{}_3 G - \sqrt{\frac{2I(s^1;s^2)}{m}}.$$

Proof. The strategy $s^3 \in \Sigma^3_m$ myopically best-responds to (s^1,s^2) on any possible history. Formally, s^3 is defined recursively as follows. Suppose s^3 is already defined on $A^{<t-1}$, for some $1 \leq t < m$. Then, s^1, s^2, and s^3 induce a random play $\bar{a}_{t-1} = (a_1,\ldots,a_{t-1}) \in A^{t-1}$ and random actions for 1 and 2 at time t, a^1_t and a^2_t. We define s^3 on A^{t-1} by choosing

$$s^3(h_{t-1}) \in \arg\max_{a^3\in A_3} \mathbb{E}[g^3(a^{-3}_t,a^3)\mathbf{1}_{\{\bar{a}_{t-1}=h_{t-1}\}}], \quad \forall h_{t-1} \in A^{t-1}.$$

Define $Y(h_{t-1}) = I(a_t^1; a_t^2 | \bar{a}_{t-1} = h_{t-1})$, for every $t \in [m]$ and $h_{t-1} \in A^{t-1}$ such that $\mathbf{P}(\bar{a}_{t-1} = h_{t-1}) > 0$. By Lemma 4.2,

$$\mathbb{E}[g^3(a_t)|\bar{a}_{t-1}] \geq \min\max_3 G - \sqrt{2Y(\bar{a}_{t-1})},$$

for every $t \in [m]$ (recalling that $[x]$ denotes the set $\{1, \ldots, x\}$).

Now, take $\hat{t}$ to be a random variable uniformly distributed in $[m]$ independently of (s^1, s^2). Let $Y = Y(\bar{a}_{\hat{t}})$. Then, by Jensen Inequality,

$$\frac{1}{m}\sum_{t=1}^{m} \mathbb{E}[g^3(a_t)] = \mathbb{E}[g^3(a_{\hat{t}})] \geq \min\max_3 G - \mathbb{E}\left[\sqrt{2Y}\right] \geq \min\max_3 G - \sqrt{2\mathbb{E}[Y]}.$$

Since $\mathbb{E}[Y] = \frac{1}{m}\sum_{t=1}^{m} I(a_t^1; a_t^2|\bar{a}_{t-1})$, it remains to show that

$$\sum_{t=1}^{m} I(a_t^1; a_t^2|\bar{a}_{t-1}) \leq I(s^1; s^2).$$

We use the inequality

$$H(U) \leq I(V; W) + H(U|V) + H(U|W)$$

with $U = \bar{a}_m$, $V = s^1$, and $W = s^2$ to conclude

$$\begin{aligned}
&\sum_{t=1}^{m} I(a_t^1; a_t^2|\bar{a}_{t-1}) \\
&\quad = \sum_{t=1}^{m} H(a_t^1, a_t^2|\bar{a}_{t-1}) - \sum_{t=1}^{m} H(a_t^1|a_t^2, \bar{a}_{t-1}) - \sum_{t=1}^{m} H(a_t^2|a_t^1, \bar{a}_{t-1}) \\
&\quad = H(\bar{a}_m) - \sum_{t=1}^{m} H(a_t^1|a_t^2, \bar{a}_{t-1}) - \sum_{t=1}^{m} H(a_t^2|a_t^1, \bar{a}_{t-1}) \\
&\leq I(s^1; s^2) + H(\bar{a}_m|s^1) + H(\bar{a}_m|s^2) - \sum_{t=1}^{m} H(a_t^1|a_t^2, \bar{a}_{t-1}) - \sum_{t=1}^{m} H(a_t^2|a_t^1, \bar{a}_{t-1}) \\
&\quad = I(s^1; s^2) + \sum_{t=1}^{m} \left[H(a_t|s^1, \bar{a}_{t-1}) - H(a_t^2|a_t^1, \bar{a}_{t-1})\right] \\
&\qquad + \sum_{t=1}^{m} \left[H(a_t|s^2, \bar{a}_{t-1}) - H(a_t^1|a_t^2, \bar{a}_{t-1})\right] \leq I(s^1; s^2),
\end{aligned}$$

