Towards a theory of repeated games with bounded memory

Gilad Bavly*† Ron Peretz*‡

December 27, 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 ("punishment level") of each player.

In contrast to the classic folk theorem, in which players are unrestricted, punishing a bounded player may involve correlation between the punishers' actions. We show that the extent of such correlation is at most proportional to the ratio between the recall capacity of the punishers and the punishee. Our result extends to an arbitrary number of players and to other models of bounded complexity.

Keywords: repeated games, bounded complexity, equilibrium payoffs, bounded recall, bounded automata, concealed correlation.

JEL classification: C72, C73.

*Bar Ilan University, Ramat Gan, Israel.

†gilad.bavly@gmail.com

‡ronprtz@gmail.com

1 Introduction

It has long been asserted that, in many economic contexts, not all courses of action are feasible (e.g., Simon (1955, 1972); Kalai (1990)). Many times it is reasonable to expect simple strategies to be employed, or at least strategies that are not immensely complex. The issue is clearly manifest in repeated games, as even a finite repetition gives rise to strategies that one may deem unrealistically complex.

In a survey of repeated games with bounded complexity, Kalai (1990) asked:

> "what are the possible outcomes of strategic games if players are restricted to (or choose to) use 'simple' strategies?"

This question has been considered in many works through the years. Naming a few notable results, we mention Abreu and Rubinstein (1988); Aumann and Sorin (1989); Neyman (1997); Gossner and Hernández (2003); Renault et al. (2007); Neyman and Okada (2009); Lehrer and Solan (2009); Mailath and Olszewski (2011); Barlo et al. (2016).

In this paper we give a lower bound for the equilibrium payoff of each player i, by bounding the amount of correlation that the other players can effectively achieve "against" i. This bound is formulated in terms of the (average per-stage) amount of correlation between the stage actions of these players.

Introduced by Aumann (1981), the two most common models of bounded complexity in repeated games are Bounded Automata and Bounded Recall.[1]

[1]Pioneering work is Neyman (1985); Rubinstein (1986); Ben-Porath (1993) on Automata and Lehrer (1988) on Bounded recall.

Both models involve putting some bounds on the memory that can be used by the players.

Our presentation here focuses on bounded recall, a simple model in which each player i has a recall capacity k_i, where i's strategy can rely only on the previous k_i stages. However, our main result (Theorem 2.1) also has a counterpart for finite automata (Theorem 5.1), as well as for some variants (see Section 5).

No further assumptions are made besides complexity bounds, e.g., there are no external communication devices, and monitoring is perfect.

In repeated games with bounded complexity, the characterisation of the equilibrium payoffs boils down to identifying the individually rational (minmax) levels of the players[2] (Lehrer, 1988, p. 137), which need not coincide with the one-stage minmax. That is, in a sufficiently long game, any payoff profile that is feasible and above each player's minmax, is close to an approximate equilibrium payoff (or simply to an equilibrium payoff, in games with a full dimensional feasible set). Thus, we can henceforth concentrate on the minmax.

The case of two players is well understood (Lehrer, 1988; Ben-Porath, 1993; Peretz, 2012). Little is known about the minmax when there are more than two players. The difficulty lies in the possibility of correlation between a group of players.

To illustrate, let us consider a one-stage three-person Matching Pennies

[2]A.k.a. "punishment levels".

game, in which Player 3's payoffs are

	L	R
T	-1	0
B	0	0

W

	L	R
T	0	0
B	0	-1

E . (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 reason 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.[3]

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 such as

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

[3]In other words, although their actions may be correlated, they are effectively independent w.r.t. player 3.

Thus, perhaps surprisingly, 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 $-\frac{1}{2}$ (which is 3's one-stage correlated-minmax level).

Things turn out to be more complicated. 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. And even if we knew this value, it would still 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. This also sets an obstacle in the way of proving that a large m (relative to k) yields a high minmax level[4] 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 them correlate against her. We measure the correlation between 1 and 2 by the average per-stage mutual information[5] between their actions given the history recalled by 3. Theorem 2.1

[4]As increasing m also "allows" 3 to send more information to 1 and 2, who may use it against 3.

[5]Mutual information (see definition in Section 3) is a useful measure of correlation between two random variables. Independent actions such as p above have 0 bits of mutual information, whereas fully correlated actions such as c have 1 bit of mutual information. Any convex combination of p and c has a fraction of a bit of mutual information, contin-

establishes that the correlation between 1 and 2 is at most $C\frac{k}{m}$, where C is a number that depends on the number of pure actions in the one-stage game.

1.1 Equilibrium payoffs with bounded complexity

Bounded complexity could bring about new equilibrium outcomes, that were not there in the unrestricted repeated game, and it could also exclude equilibrium outcomes. The set of equilibrium payoffs would still be folk-theorem-like, i.e., approximately consist of all feasible payoffs that are above each player's minmax. The difference from the unrestricted repeated game (i.e., the classic folk theorem), is that here this minmax may be different than the minmax of the one-stage game.

One could easily imagine a player so much stronger than the others, that her minmax is very high, thus shrinking the equilibrium payoffs; and vice versa for a weak player. Moreover, even when all players are of equal strength, the minmax could drop below the one-stage minmax (Peretz, 2013). Bavly and Neyman (2014) also "expand" the equilibrium payoffs, in a different case. Our result goes in the opposite direction, implying a lower bound for the minmax which is the one-stage minmax minus a function[6] of $\frac{k}{m}$.

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 (Lehrer, 1988, Theorem 3; Bavly and Neyman, 2014, Theorem 2.3). Therefore, our result closes a significant gap in the characterization of equilibrium payoffs.

---

uously increasing as the combination moves towards c.

[6] We actually show that this function is at most proportional to the square root of k/m. In particular, it is small when k/m is small.

The scale $\frac{k}{m}$ 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 one-stage correlated minmax.

The result is tight in another sense as well: the strongest player, unless she is extremely strong, cannot hope for more than her one-stage minmax. By Lehrer (1988), the minmax of a player who isn't exponentially stronger than the other players, is at most her one-stage 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.

1.2 A few notes about the proof

The following observation, perhaps of interest on its own right, 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, from some stage t on, need not be independent,[7] it turns out they cannot be too far from independent, due to the bounded memory.

Supposing the third player uses an m-recall strategy, perhaps she can exploit this fact 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 game as follows.

Let σ_1, σ_2 be any pair of mixed k-recall strategies of players 1 and 2. The auxiliary game is a zero-sum game between Bob (the maximizer) and Alice (the minimizer). It is conceived by imagining the play of the repeated game

[7] For more on this see Section 4.2.

during m consecutive stages, starting at some arbitrary point in time t. Bob, "representing" player 3, chooses a strategy to be played during these m stages against σ_1, σ_2. However, Alice, representing players 1 and 2, gets to choose what supposedly was the k-length history preceding stage t. Moreover, she can condition her choice on the *realization* of the strategies of 1 and 2.

The resulting artificial "continuation strategies" cannot be too far from independent, as already said. We then bound the average per-stage correlation, and deduce that player 3 (or rather, Bob) does well in the auxiliary game (skipping the details of how well).

The point of the auxiliary game is that, by a minmax theorem, there is a mixed *optimal* strategy for player 3, good against any choice of Alice. Therefore, this strategy is also good against σ_1, σ_2 during m stages of the original game, starting at *any* t.

With a bounded strategy, player 3 cannot employ this optimal strategy infinitely many times *independently*. But we show that it suffices to cyclically play a long cycle consisting of many independent such employments.

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[8] between 0 and 1. The minmax-value of player $i \in \{1,2,3\}$ is defined as

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

[8]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.

where $\Delta(X)$ denotes the set of probability distributions over a finite set X.

The correlated minmax-value[9] of player $i \in \{1,2,3\}$ is defined as

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

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

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

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

where $H(\cdot)$ is Shannon's entropy function.[10]

The h-correlated minmax-value is the value that player i can defend when the other two players are allowed to correlate their actions up to the level h. It 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 $f\colon D \to \mathbb{R}$ defined on a convex set D ($\subset \mathbb{R}$), the *convexification of* f is the largest convex function below f. Namely,

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

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 $s^i : A^{<T} \to A_i$, where $A^{<T} = \bigcup_{0\leq t<T} A^t$.

[9] A.k.a. the maxmin-value.

[10] The quantity $H(a) + H(b) - H(a,b)$ is called "mutual information" (see Section 3).

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 $h_t = (a_1, \ldots, a_t) \in A^t$, 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., $s^i_{|h_t}(a'_{t+1}, \ldots, a'_{t+r}) = s^i(a_1, \ldots, a_t, a'_{t+1}, \ldots, a'_{t+r})$, for all $(a'_{t+1}, \ldots, a'_{t+r}) \in A^r$.

A *k-recall* strategy for player i is a strategy $s^i \in \Sigma^i_\infty$ that depends only on the last k periods of history. 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 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 $T \in \mathbb{N} \cup \{\infty\}$,*

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

By Peretz (2012), if $\log k_3/k_1 \to 0$ then

$$\min\max_3 G^T[k_1, k_2, k_3] \leq \min\max_3 G + \mathrm{o}(1).$$

Therefore, Theorem 2.1 is tight in the case where k_3 is superlinear in k_2 yet subexponential in k_1.

We believe that Theorem 2.1 can be extended to more than three players; see Section 5.2.

3 Preliminaries

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

Shannon's entropy[11] 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.

[11] 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:

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

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*:

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

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.[12]

Let $x_1, \ldots, x_m, y_1, \ldots, y_m$ be finite random variables, and y_0 be a random variable such that each y_i is a function of $y_0, x_1, \ldots, x_{i-1}$. Suppose t is a random variable that distributes uniformly in $[m] := \{1, \ldots, m\}$ independently of $y_0, x_1, \ldots, x_m, y_1, \ldots, y_m$. 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,[13] 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

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

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. the $\|\cdot\|_1$ norm; therefore, by Pinsker's inequality,

$$\operatorname{cor} \min \max_3 G(h) \geq \min \max_3 G - \sqrt{2h}, \ \forall h \geq 0,$$

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

[12] For a proof see (Peretz, 2012, Lemma 4.2).

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

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 and 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 and 2 and of their memories at that point. Generally, the continuation strategies of 1 and 2 need not be independent, not even conditional[14] on the memories of 1 and 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 $\sigma^i \in \Delta(\Sigma^i_{m+k})$ $(i = 1, 2)$, 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 and 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

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

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 $1 \leq t \leq m$. That is, we look at an m-fold repeated game, in which Player 3 simply employs the strategy z, and Players 1 and 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 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 $\sigma^2 \in \Delta(\Sigma^2_{k+m})$,*

$$Val(\Gamma_{\sigma^1,\sigma^2,k,m}) \geq (Vex \operatorname{cor} \min\max_3 G)\left(\frac{2k \ln |A|}{m}\right).$$

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^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*

$$\mathbb{E}\left[\frac{1}{m}\sum_{t=1}^{m} g^3(a_t)\right] \geq (\mathit{Vex}\,\mathrm{cor}\min\max_3 G)\left(\frac{I(s^1;s^2)}{m}\right) .$$

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 $\bar{a}_{t-1} = (a_1, \ldots, a_{t-1}) \in A^{t-1}$ 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,

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

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,

$$\begin{aligned}\frac{1}{m}\sum_{t=1}^{m}\mathbb{E}[g^3(a_t)] = \mathbb{E}[g^3(a_{\hat{t}})] &= \mathbb{E}\left[\mathbb{E}[g^3(a_{\hat{t}})|\bar{a}_{\hat{t}-1}]\right] \\ &\geq \mathbb{E}[\mathrm{cor}\min\max_3 G(Y)] \geq (\mathit{Vex}\,\mathrm{cor}\min\max_3 G)\left(\mathbb{E}[Y]\right).\end{aligned}$$

