Toward a theory of repeated games with bounded memory

Gilad Bavly*† Ron Peretz*‡

May 22, 2017

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. Characterizing 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 a few variations of the model, as well as to finite automata.

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

JEL classification: C72, C73.

*Bar Ilan University, Ramat Gan, Israel.

†gilad.bavly@gmail.com

‡ron.peretz@biu.ac.il

1 Introduction

It has long been asserted that in many economic contexts, not all courses of action are feasible (e.g., Simon 1955, 1972). 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. To name 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), and Barlo et al. (2016).

In this paper we bound, for each player i, the amount of correlation that the other players can effectively achieve "against" i in a repeated game with bounded complexity. The bound is formulated in terms of the (average per-stage) amount of correlation between the stage actions of the players other than i. This upper bound on correlation implies a lower bound for the equilibrium payoff of each player i.

The two most common models of bounded complexity in repeated games are finite automata and bounded recall.[1] Both models involve setting bounds

[1]These models were introduced by Aumann (1981). Other pioneering works in this area include Neyman (1985), Rubinstein (1986), and Ben-Porath (1993) on automata, and

on the memory of the players.[2]

For clarity's sake, our presentation here focuses on the simple model of bounded recall, 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 also applies to finite automata (Theorem 5.1), as well as to some variants of the bounded recall model (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 characterization of the equilibrium payoffs boils down to identifying the individually rational (minmax) levels of the players[3] (Lehrer 1988, p. 137), which need not coincide with the individually rational levels of the one-stage game. 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 in a group of players. Even though the players employ uncorrelated mixed strategies at the beginning of the game, their actions can become correlated in the course

Lehrer (1988) on bounded recall.

[2]These bounds are not necessarily small, and hence bounded complexity does not imply that strategies are necessarily "simple" in everyday terms.

[3]Also known as "punishment levels."

of the game due to imperfect recall.[4,5]

To illustrate the role of correlation in our results, let us consider a one-stage three-person Matching Pennies 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 in any period.[6]

But suppose that the players have bounded recall and, therefore, they do not observe the entire history. For example, let us say that both 1 and 2

[4] The subtlety of games with imperfect recall was demonstrated by Piccione and Rubinstein (1997) even in the case of a single player.

[5] The possibility of correlation between players was demonstrated by Peretz (2013) who showed that for every $C > 0$, there is a game in which two players with recall capacities k can correlate their actions against a third player whose recall capacity is Ck, for k large enough.

[6] In other words, although the actions of 1 and 2 may be correlated, they are independent in the eyes of 3 (given his information about the entire history).

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 $2k$-periodic sequence of actions whose period consists of $2k$ independent repetitions of a correlated action profile such as

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

Therefore, perhaps surprisingly, even for m larger than k, as long as $m < 2k$ an agent who observes only the last m actions of 1 and 2 faces the correlated action c in 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).

Yet things turn out to be more complicated. Player 3 observes not just the actions of 1 and 2, but her own actions as well. By playing a certain pattern of actions, 3 can encode information about the past actions of 1 and 2, and when these actions are repeated, 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 she maintains a given payoff level. Furthermore, 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. Even if m is larger than k, 3 needs to use her advantageous information with care so as not to reveal it to 1 and 2.

Our main technical contribution is to devise and analyze 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 correlate against her. We measure the correlation between 1 and 2 (against 3) by the average per-stage mutual information[7] of their joint actions given the history recalled by 3. Theorem 2.1 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. In particular, if $m \gg k$ then 1 and 2 cannot correlate against 3; therefore, 3's minmax level is asymptotically at least her minmax level in the one-stage game.[8]

In Section 5.2 we extend our result to games with any number of players.

1.1 Equilibrium payoffs with bounded complexity

Bounded complexity may give rise to new equilibrium outcomes that were not present in the unrestricted repeated game, and it may also exclude equilibrium outcomes. The set of equilibrium payoffs would still be folk-theorem-like; i.e., it would consist of approximately 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 the minmax under bounded complexity may be different from the minmax of the one-stage game.

For example, consider a player who is stronger than the other players, i.e.,

[7] Mutual information (see Section 3 for definition) 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, which is continuously increasing as the combination moves toward c.