where the last inequality is explained as follows. a_t^1 is a function of $\bar{a}_{t-1}$ and s^1. On the one hand, it implies that $H(a_t^2|s^1, \bar{a}_{t-1}) \leq H(a_t^2|a_t^1, \bar{a}_{t-1})$. On the other hand, combined with a_t^3 being a function of $\bar{a}_{t-1}$ it implies that $H(a_t|s^1, \bar{a}_{t-1}) = H(a_t^2|s^1, \bar{a}_{t-1})$. Therefore, $H(a_t|s^1, \bar{a}_{t-1}) \leq H(a_t^2|a_t^1, \bar{a}_{t-1})$, and similarly when switching between 1 and 2. □

Proof of Lemma 4.1. Let $\rho \in X_A$ be any strategy of Alice. Let $r = (s^1, s^2, h) \in \Sigma^1_{m+k} \times \Sigma^2_{m+k} \times A^k$ be Alice's random strategy, i.e., a random variable distributed according to ρ.

Let Bob's response to ρ be the strategy $s^3 \in \Sigma^3_m$ given by Lemma 4.3 applied to the continuation strategies $(s^1_{|h}, s^2_{|h})$. Recalling (4.1) and (4.2), the payoff in Γ is the expectation of the average m-stage payoff induced by the three strategies $s^1_{|h}, s^2_{|h}, s^3$. Therefore,

$$\Gamma(\rho, s^3) \geq \min\max{}_3 G - \sqrt{\frac{2I(s^1_{|h}; s^2_{|h})}{m}}.$$

By the chain rule of mutual information,

$$\begin{aligned} I(s^1_{|h}; s^2_{|h}) \leq I(s^1, h; s^2, h) &= I(s^1; s^2, h) + I(h; s^2, h|s^1) \\ &= I(s^1; s^2) + I(s^1; h|s^2) + I(h; s^2, h|s^1) \leq 2k \ln |A| \,, \end{aligned}$$

where the last inequality holds since s^1 and s^2 are independent, and $H(h) \leq k \ln |A|$. It follows that

$$\Gamma(\rho, s^3) \geq \min\max{}_3 G - 2\sqrt{\frac{k \ln |A|}{m}}.$$

□

4.2 The maximising strategy

We now return to the repeated game of Theorem 2.1. Assume w.l.o.g. that k_1 is as large as k_2, and denote $k = k_1 = k_2$. For now let m be roughly equal to k_3. We give the exact value of m in Section 4.2.2.

For any pair of mixed strategies $\sigma^i \in \Sigma^i(k)$ $(i = 1, 2)$ we describe a strategy $\sigma^3 \in \Sigma^3(m)$ that achieves the required expected payoff against these σ^1 and σ^2. Note that σ^3 will in fact be a mixed strategy. Although, of course, the existence of a good mixed response σ^3 implies the existence of a good pure response s^3, our proof does not single out such an s^3.

Consider the T-fold repeated game $G^T[k, k, m]$. We assume first that T is either a multiple of m^3 or $T = \infty$. The other values of T are treated later. For now, let us just hint that the case of $T < m^3$ is relatively simpler, and that any finite T can be divided into $T = T_1 + T_2$; where T_2 is a multiple of m^3 and $T_1 < m^3$.

We divide the stages of the repeated game into blocks of size m. For any block, let $h \in A^k$ be the last k actions played before that block, and consider the random continuation strategies $s^1_{|h}$ and $s^2_{|h}$. Although s^1 and s^2 are independent, $s^1_{|h}$ and $s^2_{|h}$ need not be (not even conditional on h or on the memory of Player 3), because there may be some interdependence between s^1, s^2 and h. Player 3, having a finite recall, may not know exactly what this interdependence is since the joint distribution of s^1, s^2 and h may differ from one block to the other. But now consider the corresponding auxiliary game $\Gamma_{\sigma^1,\sigma^2,k,m}$. The point is that in Γ, being a zero-sum game, there is a (possibly mixed) optimal strategy ζ^* of Bob, that guarantees the value against anything in X_A, namely against any possible distribution of s^1, s^2 and h.

