Games and Economic Behavior 88 (2014) 71–89

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Online concealed correlation and bounded rationality ☆

Gilad Bavly [a,*], Abraham Neyman [b]

[a] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[b] Institute of Mathematics and Center for the Study of Rationality, Hebrew University, 91904 Jerusalem, Israel

A R T I C L E I N F O

Article history:
Received 18 July 2006
Available online 27 August 2014

JEL classification:
C72
C73

Keywords:
Repeated games
Bounded recall
Finite automata
Concealed correlation
Folk theorem
Equilibrium payoffs

A B S T R A C T

Correlation of players' actions may evolve in the common course of the play of a repeated game with perfect monitoring ("online correlation"). In this paper we study the concealment of such correlation from a boundedly rational player. We show that "strong" players, i.e., players whose strategic complexity is less stringently bounded, can orchestrate the online correlation of the actions of "weak" players, where this correlation is concealed from an opponent of "intermediate" strength. The feasibility of such "online concealed correlation" is reflected in the individually rational payoff of the opponent and in the equilibrium payoffs of the repeated game.

This result enables the derivation of a folk theorem that characterizes the set of equilibrium payoffs in a class of repeated games with boundedly rational players and a mechanism designer who sends *public* signals.

The result is illustrated in two models, bounded recall strategies and finite automata.

1. Introduction

Consider a group of agents interacting with each other sequentially, where the overall strategies employed by these agents are independent. Some level of correlation between their actions may still evolve; i.e., if we look at the actions that they take at some point in time, these actions may be correlated.[1] On the other hand, as every mixed strategy (in a game with perfect recall) is equivalent to a behavioral one (Kuhn, 1953), conditional on the full history up to that point in time, these actions are independent.[2] Therefore, the correlation of actions is not *concealed* from a fully rational observer: this observer has the resources to make a statistical prediction of the coming tuple of actions, so that conditional on his prediction, the actions are independent. This fact plays an important, albeit implicit, role in various folk theorems of repeated games.

☆ This research was supported in part by Israel Science Foundation grants 382/98, 263/03, 1123/06, 1596/10, and 538/11, by United States–Israel Binational Science Foundation Grant 2010253, and by the Zvi Hermann Shapira Research Fund grant 82893. Helpful discussions and comments by Elchanan Ben-Porath, Olivier Gossner, Penelope Hernandez, Rann Smorodinsky, and Eilon Solan are gratefully acknowledged.

* Corresponding author.
E-mail addresses: gilbav@math.huji.ac.il (G. Bavly), aneyman@huji.ac.il (A. Neyman).
URL: http://www.ratio.huji.ac.il/neyman (A. Neyman).

[1] For example, if player 1 chooses with probability 1/2 to play repeatedly the action α, and with probability 1/2 to play repeatedly the action β, and player 2 imitates at each stage the previous action of player 1, then at any stage $t \geq 2$, the probability distribution over the action pair (a_t^1, a_t^2) is correlated: the probability of (α, α) is $\frac{1}{2}$, and so is the probability of (β, β).

[2] E.g., in the previous example, at any stage $t \geq 2$, conditional on the history $(a_1, \ldots, a_{t-1})$, the distribution of the action pair (a_t^1, a_t^2) is uncorrelated: either the history tells you (from stage 2 on) that the forthcoming pair is (α, α) with probability 1, or it tells you that it is (β, β) with probability 1.

http://dx.doi.org/10.1016/j.geb.2014.08.004
0899-8256/

The current paper is concerned with repeated games in which players are not fully rational. We consider the impact on equilibrium when the players' ability to gather and process information and make computations is not unlimited. The literature contains various models of such bounded complexity in repeated games. One prominent model is that of finite automata (see, e.g., Ben-Porath, 1986; Neyman, 1985, 1998; Kalai, 1990; Papadimitriou and Yannakakis, 1994), where each player has some finite capacity for memory storage. Another prominent model is bounded recall (see, e.g., Aumann, 1981; Lehrer, 1988), where each player has a finite number of the *last* previous stages that he is able to recall. There are, of course, variations to these models (e.g., Rubinstein, 1986), as well as other interesting models (e.g., Gilboa and Schmeidler, 1994; Gossner et al., 2006). The paper is focused on the bounded recall model, for which the main results of the paper are discussed, stated, and proved, in detail. The analogous results for the variations of the bounded recall model, as well as for the finite automata model, are stated and derived in the last section of the paper.

For the case of more than two players, little is known about the equilibrium payoffs of repeated games with boundedly rational players. The main difficulty in the characterization of equilibrium payoffs of infinitely repeated games lies in the identification and quantification of feasible punishments, which depend on the possibilities of concealed correlation that a group of players may have.

The possibility of concealing correlation certainly depends upon the capabilities of the players. It seems reasonable, and is demonstrated in various models, that "stronger" players (i.e., players with higher capabilities) can out-strategize "weaker" opponents, and the concealment of correlation may be one manifestation of this. In this paper we focus on the surprising, indeed counter intuitive, possibility of concealing correlation of players' actions from a stronger opposition.

A group of players, called the *concealing group*, can *conceal* a distribution D over tuples of its members' actions from other players (or observers), if they have a profile of (independent) strategies such that (1) for every strategy profile of the other players, the empirical distribution of the sequence of their actions is close to D, and (2) this sequence appears, in the eyes of the other players, to be close to an i.i.d. play with that distribution D.

We can distinguish two instances of the problem of concealing a *correlated* distribution of action profiles. The distinction is based on the relations between the recall capacities of the concealing group and that of the other players. For clarity's sake we assume[3] that there is only one other player, called the *opponent*. The first instance is where the concealing group consists of relatively weak players and the opponent is stronger.[4] The present paper focuses on the second instance, where the concealing group of players contains both weak and strong players relative to the opponent.

Suppose (for example) that the concealing group contains two players, whose recall is shorter than that of the opponent, and an additional player, and the group tries to conceal from the opponent a distribution of action profiles in which the actions of the two weak players are correlated. Then why should this additional player matter at all? After all, in our model there is no pre-game communication, and all communication opportunities are embedded within the actions available to the players, and these actions are public. Therefore, it seems a priori, that the opponent, whose recall is longer than that of the weak players, can untangle whatever use the two weak players can make of the actions of the strong player.

The main message of this paper is that a concealing group that includes a player whose recall is longer than that of the opponent (or observer) can conceal from the opponent a distribution of the group's action profiles in which the actions of the "weak" players – those whose recall is shorter than that of the opponent – are correlated (Theorems 2.1 and 2.2).

Theorem 2.1, which is a special case of the more general Theorem 2.2, applies to a four-player game: the concealing group consists of two players of relatively short recall, called the *weak* players, and a player with a long recall, called the *strong* player, and the opponent has intermediate recall. The theorem illustrates the possibility that the concealing group can conceal from the opponent a distribution (over tuples of its members' actions) in which the actions of the two weak players are correlated. Moreover, it specifies simply stated conditions on the stage game and the recall capacities of the players that enable the concealing group to conceal from the opponent *any* correlated distribution of the two weak players' actions.