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

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

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 $r = (s^1, s^2, h) \in \Sigma_{m+k}^1 \times \Sigma_{m+k}^2 \times A^k$ 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,

$$\Gamma(\rho, s^3) \geq (\mathit{Vex}\ \mathrm{cor} \min\max_3 G)\left(\frac{I(s_{|h}^1; s_{|h}^2)}{m}\right).$$

By the chain rule of mutual information,

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

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

$$\Gamma(\rho, s^3) \geq (\mathit{Vex}\, \mathrm{cor} \min \max_3 G) \left(\frac{2k \ln |A|}{m} \right) .$$

□

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 $s^1_{|h}$ 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 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 $z_1, \ldots, z_{m^2}$ be i.i.d. variables taking values in Σ^3_m, with distribution $\zeta^* \in \Delta(\Sigma^3_m)$. During any block $B_i = ((i-1)m+1, \ldots, im)$, 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 $B_1, B_2, \ldots, B_{m^2}$. 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,

$$\Gamma_{\sigma^1,\sigma^2,k,m}(\rho \otimes \zeta) \geq \mathit{Val}\, \Gamma_{\sigma^1,\sigma^2,k,m} \geq (\mathit{Vex}\, \mathrm{cor} \min \max_3 G)\left(\frac{2k \ln |A|}{m}\right).$$

We regard the games played at each block $B_1, B_2, \ldots, B_{m^2}$ 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 $z_1, \ldots, z_{i-1}$,

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

It follows that

$$r^3 \geq (\mathit{Vex}\, \mathrm{cor} \min \max_3 G)\left(\frac{2k \ln |A|}{m}\right) - \sqrt{2 \ln |A| / k_0}\ .$$

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 $s^3 \in \Sigma^3_{T_1}$ 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.[15] As above, the optimality of ζ^* implies that the expected average payoff in each B_i is $\geq (\mathit{Vex}\operatorname{cor}\min\max_3 G)\left(\frac{2k\ln|A|}{m}\right)$.

Overall, the expected average payoff is at least

$$(\mathit{Vex}\operatorname{cor}\min\max_3 G)\left(\frac{2k\ln|A|}{m}\right) - \sqrt{2\ln|A|/k_0}$$

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,

[15] 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.

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 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 (i.e., 0-recall) normal sequence of actions of players 1 and 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*. $q : Z \times A^{-i} \to Z$ 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 are 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 $T \in \mathbb{N} \cup \{\infty\}$,*

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

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

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.

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 $k \to \infty$, and, if so, what is the limit?

6 Acknowledgement

We thank Yuval Heller and Eilon Solan for helpful suggestions. G. Bavly acknowledges support from Sir Isaac Wolfson Chair in Economics and Business Administration, the Department of Economics in Bar-Ilan University, and ISF grant 1188/14.

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.

Olivier Gossner and Penélope Hernández. On the complexity of coordination. *Mathematics of Operations Research*, 28(1):127–140, 2003.

Ehud Kalai. Bounded rationality and strategic complexity in repeated games. In Tatsuro 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.

Ehud Lehrer and Eilon Solan. Approachability with bounded memory. *Games and Economic Behavior*, 66(2):995–1004, 2009.

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.

Ron Peretz. The strategic value of recall. *Games and Economic Behavior*, 74(1):332 – 351, 2012.

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

Jérôme Renault, Marco Scarsini, and Tristan Tomala. A minority game with bounded recall. *Mathematics of Operations Research*, 32(4):873–889, 2007.

Ariel Rubinstein. Finite automata play the repeated prisoner's dilemma. *Journal of economic theory*, 39(1):83–96, 1986.

Herbert A Simon. A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118, 1955.

Herbert A Simon. Theories of bounded rationality. *Decision and organization*, 1(1):161–176, 1972.