That is very well for one block. Had 3 acted exactly the same in every block, s^1 and s^2 may have been able to learn something about this along the game. And 3 cannot play infinitely many *independent* copies of ζ^*, as we did not allow 3's strategies to be behavioral. Nevertheless, we show that it is sufficient that 3 plays a long period of independent copies cyclically.

Thus, the mixed strategy σ^3 is defined as follows. Let $z_1, \dots, z_{m^2}$ be i.i.d. variables taking values in Σ^3_m, with distribution $\zeta^* \in \Delta(\Sigma^3_m)$. During any block $B_i = ((i-1)m+1, \dots, im)$, 3 plays according to $z_i := z_{i \bmod m^2}$.

We examine the play inside any block B_i. Denote the last k periods of play before B_i by h_i. Denote the realisations of σ^1 and σ^2 by s^1 and s^2 respectively. Since s^1 and s^2 are k-recall strategies, the play during B_i is induced by $s^1_{|h_i}$, $s^2_{|h_i}$ and z_i. Furthermore, we only care about how s^1 and s^2 behave in the first $k+m$ periods. Denote the restriction of each s^j to $A^{<k+m}$ by $s'^j \in \Sigma^j_{k+m}$ $(j = 1, 2)$.

Let us now analyse the average per-stage payoff r^3 that 3 receives in m^2 consecutive blocks, say $B_1, B_2, \dots, B_{m^2}$. The analysis is made by taking a random variable $\hat{i}$ uniformly distributed on $[m^2]$ independently of $\sigma^1, \sigma^2, \sigma^3$ and estimating the expectation of the average per-stage payoff in $B_{\hat{i}}$.

Let $(\rho, \zeta) \in \Delta(\Sigma^1_{k+m} \times \Sigma^1_{k+m} \times A^k \times \Sigma^3_m)$ be the joint distribution of $(s'^1, s'^2, h_{\hat{i}}, z_{\hat{i}})$, where ρ is the joint distribution of $(s'^1, s'^2, h_{\hat{i}})$ and $\zeta = \zeta^*$ is the distribution of $z_{\hat{i}}$. Since ρ is a possible strategy for Alice in the auxiliary game (i.e., $\rho \in X_A$), and ζ^* is optimal for Bob,

$$\Gamma_{\sigma^1,\sigma^2,k,m}(\rho \otimes \zeta) \geq Val\,\Gamma_{\sigma^1,\sigma^2,k,m} \geq \min\max{}_3\, G - 2\sqrt{\frac{k \ln |A|}{m}}.$$

We regard the games played at each block $B_1, B_2, \dots, B_{m^2}$ as stages of an m^2-fold repeated meta game. Recall that r^3 is the expected average per-stage payoff of the meta game which is also the expected payoff in $B_{\hat{i}}$. Introduce a third player to Γ who has only one action and whose payoff is the same as

Bob's. By Lemma 4.2 applied to the that modification of Γ,

$$r^3 = \Gamma_{\sigma^1,\sigma^2,k,m}(\rho,\zeta) \geq \Gamma_{\sigma^1,\sigma^2,k,m}(\rho \otimes \zeta) - \sqrt{2I(s'^1, s'^2, h_{\hat{i}}; z_{\hat{i}})}.$$

By Neyman-Okada Lemma (Inequality 3.1), since each h_i is a function of (s'^1, s'^2, h_1) and $z_1, \ldots, z_{i-1}$,

$$\begin{aligned} I(s'^1, s'^2, h_{\hat{i}}; z_{\hat{i}}) &\leq \frac{1}{m^2} I(s'^1, s'^2, h_1; z_1, \ldots z_{m^2}) \\ &= \frac{1}{m^2}\left(I(s'^1, s'^2; z_1, \ldots z_{m^2}) + I(h_1; z_1, \ldots z_{m^2} | s'^1, s'^2)\right) \\ &= \frac{1}{m^2} I(h_1; z_1, \ldots, z_{m^2} | s'^1, s'^2) \leq \frac{k \ln |A|}{m^2}. \end{aligned}$$

It follows that

$$r^3 \geq \min\max{}_3 G - 2\sqrt{\frac{k \ln |A|}{m}} \left(1 + \frac{1}{\sqrt{2m}}\right).$$

4.2.1 Other values of T

If T is finite and not a multiple of m^3, let $T = T_1 + T_2 + T_3$, where $T_1 + T_2 < m^3$, and m^3 divides T_3; $T_1 < m$, and m divides T_2. During the last T_3 stages, σ^3 is defined as above, and the analysis is unaffected.