The more general result, Theorem 2.2, applies to repeated games with an arbitrary number of players: the opponent has intermediate-length recall, and the concealing group, which consists of the set of all other players, is composed of two groups: weak players with relatively short recall and strong players with long recall. The theorem specifies simply stated conditions on their recall capacities and an information-theoretic condition on a (possibly correlated) distribution of joint actions of the concealing group, so that the concealing group can conceal this distribution from the opponent.

In both theorems it is assumed that the recall of the opponent is subexponential in the recall of the weak players. A natural question that arises is whether this "subexponential" condition is essential for the conclusions of the theorems. Theorem 2.3 states that when the recall of the opponent is longer than some exponential function of the recall of the weak players, then the concealing group cannot conceal more than a negligible amount of correlation of the weak players' actions. Moreover, this conclusion holds even if the concealing group can correlate their strategies before the start of the game, and even if their strategies are allowed to choose a randomized action as a function of the recalled past, and the strong players are fully rational.

The reader may by now be wondering why allies need to correlate their actions "online," i.e., in the common course of play, by the *public* sequence of actions taken. Can they not do it "offline" by communicating through private channels? The

[3] It can be shown that this assumption is w.l.o.g.

[4] The simple example of this in Section 8.3 demonstrates that things are not so easy for the opponent.

fact of the matter is that sometimes they cannot. Some examples are anti-cartel regulations, or various types of multistage auctions, where "the rules of the game" forbid collusion.

Online concealed correlation affects the individually rational payoff (i.r.p.) in the repeated game. In fact, the concealment of correlation in Theorems 2.1 and 2.2 is stated in terms of the i.r.p. Section 8.1 defines a notion of concealment that depends upon the strategies at the players' disposal (and involves no payoffs). This notion enables us to reformulate the conclusion of the main result in a form that emphasizes the fact that the result concerns the concealment of a *distribution*, independently of the payoffs.

The rest of the paper is organized as follows. Section 2 describes the model and the main results. In Section 3 we prove Theorem 2.3. In Section 4 we prove an instance of Theorem 2.1, for the three-player matching pennies game. This instance provides a relatively simple counter-example to Lehrer (1994, Theorem 1) (see also comment 1 in Section 8.5), and serves as a helpful introduction to some of the ideas in the general results.

In Section 5 we prove Theorem 2.1 by showing how the proof for the three-player matching pennies game can be extended to other four-player games. The proof of Theorem 2.1 presents some of the ideas that appear in greater generality in the proof of Theorem 2.2. In addition, there are a few differences between the two proofs. Therefore each of these proofs may be a step towards different future extensions and related results. In Section 6 we prove Theorem 2.2. In Section 7 we discuss the implications of our results for equilibrium payoffs.

Section 8.2 discusses variations of bounded recall strategies, and Section 8.3 contains an example of a three-player game, where the players each have equal strength and can recall the last actions of only the other players, and two players can conceal a correlation of their joint actions from the third player. Section 8.4 explains how the main results yield analogous results in the model of repeated games with finite automata. Section 8.5 contains several additional remarks.

2. The model and the main result

Our discussion and results are stated for undiscounted infinitely repeated games. However, it will be clear from the tools we use that the discussion and results apply to sufficiently long finitely repeated games, as well as to infinite discounted games with a large enough discount factor (see comment 2 in Section 8.5).

2.1. Infinitely repeated games

Let $G = (N, A, r)$ be an n-player game in strategic form. $N = \{1, 2, \ldots, |N|\}$ is the finite set of players, $A = \bigtimes_{i\in N} A_i$ where A_i is the set of actions of player i, and $r = (r^i)_{i\in N}$ where $r^i : A \to \mathbb{R}$ is the payoff function of i. The linear extension of r^i to the functions defined on $\Delta(A)$, where $\Delta(*)$ stands for the probability distributions over the set $*$, is also denoted by r^i, and for a distribution y on a set B and a distribution x on a set C we denote by (y, x) the product distribution $y \otimes x$ on $B \times C$.

Player i's *individually rational payoff* (i.r.p.) in the mixed extension of G is $\bar{v}^i = \min_{y\in\bigtimes_{j\neq i}\Delta(A_j)} \max_{x\in A_i} r^i(y, x)$.

The game G_∞ denotes the infinite repetition of G with perfect monitoring. At each stage $t = 1, 2, \ldots,$ the players play the game G (the *stage game*); i.e., at stage t, player i chooses an action $a^i_t \in A_i$.

A *play* of G_∞ is an infinite sequence $a_1, a_2, \ldots,$ where $a_t = (a^i_t)_{i\in N} \in A$. In our context, the payoff for i in G_∞ is the "limit" of his average payoff along the play,[5] namely, "$\lim$"$_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} r^i(a_t)$. However, we qualify "limit," since a sequence of payoffs $(r^i(a_t))_{t\geq 1}$ induced by some play $(a_t)_{t\geq 1}$ need not have a limit of means (Cesaro limit). Nevertheless, such a limit does exist for any play that is defined by stationary bounded recall strategies, or finite automata strategies.

The set of all histories that may be played at the first $t-1$ stages is A^{t-1} (where A^0 stands for $\{\emptyset\}$), and $A^* = \bigcup_{t=1}^{\infty} A^{t-1}$ is the set of all possible histories, of any length. A pure strategy of i in the repeated game with perfect monitoring G_∞ is a function $\sigma^i : A^* \to A_i$. For any history $h = (a_1, \ldots, a_{t-1}) \in A^{t-1}$, $\sigma^i(h)$ is the action that player i will take at stage t, if the history at that stage is h. A profile $\sigma = (\sigma^i)_{i\in N}$ of pure strategies in G_∞ defines a play $(a_t(\sigma))_{t\geq 1}$ by induction on t: $a^i_1 = \sigma^i(\emptyset)$, and $a^i_{t+1}(\sigma) = \sigma^i(a_1(\sigma), \ldots, a_t(\sigma))$.

2.2. Bounded recall strategies

A *stationary bounded recall* (*bounded recall* for short, or SBR) strategy for player i in G_∞ assumes that i's play at any given stage relies only on the last m_i actions played.

A (pure) *m-recall strategy for i* is a pure strategy σ^i such that $\sigma^i(a_1, \ldots, a_{t-1}) = \sigma^i(a_{t-m}, \ldots, a_{t-1})$ for $t > m$.

Denote by $BR^i(m)$ the set of all m-recall strategies of player i in G_∞. Note that any m-recall strategy is, in particular, a k-recall strategy for any $k > m$, i.e., $BR^i(k) \supset BR^i(m)$.

For a tuple $\vec{m} = (m_i)_{i\in N}$, the game $G(\vec{m})$ is defined as the infinite repetition of G, but where the strategies of player i are his m_i-recall strategies. I.e.,

$$G(\vec{m}) = \big(N, \big(BR^i(m_i)\big)_{i\in N}, \bar{r}\big),$$

[5] The payoff in infinitely repeated games is sometimes taken to be the discounted average of payoffs, namely, $\lambda\sum_{t=1}^{\infty}(1-\lambda)^{t-1} r^i(a_t)$, for some $0 < \lambda < 1$. In this case the payoff is called *discounted*.

where $\bar{r}$ is defined for a tuple $\phi = (\phi_i)_{i\in N} \in \bigtimes_{j=1}^{n} BR^j(m_j)$ by $\bar{r}(\phi) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} r(a_t(\phi))$, where $(a_t(\phi))_{t\geq 1}$ is the play defined by the strategy profile ϕ.

Here we can write lim without reservations, since any play resulting from SBR strategies will be periodic. Let M be an integer s.t. $\forall i\ m_i \leq M$. Since there are at most $|A|^M$ possible memories of length M, in any play one of them is bound eventually to appear twice, say at stages t_1, t_2. Then, by induction on $t \geq 1$, every player takes at $t_2 + t$ the same action he took at $t_1 + t$, and therefore the play enters a cycle. The limiting average payoff will then simply be the average payoff over the cycle.

In the sequel, we examine player i's i.r.p. in the mixed extension of $G(\vec{m})$, i.e.,

$$\bar{v}^i(\vec{m}) = \min_{\sigma\in\bigtimes_{j\neq i}\Delta(BR^j(m_j))}\ \max_{\tau\in BR^i(m_i)} \bar{r}^i(\sigma,\tau),$$

where $\bar{r}$ is now extended to mixed strategies; that is, it is the expectation of our previous $\bar{r}$.

Let $\underline{v}^i(\vec{m})$ denote the maxmin level, namely,

$$\underline{v}^i(\vec{m}) = \max_{\tau\in\Delta(BR^i(m_i))}\ \min_{\sigma^{-i}\in\bigtimes_{j\neq i} BR^j(m_j)} \bar{r}^i\big(\tau,\sigma^{-i}\big).$$

Remark. A *behavioral* strategy for player i in G_∞ is a function $\sigma^i : A^* \to \Delta(A_i)$; i.e., for every possible history it specifies a probability distribution over i's actions. $BR^i_b(m)$ denotes the set of all m-recall behavioral strategies of player i.

Our main result also holds for behavioral bounded recall strategies. In particular, if we define $\bar{v}^i_*(\vec{m})$ the same way as $\bar{v}^i(\vec{m})$ except that the maximum is taken over $BR^i_b(m_i)$, and define $\bar{v}^i_b(\vec{m})$ with behavioral strategies for all players, i.e.,

$$\bar{v}^i_b(\vec{m}) = \min_{\sigma\in\bigtimes_{j\neq i}\Delta(BR^j_b(m_j))}\ \max_{\tau\in BR^i_b(m_i)} \bar{r}^i(\sigma,\tau),$$

then the result holds for $\bar{v}^i_*$ and hence for $\bar{v}^i_b$ too (since of course $\bar{v}^i_* \geq \bar{v}^i_b$).

Notation: for two functions $g, f : \mathbb{N} \to \mathbb{R}_+$ we write $g \ll f$ or $g(n) \ll f(n)$ if $g(n)/f(n) \to_{n\to\infty} 0$.

2.3. The main result

In the main results, the set of players N is fixed, and the recall capacities m_i, $i \in N = \{1, 2, \ldots, |N|\}$, will depend on a parameter $n \in \mathbb{N}$.

For a gradual presentation of the main result, we start with a special case of the result and a restriction of the number of players to four. The general result will be stated thereafter.

Theorem 2.1. *Let $G = (N, A, r)$ be a four-player game, and let m_i: $\mathbb{N} \to \mathbb{N}$, $i \in N = \{1, 2, 3, 4\}$, with $\lim_{n\to\infty} m_i(n) = \infty$, satisfy*

(a) $|A_4| \geq \min(|A_1|, |A_2|)$
(b) $m_3 \ll m_4$
(c) $\log m_3 \ll \min(m_1, m_2)$.

Then,

$$\limsup_{n\to\infty} \bar{v}^3(\vec{m}) \leq \min_{y\in\Delta(A_1\times A_2)}\ \max_{x\in A_3} r^3(y,x,z), \tag{2.1}$$

where z is the uniform distribution over A_4, and $\vec{m} = (m_1, m_2, m_3, m_4)$.

In particular, if, in addition, the payoff function r^3 is independent of the actions of player 4, then

$$\limsup_{n\to\infty} \bar{v}^3(\vec{m}) \leq \min_{y\in\Delta(A_1\times A_2)}\ \max_{x\in A_3} r^3(y,x)$$

and if, in addition, $\log m_4 \ll m_3$, then

$$\lim_{n\to\infty} \bar{v}^3(\vec{m}) = \lim_{n\to\infty} \underline{v}^3(\vec{m}) = \min_{y\in\Delta(A_1\times A_2)}\ \max_{x\in A_3} r^3(y,x). \tag{2.2}$$

Inequality (2.1) asserts that for every distribution y over action pairs of players 1 and 2, $\limsup_{n\to\infty} \bar{v}^3(\vec{m})$ is less than or equal to $\max_x r^3(y, x, z)$, where z is the uniform distribution over the actions of player 4.

The next theorem generalizes inequality (2.1) in two important directions. First, z need not be the uniform distribution, and y need not be independent of z. We provide a condition on the distribution D^{-3} of the actions of players 1, 2, and 4, such that for any distribution D^{-3} that satisfies the condition, we have $\limsup_{n\to\infty} \bar{v}^3(\vec{m}) \leq \max_{x\in A_3} r^3(D^{-3}, x)$. In addition, the number of players need not be 4; we allow for an arbitrary number of players (and more than one of them may be strong).

The condition on the distribution D^{-3}, or on D^{-k} in the case of an arbitrary number $|N| \geq k$ of players, is stated by using the following information-theoretic notion (see, e.g., Cover and Thomas, 1991).

For a finite set A, and an A-valued random variable x, denote $P_a = \Pr\{x = a\}$. The *entropy* of x is defined[6] as $H(x) = -\sum_{a \in A} P_a \log(P_a)$. Note that if x is uniformly distributed, then $H(x) = \log|A|$, and that if x, y are independent then $H(x, y) = H(x) + H(y)$.

The assumptions on the recall capacities m_i, $i \in N = \{1, 2, \ldots, |N|\}$, which depend on a parameter $n \in \mathbb{N}$, are:

(A1) $m_{i+1}(n) \geq m_i(n) \to_{n\to\infty} \infty$
(A2) $m_k \ll m_{k+1}$
(A3) $\log m_k \ll m_1$.

The inequality $m_{i+1}(n) \geq m_i(n)$ in assumption (A1) orders the players in N according to their recall capacity. The assumptions (A2) and (A3) single out one player $k \in N$, and partition the other players into two groups: $J^+ = \{i \in N : i > k\}$ and $J = \{i \in N : i < k\}$. Assumption (A2) asserts that each player $i \in J^+$ has a much longer recall than that of player k, and assumption (A3) asserts that the length of player k's recall is subexponential in that of each one of the players $j \in J$.

For $k \in N$ we denote by $-k$ the subset $\{j \in N : j \neq k\}$ of players, and for a nonempty subset $S \subset N$ we denote by D^S a distribution on $A_S := \times_{j\in S} A_j$. For every $j \in S$, respectively a nonempty subset $S^* \subset S$, the marginal distribution of D^S on A_j, respectively on A_{S^*}, is denoted D^j, respectively D^{S^*}.

Theorem 2.2. *Let (N, A, r) be a game with $N \supset \{1, 2, \ldots k\}$, and assume that the recall capacities $m_i(n)$ $(i \in N)$ satisfy (A1), (A2), and (A3). Let D^{-k} be a distribution on A_{-k} with marginals D^j on A_j. If*

$$H\big(D^{-k}\big) \geq \sum_{j<k} H\big(D^j\big), \tag{2.3}$$

then

$$\limsup_{n\to\infty} \bar{v}^k(\vec{m}) \leq \max_{x\in A_k} r^k\big(D^{-k}, x\big). \tag{2.4}$$

Note that Theorem 2.2 generalizes Theorem 2.1. Indeed, if (1) $N = \{1, 2, 3, 4\}$, (2) z is the uniform distribution on A_4, (3) y is a distribution on $A_1 \times A_2$, and (4) $D^{-3} = y \otimes z$ is the product distribution (on $A_{\{1,2\}} \times A_4$) of y and z, then $H(D^{-3}) = H(z) + H(D^{\{1,2\}}) = \log|A_4| + H(D^{\{1,2\}}) \geq \log|A_4| + H(D^i)$ $(\forall i \in \{1, 2\})$, and therefore $H(D^{-3}) \geq H(D^1) + H(D^2)$ whenever $|A_4| \geq \min(|A_1|, |A_2|)$.

There are other special cases of Theorem 2.2 that are of independent interest. One such special case is when there is only one "strong player," i.e., when $|N| = k + 1$.

Theorems 2.2 and 2.1 have been stated in terms of the i.r.p. of the repeated game. An alternative, and more general, concept of an $N \setminus \{k\}$ strategy profile σ^{-k} concealing a distribution D^{-k} from a class of strategies of player k is presented in Section 8.1.

Our proof of Theorem 2.2 shows that for any distribution D^{-k} that satisfies conditions (A1), (A2), and (A3), there is an $N \setminus \{k\}$ strategy profile $\sigma^{-k} = (\sigma^j)_{j\neq k}$ of mixtures of m_j-recall strategies that (asymptotically, i.e., for sufficiently large n, it ε-) conceals D^{-k} from the class of m_k-recall strategies of player k. Informally, it shows that for every m_k-recall strategy τ^k of player k, the distribution that is defined by $\sigma = (\sigma^{-k}, \tau^k)$ on plays of the repeated game, is such that for almost all t, the conditional distribution of a_t^{-k}, given a_t^k, is close to D. It follows that $\limsup_{n\to\infty} \max_{\tau^k \in BR^k(m_k)} \bar{r}(\sigma^{-k}, \tau^k) \leq \max_{x\in A_k} r(D^{-k}, x)$ holds for any payoff function $r : A \to \mathbb{R}$.

2.4. Tightness of the subexponential condition

The above results demonstrate the feasibility of concealing correlation against a subexponential opponent. The following theorem shows that this subexponential condition is tight, in the following sense: there exists an exponential relation s.t. a group of players cannot conceal more than a negligible amount of correlation against an opponent, whose strength stands in that relation to their strength, even if other players help them.[7] Moreover, this is the case even if the players use behavioral bounded recall strategies (note that a mixture of m-recall behavioral strategies need not be equivalent to an m-recall behavioral strategy).

[6] Henceforth $\log = \log_2$.
[7] In fact, the conclusion holds even when this strong "opponent" does not actually participate in the game, but merely observes it.

The theorem states that for every $\varepsilon > 0$ there are universal prediction functions $\mu_i : A^k \to \Delta(A_i)$ $(i \in N)$, such that for any subset J of players and a correlated strategy σ with σ^J being a mixture of $(1-\varepsilon)\frac{\log k}{\log|A|}$-recall behavioral strategies, the empirical distance between the conditional σ-distribution of a_t^J given $(a_1, \ldots, a_{t-1})$ and the product distribution $\bigotimes_{j \in J} \mu_j(a_{t-k}, \ldots, a_{t-1})$ is close to 0.

The theorem may be viewed as an asymptotic version of one aspect of Kuhn's theorem (Kuhn, 1953), namely, that conditioning on the full history results in a product distribution. But, in addition, here the prediction strategy is time-independent and universal (i.e., it is independent of the strategies of the players), and the theorem holds even if players are allowed to correlate their randomized strategies.

The allowance for behavioral strategies only serves to strengthen the result. The proof is simpler if one considers only randomization over pure strategies.

Let Σ^i denote the set of all (not necessarily bounded recall) behavioral strategies of player i in the repeated game. For a subset $J \subset N$ of players and a profile $a \in A$ of actions, Σ^J is the cartesian product $\bigtimes_{j \in J} \Sigma^j$, $BR_b^J(m)$ is the cartesian product $\bigtimes_{j \in J} BR_b^j(m)$, and a^J is the list $(a^j)_{j \in J}$ of the J coordinates of a.

Each strategy profile $s \in BR_b^J(m) \times \Sigma^{N\setminus J}$ defines a probability distribution P_s on plays of the repeated game, and in particular P_s is a probability distribution on any finite list of play coordinates. For each ($t \ge 1$ and) fixed point $(b_1, \ldots, b_{t-1}) \in A^{t-1}$, we denote by $D_s^J(b_1, \ldots, b_{t-1})$ the P_s conditional distribution of a_t^J given $(a_1, \ldots, a_{t-1}) = (b_1, \ldots, b_{t-1})$. Note that for each $(b_1, \ldots, b_{t-1}) \in A^{t-1}$, $D_s^J(b_1, \ldots, b_{t-1})$ is a probability distribution on A_J, and thus $(a_1, \ldots, a_{t-1}) \mapsto D_s^J(a_1, \ldots, a_{t-1})$ is a random variable defined over the space of plays of the repeated game (and with values in the distributions over A_J).

Similarly, each correlated strategy $\sigma \in \Delta(BR_b^J(m) \times \Sigma^{N\setminus J})$, which is a distribution over all $s \in BR_b^J(m) \times \Sigma^{N\setminus J}$, defines the probability distribution $P_\sigma = \int P_s \, d\sigma(s)$ on plays of the repeated game. For each ($t \ge 1$ and) fixed point $(b_1, \ldots, b_{t-1}) \in A^{t-1}$, we denote by $D_\sigma^J(b_1, \ldots, b_{t-1})$ the P_σ conditional distribution of a_t^J given $(a_1, \ldots, a_{t-1}) = (b_1, \ldots, b_{t-1})$, and thus $(a_1, \ldots, a_{t-1}) \mapsto D_\sigma^J(a_1, \ldots, a_{t-1})$ is a $\Delta(A_J)$-valued random variable defined over the space of plays of the repeated game. For each $t > k$ and $(b_1, \ldots, b_{t-1}) \in A^{t-1}$, we denote by $b_t[k]$ the string of the last k action profiles $(b_{t-k}, \ldots, b_{t-1})$.

Theorem 2.3. *Let $G = (N, A, r)$. For any positive integers m and k, there exist maps $\mu^i : A^k \to \Delta(A_i)$, $i \in N$, such that for every $J \subset N$ with $|A_J|^6 \le k/|A|^m$, correlated strategy $\sigma \in \Delta(BR_b^J(m) \times \Sigma^{N\setminus J})$, and $T \ge 2k$, we have*

$$\frac{1}{k} \sum_{t=T-k+m}^{T} E_\sigma \left\| D_\sigma^J(a_1, \ldots, a_{t-1}) - \mu^J\big(a_t[k]\big) \right\| \le \frac{|A|^{m/3}}{k^{1/3}} \left(\sqrt{\ln k - m \ln|A|} + 2|A_J|^3 + 1\right), \tag{2.5}$$

where $\mu^J(a_t[k])$ is the product distribution $\bigotimes_{j \in J} \mu^j(a_t[k])$ and $\| \ \|$ denotes the L_1-norm.

Note that the theorem asserts that (2.5) holds for any correlated strategy $\sigma \in \Delta(BR_b^J(m) \times \Sigma^{N\setminus J})$. The implication of the theorem for the special case of an uncorrelated strategy $\sigma \in (\bigtimes_{j \in J} \Delta(BR_b^j(m)) \times \Sigma^{N\setminus J})$ is, by itself, of interest.

We note that if all the players use m-recall behavioral strategies, then the memories of length m along the play, $a_t[m] = (a_{t-m}, \ldots, a_{t-1})$, constitute a stationary Markov chain. And if all the players use mixtures of such strategies, then $(a_t[m])$ is a mixture of such chains. But no such property holds in the setup of this theorem, where some players may use arbitrary strategies.

3. Proof of tightness

Note that $D_\sigma^J(a_1, \ldots, a_{t-1}) = \int D_s^J(a_1, \ldots, a_{t-1}) \, dP_\sigma(s \mid a_1, \ldots, a_{t-1})$, and since μ^J depends only upon $a_t[k]$ we get $\mu^J(a_t[k]) = \int \mu^J(a_t[k]) \, dP_\sigma(s \mid a_1, \ldots, a_{t-1})$. Therefore by the triangle inequality

$$\begin{aligned} &\left\| D_\sigma^J(a_1, \ldots, a_{t-1}) - \mu^J\big(a_t[k]\big) \right\| \\ &\quad \le \int \left\| D_s^J(a_1, \ldots, a_{t-1}) - \mu^J\big(a_t[k]\big) \right\| dP_\sigma(s \mid a_1, \ldots, a_{t-1}). \end{aligned}$$

Therefore, it suffices to prove that (2.5) holds for any $\sigma = s \in BR_b^J(m) \times \Sigma^{N\setminus J}$.

In what follows we will use the following application of Hoeffding's inequality. Let $X_1, X_2, \ldots$ be a sequence of $\{0,1\}$-valued i.i.d. random variables, $S_d := \frac{X_1 + \ldots + X_d}{d}$. Hoeffding's inequality (Hoeffding, 1963, Theorem 2) implies that for every fixed positive integer d and $\delta > 0$, $P(|S_d - EX_1| > \delta) \le 2\exp(-2\delta^2 d)$. Therefore, for every positive real number $\ell > 0$, $P(\exists \ell \le d \in \mathbb{N} \text{ s.t. } |S_d - EX_1| > \delta) \le \sum_{d: \ell \le d \in \mathbb{N}} 2\exp(-2\delta^2 d) \le 2\exp(-2\delta^2 \ell)/(1 - \exp(-2\delta^2)) \le \frac{2}{\delta^2}\exp(-2\delta^2 \ell)$ for $\delta > 0$ sufficiently small, e.g., for $\delta \le \sqrt{1/2}$.

Therefore, if d is an integer-valued random variable (which can be correlated to $X_1, X_2, \ldots$), then

$$P\big(|S_d - EX_1| > \delta\big) \le P(d < \ell) + \frac{2}{\delta^2}\exp\big(-2\delta^2 \ell\big) \quad \text{for } \delta \le \sqrt{1/2}. \tag{3.1}$$

Let $s \in BR_b^J(m) \times \Sigma^{N\setminus J}$. Let $1 \le \ell \le k/|A|^m$. (Eventually, we will set $\ell = |A_J|^2(k/|A|^m)^{2/3}$.) Recall that for every $t > m$, $a_t[m]$ is the A^m-valued random variable that is defined on the space of plays by $a_t[m] = (a_{t-m}, \dots, a_{t-1})$. For $t > k$ and $f \in A^m$, let $L(t, f)$ and ℓ_t be the random variables defined on the space of plays by $L(t, f) = \{m + t - k \le d < t : a_d[m] = f\}$ and $\ell_t = |L(t, a_t[m])|$ (i.e., l_t counts the number of past appearances of the current m-length recall $a_t[m]$ in its k-recalled stages).

For every integer $T \ge 2k$ and $f \in A^m$, $|\{T - k + m \le t \le T : 0 \le \ell_t < \ell \text{ and } a_t[m] = f\}| \le \ell$. Therefore, $|\{T - k + m \le t \le T : 0 \le \ell_t < \ell\}| = |\bigcup_{f \in A^m}\{T - k + m \le t \le T : 0 \le \ell_t < \ell \text{ and } a_t[m] = f\}| \le \ell|A|^m$. Therefore, for $T \ge 2k$,

$$\sum_{t=T-k+m}^{T} P_s(\ell_t < \ell) \le \ell|A|^m. \tag{3.2}$$

We define $\mu^j : A^k \to \Delta(A_j)$ as follows. Given an element $a = (a_1, \dots, a_k) \in A^k$ and an element $b^j \in A_j$, $\mu^j(a)(b^j)$ is defined as the fraction of times that the action b^j of player j followed an appearance of $a_{k+1}[m]$ in a, i.e., $\mu^j(a)(b^j) = \frac{1}{\ell_{k+1}(a)}|\{d : m + 1 \le d \le k, a_d(m) = a_{k+1}[m], \text{ and } a_d^j = b^j\}|$ if $\ell_{k+1}(a) > 0$, and otherwise $\mu^j(a)(b^j)$ is arbitrary, e.g., $= 1/|A_j|$.

As $\|s^J(a_t[m]) - \mu^J(a_t[k])\| = \sum_{b^J \in A_J} |s^J(a_t[m])(b^J) - \mu^J(a_t[k])(b^J)|$, if $\|s^J(a_t[m]) - \mu^J(a_t[k])\| \ge \delta$ then there is $b^J \in A_J$ such that $|s^J(a_t[m])(b^J) - \mu^J(a_t[k])(b^J)| \ge \delta/|A_J|$. Therefore,

$$P_s\big(\big\|s^J\big(a_t[m]\big) - \mu^J\big(a_t[k]\big)\big\| \ge \delta\big) \le \sum_{b^J \in A_J} P_s\left(\big|s^J\big(a_t[m]\big)\big(b^J\big) - \mu^J\big(a_t[k]\big)\big(b^J\big)\big| \ge \frac{\delta}{|A_J|}\right).$$

Therefore, by inequality (3.1), for any $\ell > 1$, $\delta = |A_J|\frac{\sqrt{\ln \ell}}{\sqrt{\ell}}$, and $t > k$,

$$\begin{aligned} P_s\big(\big\|s^J\big(a_t[m]\big) - \mu^J\big(a_t[k]\big)\big\| \ge \delta\big) &\le |A_J|P_s(\ell_t < \ell) + |A_J|\frac{2|A_J|^2}{\delta^2}\exp\left(\frac{-2\delta^2}{|A_J|^2}\ell\right) \\ &= |A_J|P_s(\ell_t < \ell) + \frac{2|A_J|\ell}{\ln \ell}\exp\left(-\frac{2\ln \ell}{\ell}\ell\right) \\ &= |A_J|P_s(\ell_t < \ell) + \frac{2|A_J|}{\ell \ln \ell}. \end{aligned}$$

As $\|s^J(a_t[m]) - \mu^J(a_t[k])\| \le 2$, we deduce that

$$E_s\big(\big\|s^J\big(a_t[m]\big) - \mu^J\big(a_t[k]\big)\big\|\big) \le \delta + 2|A_J|P_s(\ell_t \le \ell) + \frac{4|A_J|}{\ell \ln \ell}.$$

Set $\ell = k^{2/3}|A_J|^2/|A|^{2m/3}$. Then, using $k/|A|^m \ge |A_J|^6$,

$$\delta = \frac{|A|^{m/3}}{k^{1/3}}\sqrt{\frac{2}{3}\ln k - \frac{2m}{3}\ln|A| + 2\ln|A_J|} \le \frac{|A|^{m/3}}{k^{1/3}}\sqrt{\ln k - m\ln|A|}.$$

Using (3.2), we have

$$\frac{1}{k}\sum_{t=T-k+m}^{T} 2|A_J|P_s(\ell_t \le \ell) \le \frac{2|A_J|\ell|A|^m}{k} = 2|A_J|^3\frac{|A|^{m/3}}{k^{1/3}}.$$

As $|A_J| \ge 2$ (w.l.o.g.) and $k \ge |A|^m|A_J|^6$,

$$\frac{4|A_J|}{\ell \ln \ell} \le \frac{2|A|^{2m/3}}{k^{2/3}} \le \frac{|A|^{m/3}}{k^{1/3}}.$$

We conclude, using $D_s^J(a_1, \dots, a_{t-1}) = s^J(a_t[m])$, that

$$\begin{aligned} \frac{1}{k}\sum_{t=T-k+m}^{T} E_s\big\|D_s^J(a_1, \dots, a_{t-1}) - \mu^J\big(a_t[k]\big)\big\| &\le \delta + 2|A_J|^3\frac{|A|^{m/3}}{k^{1/3}} + \frac{|A|^{m/3}}{k^{1/3}} \\ &\le \frac{|A|^{m/3}}{k^{1/3}}\big(\sqrt{\ln k - m\ln|A|} + 2|A_J|^3 + 1\big). \end{aligned}$$

4. An example of online concealed correlation by "weak" players

Here we present an example of online concealed correlation by weak players, with the help of a stronger one, in a specific four-player game. The construction and proof of this instance contains a few of the ideas required for the general result, and we hope that it serves as a helpful introduction to the proof of the general result.

	L	R
T	−1	0
B	0	0

E

	L	R
T	0	0
B	0	−1

W

Fig. 1. Three-player matching pennies.

4.1. The game and the strategies

Consider the normal-form game in Fig. 1, called "three-player matching pennies," in which player 1, Rowena, chooses a row (Top or Bottom), player 2, Colin, chooses a column (Left or Right), and player 3, Matt, chooses a matrix (East or West). The numbers in the matrix are the payoffs of Matt. The payoffs of Rowena and Colin are not specified. The i.r.p. of Matt in this (stage) game is $-\frac{1}{4}$.[8]

Now add a fourth player, Forest, to the game. Forest's recall will be longer than Matt's, but he will have no influence on the payoffs of the stage game, or, in particular, on Matt's payoff. Therefore Matt should care only about the actions of Rowena and Colin, and his i.r.p. in the stage game remains $-\frac{1}{4}$.

Rowena and Colin will conceal the correlation of their actions from Matt, so as to bring Matt's payoff down to around $-\frac{1}{2}$, which equals Matt's minmax in correlated actions, i.e., $\min_{y\in\Delta(A_1\times A_2)}\max_{x\in A_3} r^3(y,x) = -\frac{1}{2}$. The key point is that Forest's sequence of actions assists Rowena and Colin in correlating their own actions, while these "signals" remain unintelligible to Matt.

Since we are interested in the asymptotic behavior, let the recall capacities be functions of a parameter n: player i has a recall of length $m_i(n)$, and $\forall i\ \lim_{n\to\infty} m_i(n) = \infty$. Still, $m_1 \le m_2 \le m_3 \le m_4$, and the functions m_i are assumed to retain some relations among themselves, to be specified shortly. Rowena's, Colin's, and Forest's mixed strategies will assure that Matt's maximum expected payoff approaches $-\frac{1}{2}$, as n goes to infinity.

Following is a general description of the scheme. Colin's play will approximate a long cycle of random i.i.d. actions, distributed $\frac{1}{2}-\frac{1}{2}$. Forest, who has a relatively large memory, will be able to remember the whole cycle. Forest's actions will be used by Rowena as instructions on how she should play, so that her actions will coincide with Colin's. However, if these instructions were simply the forthcoming actions of Colin, or any deterministic function of them, the correlation would not be concealed from Matt: like Rowena, he would be able to foresee Colin's actions, and play his best response against them. Therefore, Rowena chooses a random "dictionary," each entry of which translates a finite sequence (block) of Forest's actions to a block of her own actions. That is, she randomly chooses her own interpretation of Forest's instructions, and counts on Forest to figure it out. So Forest has the task of finding out which instructions (block of actions) he should play, so that Rowena's interpretation (her own block of actions) matches what is meant (Colin's block of actions). Forest does not have to know the whole dictionary chosen by Rowena – only those entries in the dictionary he actually uses. He will learn each such entry simply by consecutively trying different blocks, till he hits upon the right one.

Now we give a more detailed description of the strategies. We require that the following relations hold between the recall capacities $m_1 \le m_2 \le m_3 \le m_4$:

$$m_3 \ll m_4 \quad \left(\text{i.e., } \lim_{n\to\infty}\frac{m_3(n)}{m_4(n)} = 0\right) \tag{4.1}$$

$$\log m_3 \ll m_1, m_2 \quad \big(m_3 \text{ is “subexponential” in } m_1, m_2\big). \tag{4.2}$$

Actually, instead of (4.2), we can settle for the following two weaker requirements:

$$(m_3)^6 \ll 2^{m_2} \tag{4.2a}$$

$$(m_3)^4 \ll (m_1)^4 \cdot 2^{m_1}, \tag{4.2b}$$

which are both implied by (4.2).

The available actions for Rowena, Colin, and Matt are $A_1 = A_2 = A_3 = \{0,1\}$ (instead of $\{T,B\}$, $\{L,R\}$, and $\{E,W\}$). For simplicity, let Forest's actions be $A_4 = \{0,1,x\}$.

$$\text{Let } K(n) = \frac{m_1(n)}{2} - 1 \tag{4.3}$$

K will be the size of a block (as we consider the asymptotic behavior, we may assume w.l.o.g. that K is an integer). We can choose an integer-valued function $L(n)$ s.t. $(K+1)$ divides L, $L = c(K+1)$, and

$$L \le m_4 - K \tag{4.4}$$

$$m_3 \ll L \tag{4.5}$$

[8] For every pair of mixed actions, $(x, 1-x)$ for Rowena, and $(y, 1-y)$ for Colin, either $xy \le \frac{1}{4}$ or $(1-x)(1-y) \le \frac{1}{4}$. Thus, by playing either E or W, the payoff to Matt is at least $-\frac{1}{4}$. On the other hand, if $x = \frac{1}{2}$ and $y = \frac{1}{2}$, the expected payoff to Matt is exactly $-\frac{1}{4}$.

$$L^6 \ll 2^{m_2} \tag{4.6a}$$

$$L^4 \ll (m_1)^4 \cdot 2^{m_1}, \tag{4.6b}$$

namely, L obeys the same magnitude restrictions, compared to m_1, m_2, as m_3 does, but L's magnitude is larger than m_3 (and L is at least somewhat smaller than m_4). L will be the length of a cycle.

Let Colin choose at random an L-periodic sequence $x_1, x_2, \dots$ ($x_i \in \{0,1\}$), where the distribution of $x_1, \dots, x_L$ is the conditional distribution of an i.i.d. sequence distributed $\frac{1}{2} - \frac{1}{2}$, given that

$$\forall s,t \quad \text{s.t.} \quad 1 \le s < t \le L \quad \exists 0 \le i < m_2 \quad \text{s.t.} \quad x_{s+i} \ne x_{t+i}. \tag{4.7}$$

Colin's strategy, σ^2, is to play the chosen sequence (an alternative description of σ^2: choose at random an L-periodic sequence $x_1, x_2, \dots$ that obeys (4.7), with uniform probability over all such sequences; then play this sequence).[9]

Condition (4.7) means that no identical m_2-length memories appear twice within the period. Therefore, any m_2 consecutive terms within the period uniquely determine the next term. Hence, playing such a sequence does not require more than m_2-recall.

Rowena randomly chooses a 1–1 function $f : \{0,1\}^K \to \{0,1\}^K$ with uniform probability over all such functions. Her play depends solely on the past actions of Forest: she plays a block of K actions, as a function of the previous K-block of Forest's actions (Forest's block is identified by the x he played before the beginning of the block). Before her block, she plays an arbitrary action (say 1).

Rowena	....			1	$f(y_1 \dots y_K)$	....
Forest	....	x	$y_1 \dots y_K$			....

Thus, if Forest's block was $(y_1, \dots, y_K)$, then the block Rowena plays is $f(y_1, \dots, y_K)$. Equality (4.3) guarantees that, for every f, this will be an m_1-recall strategy (Rowena's strategy, σ^1, is mixed, according to the random choice of f).

Forest chooses some order $\mathcal{R}$ on $\{0,1\}^K$. His strategy, σ^4, is to play his blocks for Rowena to interpret, each block preceded by an x.

Rowena	......			1	$f(y_1 \dots y_K)$	....		
Colin	......			$*$	$z_1 \dots z_K$	....		
Forest	......	x	$y_1 \dots y_K$			....	x	$\alpha_1 \dots \alpha_K$
		L (brace)						↑ Current

He finds his own block that is L stages back ((4.4) assures that he can do so) and checks whether that block worked. Suppose he played there $\bar{y} = (y_1, \dots, y_K)$. This made Rowena play $f(\bar{y})$ in the next block, and suppose Colin played there $\bar{z} = (z_1, \dots, z_K)$. If $f(\bar{y}) = \bar{z}$, then Forest plays the same in the current block, i.e., $(\alpha_1, \dots, \alpha_K) = \bar{y}$. Otherwise, he plays the next block, according to his pre-defined order $\mathcal{R}$, i.e., $(\alpha_1, \dots, \alpha_K) =$ "$\bar{y} + 1$".

Eventually, for every slot of K stages within the cycle of length L, Forest hits upon the right block to play, and plays it thereafter. When this process is carried out for all the blocks in the cycle, the play of Rowena, Colin, and Forest enters a cycle of this form:

Rowena	...	$1\| \ z_1^1 \dots z_K^1$	$1\| \ z_1^2 \dots z_K^2$	...	$1\| \ z_1^c \dots z_K^c$	...
Colin	...	$*\| \ z_1^1 \dots z_K^1$	$*\| \ z_1^2 \dots z_K^2$	...	$*\| \ z_1^c \dots z_K^c$	...
Forest	...	$x\| f^{-1}(z_1^2 \dots z_K^2)$	$x\| f^{-1}(z_1^3 \dots z_K^3)$	...	$x\| f^{-1}(z_1^1 \dots z_K^1)$	...
		L (brace)				

4.2. The payoff

Now we claim that, given the strategies described above, Matt has no m_3-recall strategy that correctly "predicts" this pair of actions (i.e., plays the opposite action) more than $\frac{1}{2} + \varepsilon$ of the time, for n large enough, and therefore Matt's payoff will be $\le -\frac{1}{2} + \varepsilon$. In other words, let $\sigma^{-3} = (\sigma^1, \sigma^2, \sigma^4)$. Then,

$$\lim_{n \to \infty} \max_{\tau \in BR^3(m_3)} \bar{r}^3(\tau, \sigma^{-3}) = -\frac{1}{2}. \tag{4.8}$$

We make two main points in the proof. (1) Along the L-cycle, Colin's actions (which coincide with Rowena's actions) approximate a random i.i.d. sequence, distributed $\frac{1}{2} - \frac{1}{2}$, and Forest's actions are almost independent of this sequence.

[9] The equivalence of these two descriptions relies on the uniformity of the distribution $\frac{1}{2} - \frac{1}{2}$. The first description easily generalizes to nonuniform distributions.

(2) Due to his bounded recall, Matt is unable, at any stage, to gather too much information about the actual realization of Rowena's and Colin's strategies. For example, he is unable to learn enough about the realization of the above random sequences from the initial phase of the game, before the play of Rowena and Forest stabilizes, i.e., before Forest has learned how Rowena wants her instructions.

Colin's strategy, as described above, is to choose, with uniform probability, any sequence $x_1, \dots, x_L$ that satisfies (4.7), and play it periodically. To verify that the distribution of this sequence is arbitrarily close to an i.i.d. $\frac{1}{2} - \frac{1}{2}$ sequence, for n large enough, it suffices to show that the probability that such a sequence obeys (4.7) is arbitrarily close to 1. For this we can use the following, more general, claim (see Neyman, 1997, pp. 247–248):

Lemma 4.1. *Let $l : \mathbb{N} \to \mathbb{N}$, and let $x_1, x_2, \dots$ be an $l(n)$-periodic i.i.d. sequence, where the support of x_i contains at least two elements. Then for $0 < \alpha < 1$, s.t. for $1 \le i \ne j \le l(n)$ $\Pr(x_i = x_j) \le \alpha$, we have*

$$\Pr\big(\exists s, t \text{ s.t. } 1 \le s < t \le l(n) \text{ s.t. } \forall 0 \le i < n\ x_{s+i} = x_{t+i}\big) < l^2(n)\alpha^{[n/3]}.$$

In our case, we may view L as a function of m_2, and take $\alpha = \frac{1}{2}$, so that we get

$$\Pr(\exists s, t \text{ s.t. } 1 \le s < t \le L \text{ s.t. } \forall 0 \le i < m_2\ x_{s+i} = x_{t+i}) < \frac{L^2}{2^{[m_2/3]}}$$

and by (4.6a), this probability converges to 0.

Now, disregarding the beginnings of blocks, examine the sequence $\bar{\alpha}$ played by Forest along the L-cycle. A realization of $\bar{\alpha}$ may be almost any sequence; the only restriction is induced by the fact that f, Rowena's function, is a 1–1 function; i.e., if two distinct K-blocks of Forest's are identical, then so are Rowena's, and vice versa. In particular, any sequence in which all of Rowena's blocks are different, and likewise Forest's, may be realized. But most random sequences are like that: let $\bar{z}$ be a random L-length (more precisely, L minus the beginnings of blocks) $\frac{1}{2} - \frac{1}{2}$ i.i.d. sequence. Then, $\Pr(\text{Block } i = \text{Block } j) = 2^{-K}$; hence $\Pr(\bar{z} \text{ contains two identical blocks}) \le \binom{c}{2}2^{-K}$, where $c = \frac{L}{K+1}$ is the number of blocks, and this probability converges to 0, by (4.3) and (4.6b). Now, the sequence $x_1, \dots, x_L$ played by both Rowena and Colin approximates an i.i.d. $\frac{1}{2} - \frac{1}{2}$ sequence. Combining this with the fact that f is chosen with uniform probability, we get that the distribution of $\bar{\alpha}$ is arbitrarily close to that of a random $\frac{1}{2} - \frac{1}{2}$ i.i.d. sequence, which is independent of the sequence $x_1, \dots, x_L$.

Thus, if we view the sequences played by Rowena, Colin, and Forest along the cycle as one random variable, then the distribution of this random variable is arbitrarily close to the distribution of a random variable of the ideal form (i.e., where the actions of Rowena and Colin coincide, they are i.i.d. $\frac{1}{2} - \frac{1}{2}$, and Forest's sequence is independent of theirs). Therefore, for any strategy σ^3 of Matt, his expected payoff when playing σ^3 against this ideal sequence is close to his expected payoff when playing σ^3 here.

We have disregarded the beginnings of blocks, but as their frequency converges to 0 (since $K \to \infty$), the following claim then suffices to prove (4.8):

Claim 4.2. *Let $\sigma^{-3} = (\sigma^1, \sigma^2, \sigma^4)$ be a tuple of mixed strategies that eventually end up playing a cycle of length L, during which the actions $\bar{x}$ of Rowena and Colin coincide, the actions of Forest $\bar{\alpha}$ are independent of $\bar{x}$, and $\bar{x}$ is a random i.i.d. $\frac{1}{2} - \frac{1}{2}$ sequence. Then* $\lim_{n\to\infty} \max_{\sigma^3 \in BR^3(m_3)} \bar{r}^3(\sigma^3, \sigma^{-3}) = -\frac{1}{2}$.

Proof. Let[10] $B = (t_1, \dots, t_L)$ be any sequence of L consecutive stages that occur after the play of Rowena, Colin, and Forest stabilizes into playing the cycle. Let τ be any pure strategy (not necessarily SBR) for Matt that *begins at the beginning of the cycle B*; i.e., τ is not contingent on the history prior to B. The random variables $a_t(\tau, \sigma^{-3})$ denote the play at stage t, induced by τ and σ^{-3}. Since the actions $\bar{x}$ are independent throughout B (and independent of $\bar{\alpha}$), $\forall t \in B$ $E_{\sigma^{-3}}(r^3(a_t) \mid a_{t_1}, \dots, a_{t-1}) = -\frac{1}{2}$ ($\bar{x}$ and $\bar{\alpha}$ may very well depend on the history prior to B, but here we are concerned only with their a priori distribution). Hence, $(r^3(a_t) + \frac{1}{2})_{t\in B}$ is a sequence of bounded martingale differences; therefore Azuma's inequality (see, e.g., Alon and Spencer, 1992, p. 79) implies that for every $\varepsilon > 0$, there exists $C(\varepsilon) > 0$, s.t.

$$\Pr_{\sigma^{-3}}\left(\frac{\sum_{t\in B} r^3(a_t(\tau, \sigma^{-3}))}{L} \ge -\frac{1}{2} + \varepsilon\right) \le e^{-C(\varepsilon)\cdot L}.$$

Hence, for any finite set Θ of strategies τ as above,

$$\Pr_{\sigma^{-3}}\left(\max_{\tau\in\Theta} \frac{\sum_{t\in B} r^3(a_t(\tau, \sigma^{-3}))}{L} \ge -\frac{1}{2} + \varepsilon\right) \le |\Theta| \cdot e^{-C(\varepsilon)\cdot L}.$$

At the beginning of the game, Matt chooses a pure m_3-recall strategy σ^3. The subgame strategy following a history h, $\sigma^3 \mid h$, depends on h only via Matt's memory $\alpha(=\alpha(h))$, i.e., the last m_3 stages within h. Hence, we denote $\sigma^3 \mid h$ by $\tau(\sigma^3, \alpha)$.

[10] The proof is an adaptation of the proof in Neyman (1997, p. 238) of Ben-Porath's results for finite automata Ben-Porath (1986, 1993).

In particular, when reaching the beginning of the cycle B, Matt's subgame strategy, $\tau(\sigma^3, \alpha)$, is completely determined by the memory α he may have at this point. Let $\Theta = \Theta(\sigma^3)$ be the set of all possible strategies thus determined before B. Then,

$$\Pr_{\sigma^{-3}}\left(\frac{\sum_{t\in B} r^3(a_t(\sigma^3, \sigma^{-3}))}{L} \geq -\frac{1}{2} + \varepsilon\right) \leq |\Theta| \cdot e^{-C(\varepsilon)\cdot L}.$$

There are no more than $|A|^{m_3}$ possible memories; hence $|\Theta| \leq |A|^{m_3}$. Therefore, by (4.5), $|\Theta| \cdot e^{-C(\varepsilon)\cdot L} \to 0$, and the result follows. □

An alternative way of looking at the proof is as follows. Matt may choose σ^3 so that along the play it "encodes," through his own actions, information about the realization of σ^{-3}. His information also includes the past m_3 actions of his adversaries. However, all this information is limited, due to his bounded recall. Thus, even in the best imaginable case for him, in which σ^3 arrives at B with optimal memory, and σ^3 also makes optimal use of that memory along B, it will still do "well" only against a minor fraction of the realizations of σ^{-3}.

5. Proof of Theorem 2.1

In Section 4 we have proved one specific example, based on the three-player matching pennies stage game. Now we show how the result, that the payoff of the third player in the infinitely repeated game is not more than his minmax in correlated actions, extends to a general class of four-player games (when the appropriate relations between strength levels obtain) in which $|A_4| \geq |A_1|$ or $|A_4| \geq |A_2|$.

First, let us point out that in Section 4, the set of actions available to Forest was taken to be $A_4 = \{0, 1, x\}$ just for convenience. The extra action, x, was used to designate the beginnings of blocks. The use of this extra action can be dispensed with. