

Towards a theory of repeated games with bounded recall

Gilad Bavly Ron Peretz*

October 18, 2016

Abstract

We study repeated games in which each player i is restricted to (mixtures of) strategies that can recall up to k_i stages of history. Characterising the set of equilibrium payoffs boils down to identifying the individually rational level of each player.

In contrast to the classic folk theorem, in which players have unrestricted recall, punishing a bounded-recall player may involve correlation between the punishers' actions. We quantify this correlation in terms of the average per-stage mutual information between the punishers' actions conditioned on the punishee's recalled history. The amount of correlation is proportional to the ratio between the recall capacities of the punishers and the punishee times the logarithm of the number of actions in the stage game. Our result extends to an arbitrary number of players and to other models of bounded memory.

Keywords: repeated games, bounded complexity, bounded recall, concealed correlation.

JEL classification: C72, C73.

1 Introduction

Bounded memory is a central models in Game Theory when it comes to modeling agents with bounded cognitive abilities (see e.g., Aumann (1981); Abreu and Rubinstein (1988); Neyman (1997)). Kalai (1990) posed four fundamental question regarding repeated games with bounded memory. The first question was

*Bar Ilan University, Israel.

“what are the possible outcomes of strategic games if players are restricted to (or choose to) use ‘simple’ strategies?”

This question has been attempted by many authors through the years, but a complete answer is yet unknown. A few advances include: Aumann and Sorin (1989); Neyman (1985); Lehrer (1988); Ben-Porath (1993); Neyman and Okada (2009); Neyman and Spencer (2010); Mailath and Olszewski (2011); Barlo et al. (2016).

Most of the past results concern two-person games. Until recently little was known about games with more than two players. The case of three players (or more) is inherently more difficult than that of two players. Peretz (2013) and Bavly and Neyman (2014) shed some light on games with three (or more) players. This paper makes further progress on three-player games which we believe is significant and can inspire future progress on Kalai’s question and repeated games in general.

Before we get to the state of the art, let us first brief through some previously known facts about repeated games in which the players are restricted to strategies of bounded complexity. In such games the characterisation of the equilibrium payoffs boils down to identifying the individually rational (minmax) levels of the players (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 approximate equilibrium payoff (strict equilibrium in games with a full dimensional feasible set).

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 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); Peretz (2012)). When there are more than two players, the difficulty in evaluating the minmax lies in the possibility of correlation between a group of players.

For illustration, let us consider a one-stage three-person Matching Pennies game, in which Player 3’s payoffs are

$$\begin{array}{c}
 \begin{array}{cc} & \text{L} & \text{R} \\ \text{T} & \boxed{-1} & \boxed{0} \\ \text{B} & \boxed{0} & \boxed{0} \end{array} & & \begin{array}{cc} & \text{L} & \text{R} \\ \text{T} & \boxed{0} & \boxed{0} \\ \text{B} & \boxed{0} & \boxed{-1} \end{array} . \\
 \text{W} & & \text{E}
 \end{array} \tag{1.1}$$

Player 3's minmax level is $-\frac{1}{4}$, which is obtained when 1 and 2 play

$$p = \begin{array}{|c|c|} \hline \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{4} \\ \hline \end{array}.$$

Now consider a repetition of the one-stage game (finitely or infinitely many times). Although the strategy space of the repeated game may be quite complicated, 3's minmax level remains $-\frac{1}{4}$. The explanation is that conditioned on any (finite) history, the actions of 1 and 2 right after that history are independent; therefore, 3, who observes the history, need only respond to product distributions at any period.

But now suppose the players have bounded recall, and therefore, they do not observe the entire history. Say, both 1 and 2 have recall capacity k and 3 has recall capacity m . In Peretz (2013) it was shown that (up to an approximation) 1 and 2 can implement a $2k$ - (but not more than $2k$) periodic sequence of actions whose period consists of $2k$ independent repetitions of a correlated action profile

$$c = \begin{array}{|c|c|} \hline \frac{1}{2} & 0 \\ \hline 0 & \frac{1}{2} \\ \hline \end{array}.$$

If $m < 2k$, then an agent who observes only the last m actions of 1 and 2 faces the correlated action c at every period; therefore 3's best response conditioned on the last m actions of 1 and 2 ensures him only $-\frac{1}{2}$ (3's correlated-minmax level).

But, things are more complicated than that. Player 3 observes not just the actions of 1 and 2, but 3's own actions as well. By playing a certain pattern of actions, 3 can encode information regarding the past actions of 1 and 2, and then, when these actions repeat, 3 can predict the next move of 1 and 2.

Determining the value that 3 can defend in such a way is a delicate matter. One needs to quantify the amount of information that 3 can encode while maintaining a given payoff level. Frankly, we do not know how to do it, but even if we knew, it would not suffice for computing the minmax level. The problem is that 3's actions may send information to 1 and 2, who could use that information to enhance the correlation between their actions. Paradoxically, increasing m allows 3 to send more information to 1 and 2, who may use it against 3, and thus decrease 3's minmax level. Of course, such a paradox can be easily avoided if 3 disregards some of her recall capacity. Nonetheless, that paradox sets an obstacle in the way to proving that large m (relative to k) yields high minmax level (for 3).

Our main technical contribution is devising and analysing a strategy for 3 that allows her to exploit the limited recall capacities of 1 and 2 while not revealing any information that might help 1 and 2 to correlate against her. We measure the correlation between 1 and 2 by average per-stage mutual information between their actions given the history recalled by 3. Mutual information¹ is a useful measure of correlation between two random variables. Independent actions such as p have 0 bits of mutual information, whereas fully correlated actions such as c have 1 bit of mutual information. Any convex combination between p and c has a fraction of a bit of mutual information, which increases continuously as the combination moves towards c . Theorem 2.1 establishes that the average per-stage mutual information between the actions of 1 and 2 can be bounded above by $C \frac{k}{m}$. Where C , a number that depends on the number of pure actions in the one-stage game, is determined up to a constant multiplicative factor.

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,² it turns out they cannot be so far from independent, depending on their recall capacity (or memory size, more generally).

Supposing the third player uses an m -recall strategy, perhaps she can exploit this observation during the following m stages or so. However, this 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 compute the value of this game. 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*. We utilise Neyman-Okada’s Lemma, a potent tool from the theory of two-person zero-sum games with bounded complexity, to show that it suffices to play a long cycle consisting of independent such plays.

It remains to compute the value of the auxiliary game. That value is

¹See definition in Section 3.

²For more on this see Section 4.2.

shown to be the vexification of the h -correlated minmax value the one-stage game. The h -correlated minmax value can be described as the value of a two-stage sequential extension of the one-stage game. First, 1 and 2 choose a correlated action whose mutual information is at most h , and then 3 replies the mixed joint action of 1 and 2 (with a pure action, w.l.o.g.).

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 anything beyond the correlated 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 $G = \langle A = A_1 \times A_2 \times A_3, g : A \rightarrow [0, 1]^3 \rangle$. I.e., it is assumed that the payoffs are scaled³ between 0 and 1, namely, $0 \leq g^i(a) \leq 1$, for all $a \in A$ and $i \in \{1, 2, 3\}$. The minmax-value of player $i \in \{1, 2, 3\}$ is defined as

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

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

$$\text{cor min max}_i G := \min_{x^{-i} \in \Delta(A_{-i})} \max_{a^i \in A_i} g^i(x^{-i}, a^i),$$

where $A_{-i} := \prod_{j \neq i} A_j$.

Between the minmax and the correlated minmax-values there is a range of intermediate values. We define the h -correlated minmax-value of player $i \in \{1, 2, 3\}$ ($h \geq 0$) as

$$\text{cor min max}_i G(h) := \min_{\substack{x^{-i} \in \Delta(A_{-i}) \\ \sum_{j \neq i} H(x^j) - H(x^{-i}) \leq h}} \max_{a^i \in A_i} g^i(x^{-i}, a^i),$$

where $H(\cdot)$ is Shannon's entropy function.

³This 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.

The h -correlated minmax-value is a continuous non-increasing function of h . For $h = 0$, it is equal to the (uncorrelated) minmax-value. For h large enough (e.g., $h = \min_{j \neq i} \{\ln |A_j|\}$), it reaches its minimum which is equal to the correlated minmax-value.

Our main result uses the convexification of the h -correlated minmax-value which is defined as follows: for a bounded function $f: D \rightarrow \mathbb{R}$ defined on a convex set $D (\subset \mathbb{R})$, the *convexification* of f is the largest convex function below f . Namely,

$$(Vex f)(h) := \sup\{c(h) : c: D \rightarrow \mathbb{R}, c \text{ is convex}, c(x) \leq f(x) \forall x \in D\}.$$

For $T \in \mathbb{N} \cup \{\infty\}$, a (pure) strategy for player $i \in \{1, 2, 3\}$ in the T -fold repeated game is a function $s^i : A^{<T} \rightarrow A_i$, where $A^{<T} = \bigcup_{0 \leq t < T} A^t$.