During the first T_1 stages, σ^3 can simply play perfectly against (σ^1, σ^2). By Lemma 4.3, there is a strategy $s^3 \in \Sigma^3_{T_1}$ that yields an expected average payoff of at least $\min\max_3 G$ during these stages, since σ^1 and σ^2 are independent. Therefore, a perfect play yields at least that much.

The next T_2 stages are divided into blocks of length m, and an independent copy of ζ^* is played for each block. Namely, Let $z_1, \ldots, z_{T_2/m}$ be i.i.d. variables taking values in Σ^3_m, with distribution ζ^*. During each block B_i, σ^3 plays according to z_i.[8] As above, the optimality of ζ^* implies that the expected average payoff in each B_i is $\geq \min\max_3 G - 2\sqrt{\frac{k \ln |A|}{m}}$.

Overall, the expected average payoff is at least

$$\min\max{}_3 G - 2\sqrt{\frac{k \ln |A|}{m}} \left(1 + \frac{1}{\sqrt{2m}}\right)$$

during the last T_3 stages, and we have a better bound for the first $T_1 + T_2$ stages.

[8]Actually, if one wanted to prove Theorem 2.1 only for small values of T, for example $T < m^3$, the proof could have been significantly simpler and did not need to go through the auxiliary game.

4.2.2 Final adjustments

Strictly speaking, although the above strategy σ^3 always focuses on one block of length m, it need not be a k_3-recall strategy. To make sure that it is, we now make small modifications to σ^3, and show that their effect on the expected payoff is small.

Since the strategy σ^3 cannot rely on the time t, we will make sure that the strategy always "knows where we are", by making it play some predefined actions at some stages. Dividing T into three phases of length T_1, T_2 and T_3 as above, we need to take care of three things: knowing the index of the current block during the second phase, knowing the index modulo m^2 during the third phase, and knowing where a block begins. During the first phase the history is shorter than m, therefore we know exactly where we are.

Assume w.l.o.g. that $|A_3| \geq 2$. Let $\gamma \in A_3$ be some action of Player 3. Denote $a = \lfloor \sqrt{m} \rfloor$, $b = \lceil \log_{|A_3|}(2m^2) \rceil$. The size of a block, m, is taken as the maximal numbers such that $k_3 \geq m + \max\{a, b\}$.

Every block B_i begins with $a+1$ stages in which 3 first plays γ, and then plays some fixed action different than γ for a stages. Denote this sequence of $a+1$ actions by $\bar{\alpha}$. This is followed by a "counter" $\bar{\beta}_i$ which designates the current phase (second or third) plus the block index (absolute index in the second phase and index modulo m^2 in the third). This counter has at most $2m^2$ different possible values, therefore it requires b stages.

The choice of m ensures that σ^3 is a k_3-recall strategy, since at any point in time we can see, within the previous k_3 stages, the last completed $\bar{\alpha}$ and the last completed counter.

During the rest of the block we play normally, except that we play the action γ every a stages. This makes sure that we can find the $\bar{\alpha}$ designating the beginning of a block, because $\bar{\alpha}$ contains a consecutive stages without γ.

The only modification needed in the proof is modifying the definition of the auxiliary game Γ to that of the game $\Gamma(\bar{\alpha}, \bar{\beta}_i, \gamma)$, defined the same except that the strategies of Bob are restricted to play $\bar{\alpha}$ in the first a stages, $\bar{\beta}_i$ in the following b_i stages, and then γ every a stages. Elsewhere, a strategy is free to choose anything, as before.