[8] If, in addition, $\log m \ll k$, then 3 cannot predict the actions of 1 and 2 (by Lehrer 1988); therefore 3's minmax is asymptotically equal to her one-stage minmax.

her recall capacity is larger than that of the others. It is not hard to show that if the difference in strength is very large then her minmax is high. This shrinks the set of equilibrium payoffs, compared to that set in the classic folk theorem; and conversely for a sufficiently weaker player.

Moreover, even when all players are of equal strength (i.e., have the same recall capacity), the minmax may drop below the one-stage minmax (Peretz 2013). Bavly and Neyman (2014) also "expand" the equilibrium payoffs (i.e., give an upper bound for a player's minmax) in a different case. Our result goes in the opposite direction, since our bound on the correlation between Players 1 and 2 implies a lower bound for the minmax of Player 3, which equals the one-stage minmax minus a function[9] of $\frac{k}{m}$.

Thus far, it was known only 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.

The linear scale $\frac{k}{m}$ in our result is the best we could hope for, since Peretz (2013) showed that being linearly stronger may not be enough to defend any value 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. Lehrer (1988) showed that 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 say that, asymptotically, the minmax of a "moderately stronger"

[9]This function is at most proportional to the square root of k/m. In particular, it is small when k/m is small.

player is the same as her minmax in the one-stage game.

1.2 A few notes about the proof

What follows is not intended to be a proper "sketch" of the proof, but mainly aims at presenting some of the ideas that drive it.

The following observation, of interest in its own right, plays an important role in the proof. Consider two players who each choose a mixed k-recall strategy (in particular, their randomization is independent). Although their continuation strategies from some stage t on need not be independent,[10] it turns out they cannot be too far from it, due to their bounded memory.

Suppose that the third player uses an m-recall strategy. She can exploit this fact during the following m stages or so. However, she cannot do so directly, since her strategy cannot depend on the time t. Therefore, we first define an auxiliary game as follows.

Fix a pair σ_1, σ_2 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 original repeated game during m consecutive stages, starting at some arbitrary point in time t. Bob, "representing" 3, chooses a strategy to be played during these m stages against σ_1, σ_2. However, Alice, representing 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" (i.e., the continuation after the k-length "memory" that Alice chooses) cannot be too far from

[10]For further discussion see Section 4.2.

independent, as we already said. We then bound the average per-stage correlation, and deduce that Player 3 (more precisely, 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 Bob that is good against any choice of Alice. Therefore, this strategy (employed by 3) is also good against σ_1, σ_2 during m stages of the original game, starting at *any* t.

With (stationary) bounded recall, 3 cannot employ Bob's optimal strategy infinitely many times *independently*. But we show that it suffices to cyclically repeat a long cycle consisting of many independent instances of Bob's optimal strategy.

2 Model and Results

Throughout, a finite three-person game in strategic form is a pair $G = \langle A = A_1 \times A_2 \times A_3, g : A \to [0,1]^3 \rangle$. Namely, it is assumed that the payoffs are scaled[11] between 0 and 1. The minmax value of player $i \in \{1,2,3\}$ is defined as

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

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

The correlated minmax value[12] of player $i \in \{1,2,3\}$ is defined as

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

[11] This is merely a normalization: in games with a larger range of payoffs, some of our derived constants should simply be multiplied by that range.

[12] Also known as the maxmin value.

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

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

where $H(\cdot)$ is Shannon's entropy function.[13]

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 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, 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 $A^{<T} = \bigcup\limits_{0 \leq t < T} A^t$. 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 by 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,

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

denoted by $s^i_{|h_t}$, 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 by $\Sigma^i(k)$. For natural numbers k_1, k_2, k_3, the undiscounted T-fold repeated version of G where each player i is restricted to k_i-recall strategies is denoted by $G^T[k_1, k_2, k_3]$. The payoff in this game 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 finite three-person game $G = \langle A, g \rangle$ and every $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that for every $k_3 \geq k_2 \geq k_1 \geq k_0$ and*

$T \in \mathbb{N} \cup \{\infty\}$,

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

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

$$\operatorname{minmax}_3 G^T[k_1, k_2, k_3] \leq \operatorname{minmax}_3 G + \mathrm{o}(1).$$

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

In Section 5.2 we extend Theorem 2.1 to more than three players (Theorem 5.2).

3 Preliminaries

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

Shannon's entropy[14] 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 by $p(x)$. We have

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

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

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 analyzing repeated games with bounded memory. A key idea of theirs is captured in the following lemma.[15]

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

Let $x_1, \ldots, x_m, y_1, \ldots, y_m$ be finite random variables, and let y_0 be a random variable such that each y_i is a function of $y_0, x_1, \ldots, x_{i-1}$. Suppose that 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,[16] y_0 is a random strategy of a second player, and $y_1, \ldots, y_m$ are the actions played by the second player.

Of special interest is the case where 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,

$$\operatorname{cor} \operatorname{minmax}_3 G(h) \geq \operatorname{minmax}_3 G - \sqrt{2h}, \ \forall h \geq 0,$$

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

The main effort is to prove 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

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

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 that 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, nor even independent conditional[17] on the memories of 1 and 2.

This leads us to define and analyze the following 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 minimizer, and Bob, the maximizer (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 $X_B = \Delta(\Sigma^3_m)$.

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

[17] We elaborate on this point in Section 4.2.

$z \in \Sigma_m^3$ induces a play $a_1, \ldots, a_m$ of Players 1, 2, 3, defined by

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

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 $\sigma^2 \in \Delta(\Sigma^2_{k+m})$,*

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

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*

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

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 that 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

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

For every $t \in [m]$ and every $h_{t-1} \in A^{t-1}$ for which $\mathbf{P}(\bar{a}_{t-1} = h_{t-1}) > 0$, define $Y(h_{t-1}) = I(a^1_t; a^2_t | \bar{a}_{t-1} = h_{t-1})$. By the definition of s^3,

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

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,

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

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

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

To this end, 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 $W = s^2$, as follows.

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

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 the fact that a_t^3 is 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,

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

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

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

□

4.2 The maximizing 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 σ^1 and σ^2. Note that σ^3 is in fact a mixed strategy. Although 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 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 independent (nor even independent 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 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 next. But consider the corresponding auxiliary game $\Gamma_{\sigma^1,\sigma^2,k,m}$. The point is that Γ, being a zero-sum game, has a (possibly mixed) optimal strategy ζ^* of Bob that guarantees the value against any strategy in X_A, i.e., against any possible distribution of s^1, s^2, and h.

Employing ζ^* on a single block should do very well. But had Player 3 acted exactly the same in every block, s^1 and s^2 might have been able to learn something about this during the game. And 3 cannot play infinitely many *independent* instances of ζ^*, as we do not allow 3's strategies to be behavioral. Nevertheless, we show that it is sufficient that 3 cyclically repeats a long cycle consisting of many independent instances of ζ^*.

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 $\zeta^* \in \Delta(\Sigma^3_m)$. In any block $B_i = ((i-1)m+1, \ldots, im)$, Player 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 realizations 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 in 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 analyze the average per-stage payoff r^3 that Player 3 receives in m^2 consecutive blocks, say, $B_1, B_2, \ldots, B_{m^2}$. The analysis is performed 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 $(\rho, \zeta) \in \Delta(\Sigma^1_{k+m} \times \Sigma^2_{k+m} \times A^k \times \Sigma^3_m)$ 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,

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

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, and hence it 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 the 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

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

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: (i) $T_1 + T_2 < m^3$, (ii) m^3 divides T_3, (iii) $T_1 < m$, and (iv) m divides T_2. In the last T_3 stages, σ^3 is defined as above, and the analysis is unaffected.

In 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 $\text{minmax}_3\, G$ 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 instance of ζ^* is played for each block.[18] Formally, Let $z_1, \ldots, z_{T_2/m}$ be i.i.d. variables taking values in Σ^3_m, with distribution ζ^*. In each block B_i, σ^3 plays according to z_i. As above, the optimality of ζ^* implies that the expected average payoff in each B_i is $\geq (\mathit{Vex}\, \text{cor minmax}_3\, G)\left(\frac{2k \ln |A|}{m}\right)$.

Overall, the expected average payoff is at least

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

in the last T_3 stages, and we got 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.

[18]Proving Theorem 2.1 only for small values of T, say $T < m^3$, is significantly simpler and does not need to go through the auxiliary game. Since we needed the auxiliary game for general values of T, we might as well utilize it in this part of the proof as well.

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 in some stages. Dividing T into three phases of length T_1, T_2, and T_3 as above, we need to make sure of three things: knowing the index of the current block in the second phase, knowing the index modulo m^2 in the third phase, and knowing where a block begins. In 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 $a = \lfloor \sqrt{m} \rfloor$ and $b = \lceil \log_{|A_3|}(2m^2) \rceil$. 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 Player 3 first plays γ, and then plays some fixed action different from γ for a stages. Denote this sequence of $a + 1$ actions by $\bar{\alpha}$. This is followed by a "counter" $\bar{\beta}_i$ that designates the current (second or third) phase plus the block index (i.e., the absolute index in the second phase or the 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.

In 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 to replace the definition of the auxiliary game Γ by that of the game $\Gamma(\bar{\alpha}, \bar{\beta}_i, \gamma)$, defined the same except that the strategies of Bob are restricted to playing $\bar{\alpha}$ 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, in any block we get the same average payoff as 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 undiscounted 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 still holds for a discounted payoff, when the discount rate approaches 0.

Suppose that we allowed Players 1 and 2 to play mixtures of behavioral 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 complexity limitations of these players were used in the proof only to make the following assertion: the continuation strategy of i at any point in time depends only on the last k_i actions. The assertion is true in this model as well.

Another plausible variation of the model is to allow strategies to depend not only on the last k_i actions, but also on calendar time. Here, too, our result holds. The proof of 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 Player 3 being able to repeat a strategy indefinitely. Note that it is crucial for Player 3 to condition her actions on calendar time. Otherwise, if Players 1 and 2 conditioned on time while 3 did not, any fixed (i.e., 0-recall) normal sequence of actions of Players 1 and 2 would seem random to 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, and its elements are called *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., $z_{t+1} = q(z_t, a_t^{-i})$. In 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 allow 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 in the case of bounded recall, or the current state of i's automaton in the case of automata. 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, which is the counterpart of Theorem 2.1 for finite automata.

Theorem 5.1. *For every finite three-person game $G = \langle A, g\rangle$ and every $\epsilon > 0$ there exists $s_0 \in \mathbb{N}$ such that for every $s_3 \geq s_2 \geq s_1 \geq s_0$ and $T \in \mathbb{N} \cup \{\infty\}$,*

$$\operatorname{minmax}_3 G^T(s_1, s_2, s_3) \geq (\mathit{Vex}\ \operatorname{cor}\operatorname{minmax}_3 G)\left(\frac{2\ln s_2\ \ln|A|}{\ln s_3}\right) - \epsilon,$$

where $G^T(s_1, s_2, s_3)$ denotes the undiscounted 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., transition functions of the form $q : Z \times A^{-i} \to \Delta(Z)$.

5.2 Many players

In this Section we extend our result on the minmax in a three-player repeated game to a game with any number of players. In order to do that, we need to define a notion that extends the notion of mutual information to more than two random variables. One such extension is the following.

The *total correlation* of a tuple of discrete random variables $x_1, \ldots, x_d$ is defined as

$$C(x_1, \ldots, x_d) = \sum_{i=1}^{d} H(x_i) - H(x_1, \ldots, x_d)\,.$$

The *Kullback–Leibler divergence* from p to q (a.k.a. *relative entropy*; see, e.g., Cover and Thomas 2006, Chapter 2.3), where p and q are discrete probability distributions, is defined as $D_{KL}(p\|q) = \sum_\xi p(\xi) \ln \frac{p(\xi)}{q(\xi)}$. The total correlation of $x_1, \ldots, x_d$ also equals the divergence from the joint distribution of these variables to the product of their marginal distributions, i.e.,

$$C(x_1, \ldots, x_d) = D_{KL}\left(p(x_1, \ldots, x_d) \| \, p(x_1) \otimes \ldots \otimes p(x_d) \right). \tag{5.1}$$

We extend the notion of h-correlated minmax to an n-player game in strategic form $\langle N, A, g \rangle$ by

$$\operatorname{cor\,minmax}_i G(h) := \min_{\substack{x^{-i} \in \Delta(A_{-i}): \\ C(x^{-i}) \leq h}} \max_{a^i \in A_i} g^i(x^{-i}, a^i),$$

where x^{-i} is regarded as an $(n-1)$-tuple. Note that this is in fact the same expression used to define this notion in Section 2.

The following theorem is the n-player extension of Theorem 2.1.

Theorem 5.2. *For every finite game $G = \langle N, A, g \rangle$ and every $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that for every $k_n \geq \ldots k_2 \geq k_1 \geq k_0$ and $T \in \mathbb{N} \cup \{\infty\}$,*

$$\begin{aligned} &\operatorname{minmax}_n G^T[k_1, \ldots, k_n] \\ &\qquad \geq (\mathit{Vex} \ \operatorname{cor\,minmax}_n G)\left((n-1) \ln |A| \cdot \frac{k_{n-1}}{k_n} \right) - \epsilon \\ &\qquad\qquad \geq \operatorname{minmax}_n G - \sqrt{2(n-1) \ln |A| \cdot \frac{k_{n-1}}{k_n}} - \epsilon \, . \end{aligned}$$

5.2.1 Proof of Theorem 5.2

The conditional total correlation of $x_1, \ldots, x_d$ given z is defined by

$$C(x_1, \ldots, x_d | z) = \sum_{i=1}^{d} H(x_i|z) - H(x_1, \ldots, x_d|z) \, .$$

Lemma 5.3. (i) *If y_i is a function of x_i for every $1 \le i \le d$, then*

$$C(x_1, \ldots, x_d) - C(y_1, \ldots, y_d) \ge C(x_1, \ldots, x_d | y_1, \ldots, y_d).$$

(ii) *Moreover, if y_i is a function of z and x_i for every $1 \le i \le d$, then*

$$C(x_1, \ldots, x_d | z) - C(y_1, \ldots, y_d | z) \ge C(x_1, \ldots, x_d | y_1, \ldots, y_d, z).$$

In particular, $C(x_1, \ldots, x_d) \ge C(y_1, \ldots, y_d)$ on (i), and $C(x_1, \ldots, x_d | z) \ge C(y_1, \ldots, y_d | z)$ on (ii).

Proof. If b is a function of a, then $H(a) = H(a, b)$; therefore,

$$H(a) - H(b) = H(a, b) - H(b) = H(a|b).$$

To prove (i), write

$$\begin{aligned}
&C(x_1, \ldots, x_d) - C(y_1, \ldots, y_d) \\
&\quad = \left(\sum_{i=1}^{d} H(x_i) - H(x_1, \ldots, x_d) \right) - \left(\sum_{i=1}^{d} H(y_i) - H(y_1, \ldots, y_d) \right) \\
&\quad = \sum_{i=1}^{d} (H(x_i) - H(y_i)) - (H(x_1, \ldots, x_d) - H(y_1, \ldots, y_d)) \\
&\quad = \sum_{i=1}^{d} H(x_i | y_i) - H(x_1, \ldots, x_d | y_1, \ldots, y_d) \\
&\quad \ge \sum_{i=1}^{d} H(x_i | y_1, \ldots, y_d) - H(x_1, \ldots, x_d | y_1, \ldots, y_d) \\
&\quad = C(x_1, \ldots, x_d | y_1, \ldots, y_d).
\end{aligned}$$

The proof of (ii) is similar. □

We first prove the second inequality of Theorem 5.2, similarly to Theorem 2.1. A more general form of Pinsker's inequality states that for discrete probability distributions p and q, $\|p-q\|_1 \leq \sqrt{2D_{KL}(p\|q)}$. The payoff function g^n is 1-Lipschitz w.r.t. the $\|\cdot\|_1$ norm; therefore, by (5.1),

$$\text{for any } h \geq 0, \ \ \operatorname{cor\,minmax}_n G(h) \geq \operatorname{minmax}_n G - \sqrt{2h}\,,$$

and since the function on the right-hand side is convex, it is smaller than $(\mathit{Vex}\ \operatorname{cor\,minmax}_n G)(h)$ as well.

To prove the first inequality of Theorem 5.2, we need to review the proof of Theorem 2.1, written for three-player games, and adapt it to general n-player games. Let us start by reviewing Lemma 4.2, which turns out to require most of the work.

The general form of the lemma would state that against any tuple of random strategies $s^1, \ldots, s^{n-1}$, Player n has a pure response s^n that yields her an expected average of at least $(\mathit{Vex}\operatorname{cor\,minmax}_n G)\left(\frac{C(s^1,\ldots,s^{n-1})}{m}\right)$. Reviewing the proof of Lemma 4.2, the adaptation to n players is straightforward up to the point where the proof says that it remains to show that

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

Hence, the general proof should show that $\sum_{t=1}^{m} C(a_t^1, \ldots, a_t^{n-1} | \bar{a}_{t-1}) \leq C(s^1, \ldots, s^{n-1})$. We write it as a separate lemma:

Lemma 5.4. *Let $s^1, \ldots, s^d$ be random strategies of the players of a d-player repeated game. The play $a_1, a_2, \ldots$ induced by $(s^1, \ldots, s^d)$ satisfies that for any m,*

$$\sum_{t=1}^{m} C(a_t^1, \ldots, a_t^d | \bar{a}_{t-1}) \leq C(s^1, \ldots, s^d)\,.$$

Proof. For any $t \geq 1$ and $1 \leq i \leq d$, a_t^i is a function of $\bar{a}_{t-1}$ and s^i. By Lemma 5.3 (ii),

$$C(s^1, \ldots, s^d | \bar{a}_{t-1}) - C(a_t^1, \ldots, a_t^d | \bar{a}_{t-1}) \geq C(s^1, \ldots, s^d | \bar{a}_t)\,, \tag{5.2}$$

because $(a_t^1, \ldots, a_t^d, \bar{a}_{t-1}) = \bar{a}_t$.

Rearrange (5.2) as

$$C(a_t^1, \ldots, a_t^d | \bar{a}_{t-1}) \leq C(s^1, \ldots, s^d | \bar{a}_{t-1}) - C(s^1, \ldots, s^d | \bar{a}_t) \,,$$

and sum both sides from $t = 1$ to m, to get

$$\sum_{t=1}^{m} C(a_t^1, \ldots, a_t^d | \bar{a}_{t-1}) \leq C(s^1, \ldots, s^d | \bar{a}_0) - C(s^1, \ldots, s^d | \bar{a}_m) \,,$$

and the RHS is simply $C(s^1, \ldots, s^d) - C(s^1, \ldots, s^d | \bar{a}_m) \leq C(s^1, \ldots, s^d)$. □

Adapting the construction of the auxiliary game to n players is, again, straightforward, with Bob representing Player n and Alice representing Players $1, \ldots, n-1$. We should show that the value of this game is at least

$$(\mathit{Vex} \;\; \mathrm{cor\,minmax}_n\, G) \left((n-1) \ln |A| \cdot \frac{k_{n-1}}{k_n} \right) ,$$

generalizing Lemma 4.1.

Where the proof of Lemma 4.1 shows that $I(s^1_{|h}; s^2_{|h}) \leq 2k \ln |A|$, the adapted proof should show that $C(s^1_{|h}, \ldots, s^{n-1}_{|h}) \leq (n-1)k \ln |A|$, where h and s_i are as in the auxiliary game. $s^i_{|h}$ is a function of the pair (h, s^i), therefore,

$$
\begin{aligned}
C(s^1_{|h},\dots,s^{n-1}_{|h}) &\le C\left((h,s^1),\dots,(h,s^{n-1})\right)\\
&= \sum_{i=1}^{n-1} H(h,s^i) - H\left((h,s^1),\dots,(h,s^{n-1})\right)\\
&= \sum_{i=1}^{n-1} H(h,s^i) - H(h,s^1,\dots,s^{n-1})\\
&\le \sum_{i=1}^{n-1} \left[H(h)+H(s^i)\right] - H(s^1,\dots,s^{n-1})\\
&= (n-1)H(h) + \sum_{i=1}^{n-1} H(s^i) - H(s^1,\dots,s^{n-1})\\
&= (n-1)H(h) \;\le\; (n-1)k\ln|A|\ ,
\end{aligned}
$$

where the last inequality holds since the size of the support of h is at most $|A|^k$, and the preceding equality holds since $s^1,\dots,s^{n-1}$ are independent.

The adaptation of the rest of the proof of Theorem 2.1, once we are done with the auxiliary game, is straightforward. Against the strategies $\sigma^1,\dots,\sigma^{n-1}$, Player n derives the maximizing strategy from the auxiliary game as we describe there, and the number of players she faces is immaterial.

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 down 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 $\text{minmax}_3\, G^\infty[k, k, k]$ 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 the Department of Economics at Bar-Ilan University, the Sir Isaac Wolfson Chair in Economics and Business Administration, 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, 1993.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley Interscience, New York, 2nd edition, 2006.

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, 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, Berlin, 1997.

Abraham Neyman. Learning effectiveness and memory size. Discussion Paper 476, Center for the Study of Rationality, Hebrew University, Jerusalem, 2008.

Abraham Neyman and Daijiro Okada. Repeated games with bounded entropy. *Games and Economic Behavior*, 30(2):228–247, 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.

Michele Piccione and Ariel Rubinstein. On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1): 3–24, 1997.

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*, 69(1):99–118, 1955.

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