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 strategies for player i is denoted Σ_T^i . For a strategy s^i and a history of play $h_t = (a_1, \dots, a_t) \in A^t$, 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., $s^i_{|h_t}(a'_{t+1}, \dots, a'_{t+r}) = s^i(a_1, \dots, a_t, a'_{t+1}, \dots, a'_{t+r})$, for all $(a'_{t+1}, \dots, a'_{t+r}) \in A^r$.

A *k-recall* strategy for player i is a strategy $s^i \in \Sigma_\infty^i$ that depends only on the last k periods of history. Namely, for any two histories of any length $\bar{a} = (a_1, \dots, a_{m-1})$ and $\bar{b} = (b_1, \dots, b_{n-1})$, if $(a_{m-k}, \dots, a_{m-1}) = (b_{n-k}, \dots, b_{n-1})$ 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 is the average per-stage payoff, for $T < \infty$, 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. *For every $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that for every finite*

three-person game $G = \langle A, g \rangle$ and every $k_3 \geq k_2 \geq k_1 \geq k_0$ and $T \in \mathbb{N} \cup \{\infty\}$,

$$\begin{aligned} \min \max_3 G^T[k_1, k_2, k_3] & \\ & \geq (\text{Vex cor min max}_3 G) \left(2 \ln |A| \cdot \frac{k_2}{k_3} \right) - \epsilon \\ & \geq \min \max_3 G - 2 \sqrt{\ln |A| \cdot \frac{k_2}{k_3}} - \epsilon . \end{aligned}$$

3 Preliminaries

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

Shannon's entropy⁴ of a discrete random variable x is the following non-negative quantity

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

where $0 \ln 0 = 0$ by continuity.

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

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

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:

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

Mutual information is a useful measure of interdependence between a pair of random variables. Another useful measure of interdependence is the norm distance between the joint distribution and the product of the two marginal distributions. A relation between these measures is given by *Pinsker's Inequality*:

$$\|p(x, y) - p(x) \otimes p(y)\|_1 \leq \sqrt{2I(x; y)}.$$

⁴In the literature, a similar definition using \log_2 instead of \ln is also commonly referred to as "Shannon's entropy."

3.1 Neyman-Okada Lemma

In a sequence of papers, Neyman and Okada (Neyman and Okada, 2000, 2009; Neyman, 2008) developed a methodology for analysing repeated games with bounded memory. A key idea of theirs is captured in the following lemma.

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

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

The interpretation is that x_1, \dots, x_m is a sequence of actions played by an oblivious player⁵, y_0 is a random strategy of another player, and y_1, \dots, 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, \dots, x_m are i.i.d. . In this case we have

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

4 Proof of Theorem 2.1

The second inequality in Theorem 2.1 is an immediate corollary of Pinsker's inequality. The payoff function g^3 is 1-Lipschitz w.r.t. to the $\|\cdot\|_1$; therefore, by Pinsker's inequality,

$$\text{cor min max}_3 G(h) \geq \min \max_3 G - \sqrt{2h}, \quad \forall h \geq 0,$$

and the function on the right-hand side is convex.

The main effort is proving the first inequality 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

⁵An oblivious player is one who ignores the actions of the other players.

strategies of 1 & 2 need not be independent, not even conditional⁶ on the memories of 1 & 2.

This leads us to define and analyse 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 $\sigma^i \in \Delta(\Sigma_{m+k}^i)$ ($i = 1, 2$), 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

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

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

The strategies of Alice can also be described as follows. A pair of strategies $s^1 \in \Sigma_{m+k}^1$ and $s^2 \in \Sigma_{m+k}^2$ 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 $r = (s^1, s^2, h) \in \Sigma_{m+k}^1 \times \Sigma_{m+k}^2 \times A^k$ and $z \in \Sigma_m^3$ induces a play a_1, \dots, a_m of players 1, 2, 3, defined by

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

for any $1 \leq t \leq m$. 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_h^i).

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 by

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

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 $\sigma^1 \in \Delta(\Sigma_{k+m}^1)$ and $\sigma^2 \in \Delta(\Sigma_{k+m}^2)$,*

$$\text{Val}(\Gamma_{\sigma^1, \sigma^2, k, m}) \geq (\text{Vex cor min max}_3 G) \left(\frac{2k \ln |A|}{m} \right).$$

⁶We further elaborate on this point in Section 4.2.