Otherwise the proof proceeds as above, and the analysis of the "free" stages is unaffected. The payoff in the predetermined stages may of course be low (recall that the payoff is always between 0 and 1). Therefore, during any block we get the average payoff we got above, minus at most $\frac{1}{m}((a+1) + b + m/a) \simeq \frac{1}{m}(2\sqrt{m} + log_{|A_3|}(2m^2))$.

5 Extensions

We considered repeated games in which the payoff was the un-discounted average of the stage payoffs, or the limiting average in the case of infinite repetition. It is easily verified that the asymptotic form of our result would still hold for a discounted payoff, when the discount rate approaches 0.

Suppose we allowed the strategies of Players 1 and 2 to depend not only on their recall, but also on the time t. Our result holds in this model too, with σ^3 unaltered (in particular, σ^3 need not even rely on t). The reason is that the point in the proof where the complexity limitations of these players was exploited was only this: the continuation strategy of i at any point in time depends only on the last k_i actions. This is true here as well.

5.1 Finite automata

Finite automata are another common model of bounded complexity in repeated games. An *automaton* of player i is a tuple $\mathcal{A} = \langle Z, z_0, q, f\rangle$. Z is a finite set, whose elements are called the *states* of $\mathcal{A}$. $z_0 \in Z$ is the *initial state*. $q : Z \times A^{-i} \to Z$ is the *transition function*. $f : Z \to A_i$ is the *action function*.

$\mathcal{A}$ induces a strategy in the repeated game as follows. Let $z_t \in Z$ denote the state of the automaton at stage t. Before the game begins the state is the initial state z_0. The transition from one state to the next is determined by the current state and the actions of the other players, i.e., $z_{t+1} = q(z_t, a_t^{-i})$. At stage t, the strategy plays the action $f(z_t)$.

The complexity of a strategy is measured by the *size* (i.e., the number of states) of the smallest automaton that implements this strategy. Any m-recall strategy is implementable by an $|A|^m$-automaton (but not vice versa), simply by letting each state of the automaton correspond to a different possible recall.

If we allowed the strategies of Players 1 and 2 to be implementable by automata of size $|A|^{k_i}$, instead of k_i-recall strategies, the result still holds, with σ^3 unaltered. The reason is, again, that the continuation strategy of i at any point in time depends only on a limited source of information: the last k_i actions is now replaced by the current state of the automaton. As the automaton has only $|A|^{k_i}$ possible states, we get exactly the same information-theoretic inequalities.

On the other hand, since σ^3 is implementable by an automaton of size $|A|^{k_3}$, we get the following theorem, a counterpart of Theorem 2.1 for finite automata.

Theorem 5.1. *There exists a constant $C > 0$ such that for every finite three-person game $G = \langle A, g \rangle$ and every $s_3 \geq s_2 \geq s_1 \geq 0$ and $T \in \mathbb{N} \cup \{\infty\}$,*

$$\min\max{}_3 G^T(s_1, s_2, s_3) \geq \min\max{}_3 G - C\sqrt{\frac{\log s_2 \ln |A|}{\log s_3}}$$

where $G^T(s_1, s_2, s_3)$ denotes the un-discounted T-fold repetition of G where each player i is restricted to an s_i-automaton.

We also note that the above argument still holds if we allow for automata with stochastic transitions, i.e., whose transition functions have the form $q : Z \times A^{-i} \to \Delta(Z)$.

5.2 Many players

We believe that our result, stated about the minmax in a 3-player repeated game, can be extended to any number of players, which is the next item on our agenda. Note that our proof used the convenient notion of mutual information. This notion has no canonical extension to more than two random variables, hence the proof would require some care.

References

Robert J. Aumann. Survey of repeated games. In *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, pages 11–42. Bibliographisches Institut, Mannheim, 1981.

Gilad Bavly and Abraham Neyman. Online concealed correlation and bounded rationality. *Games and Economic Behavior*, 88:71–89, 2014.

Elchanan Ben-Porath. Repeated games with finite automata. *Journal of Economic Theory*, 59(1):17–32, February 1993.

Ehud Lehrer. Repeated games with stationary bounded recall strategies. *Journal of Economic Theory*, 46(1):130–144, October 1988.

Ron Peretz. Correlation through bounded recall strategies. *International Journal of Game Theory*, 42(4):867–890, 2013.