The rest of this section is devoted to proving Lemma 4.1. Our next lemma states that the convexification of the h -correlated minmax-value of G is at most the mh -correlated minmax-value of the m -fold repetition G^m .

Lemma 4.2. *Let s^1 and s^2 be random strategies that assume values in Σ_m^1 and Σ_m^2 , respectively. There exists a pure strategy $s^3 \in \Sigma_m^3$ such that the play a_1, \dots, a_m induced by (s^1, s^2, s^3) satisfies*

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

Proof. The strategy $s^3 \in \Sigma_m^3$ 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 $\bar{a}_{t-1} = (a_1, \dots, a_{t-1}) \in A^{t-1}$ and random actions for 1 and 2 at time t , a_t^1 and a_t^2 . We define s^3 on A^{t-1} by choosing

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

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 the definition of s^3 ,

$$\mathbb{E}[g^3(a_t) | \bar{a}_{t-1}] \geq \text{cor min max}_3 G(Y(\bar{a}_{t-1})),$$

for every $t \in [m]$.

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,

$$\begin{aligned} \frac{1}{m} \sum_{t=1}^m \mathbb{E}[g^3(a_t)] &= \mathbb{E}[g^3(a_{\hat{t}})] = \mathbb{E}[\mathbb{E}[g^3(a_{\hat{t}}) | \bar{a}_{\hat{t}-1}]] \\ &\geq \mathbb{E}[\text{cor min max}_3 G(Y)] \geq (\text{Vex cor min max}_3 G)(\mathbb{E}[Y]). \end{aligned}$$

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

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

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. \square

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

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

$$\Gamma(\rho, s^3) \geq (\text{Vex cor min max}_3 G) \left(\frac{2I(s_{|h}^1; s_{|h}^2)}{m} \right).$$

By the chain rule of mutual information,

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

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

$$\Gamma(\rho, s^3) \geq (\text{Vex cor min max}_3 G) \left(\frac{2k \ln |A|}{m} \right).$$

□

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 $\sigma^i \in \Delta(\Sigma^i(k))$ ($i = 1, 2$) we describe a strategy $\sigma^3 \in \Delta(\Sigma^3(m))$ 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 $s_{|h}^1$ and $s_{|h}^2$. Although s^1 and s^2 are independent, $s_{|h}^1$ and $s_{|h}^2$ 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 behavioural. 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_m^3 , with distribution $\zeta^* \in \Delta(\Sigma_m^3)$. 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 $s^j \in \Sigma_{k+m}^j$ ($j = 1, 2$).

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 $(\rho, \zeta) \in \Delta(\Sigma_{k+m}^1 \times \Sigma_{k+m}^1 \times A^k \times \Sigma_m^3)$ 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,

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

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 $B_{\hat{i}}$. Since Bob's payoff function in $\Gamma_{\sigma^1, \sigma^2, k, m}$ is 1-Lipschitz, by Pinsker's inequality,

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

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

$$\begin{aligned} I(s^1, s^2, h_{\hat{i}}; z_{\hat{i}}) &\leq \frac{1}{m^2} I(s^1, s^2, h_1; z_1, \dots, z_{m^2}) \\ &= \frac{1}{m^2} (I(s^1, s^2; z_1, \dots, z_{m^2}) + I(h_1; z_1, \dots, z_{m^2} | s^1, s^2)) \\ &= \frac{1}{m^2} I(h_1; z_1, \dots, z_{m^2} | s^1, s^2) \leq \frac{k \ln |A|}{m^2} \leq \frac{\ln |A|}{k_0}. \end{aligned}$$

It follows that

$$r^3 \geq (\text{Vex cor min max}_3 G) \left(\frac{2k \ln |A|}{m} \right) - \sqrt{2 \ln |A| / k_0}.$$

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 4.2, there is a strategy $s^3 \in \Sigma_{T_1}^3$ 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 $z_1, \dots, z_{T_2/m}$ be i.i.d. variables taking values in Σ_m^3 , with distribution ζ^* . During each block B_i , σ^3 plays according to z_i .⁷ As above, the optimality of ζ^* implies that the expected average payoff in each B_i is $\geq (\text{Vex cor min max}_3 G) \left(\frac{2k \ln |A|}{m} \right)$.

Overall, the expected average payoff is at least

$$(\text{Vex cor min max}_3 G) \left(\frac{2k \ln |A|}{m} \right) - \sqrt{2 \ln |A| / k_0}$$

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

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 $k_3 \geq m + \max \{a, b\}$.

⁷Actually, 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.

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 Players 1 and 2 to play mixtures of behaviour k_i -recall strategies. That is, a mixture of functions from A^{k_i} to $\Delta(A_i)$. Our result holds in this model too, with σ^3 unaltered (in particular, σ^3 need not toss coins). 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.

Another plausible variation of the model is allowing strategies to depend not only of the last k_i actions, but also on calendar time. Here, too, our result holds. The proof for this model is simpler, since we only have to consider the instance of the auxiliary game played at each block separately (Lemma 4.1).

We do not have to worry about being able to repeat a strategy indefinitely. To this end, it is crucial that Player 3 can condition her actions on calendar time. Otherwise, any fixed (0-recall) normal sequence of actions of players 1 & 2 would seem random in the eyes of Player 3.

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*. $q : Z \times A^{-i} \rightarrow Z$ is the *transition function*. $f : Z \rightarrow A_i$ 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. *For every $\epsilon > 0$ there exists $s_0 \in \mathbb{N}$ such that for every finite three-person game $G = \langle A, g \rangle$ and every $s_3 \geq s_2 \geq s_1 \geq s_0$ and $T \in \mathbb{N} \cup \{\infty\}$,*

$$\min \max_3 G^T(s_1, s_2, s_3) \geq (\text{Vex cor } \min \max_3 G) \left(\frac{2 \ln s_2 \ln |A|}{\ln s_3} \right) - \epsilon.$$

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 $q : Z \times A^{-i} \rightarrow \Delta(Z)$.

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.

5.3 An open question

Theorem 2.1 sets an asymptotic lower bound on the minmax value in the presence of bounded recall. A comparison with Peretz (2013) shows that our lower bound is of the correct order of magnitude, but it does not suggest that the bound is tight. Providing tight bounds for the minmax value (of three-person games) with bounded recall remains an open problem.

To pin-point the problem let us focus on the three-person Matching Pennies Game $G = \langle A, g \rangle$ in which Player 3's payoff function is given in (1.1). Does $\min \max_3 G^\infty[k, k, k]$ converge as $k \rightarrow \infty$, and, if so, what is the limit?

6 Acknowledgement

We thank Yuval Heller for helpful suggestions.

References

- Dilip Abreu and Ariel Rubinstein. The structure of nash equilibrium in repeated games with finite automata. *Econometrica*, 56(6):1259–1281, 1988.
- 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.
- Robert J Aumann and Sylvain Sorin. Cooperation and bounded recall. *Games and Economic Behavior*, 1(1):5–39, 1989.
- Mehmet Barlo, Guilherme Carmona, and Hamid Sabourian. Bounded memory folk theorem. *Journal of Economic Theory*, 163:728–774, 2016.

- 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 Kalai. Bounded rationality and strategic complexity in repeated games. In Tatsuhiro Ichiishi, Abraham Neyman, and Yair Tauman, editors, *Game theory and applications*, pages 131–157. Academic Press, San Diego, 1990.
- Ehud Lehrer. Repeated games with stationary bounded recall strategies. *Journal of Economic Theory*, 46(1):130–144, October 1988.
- George J Mailath and Wojciech Olszewski. Folk theorems with bounded recall under (almost) perfect monitoring. *Games and Economic Behavior*, 71(1):174–192, 2011.
- Abraham Neyman. Bounded complexity justifies cooperation in the finitely repeated prisoners’ dilemma. *Economics Letters*, 19(3):227–229, 1985.
- Abraham Neyman. Cooperation, repetition, and automata. In *Cooperation: Game Theoretic Approaches, NATO ASI Series F*, pages 233–255. Springer-Verlag, 1997.
- Abraham Neyman. Learning effectiveness and memory size. Discussion Paper 476, Center for the Study of Rationality, Hebrew University, Jerusalem, February 2008.
- Abraham Neyman and Daijiro Okada. Repeated games with bounded entropy. *Games and Economic Behavior*, 30(2):228–247, February 2000.
- Abraham Neyman and Daijiro Okada. Growth of strategy sets, entropy, and nonstationary bounded recall. *Games and Economic Behavior*, 66(1):404 – 425, 2009. ISSN 0899-8256.
- Abraham Neyman and Joel Spencer. Complexity and effective prediction. *Games and Economic Behavior*, 69(1):165–168, May 2010.
- Ron Peretz. The strategic value of recall. *Games and Economic Behavior*, 74(1):332 – 351, 2012. ISSN 0899-8256.
- Ron Peretz. Correlation through bounded recall strategies. *International Journal of Game Theory*, 42(4):867–890, 2013.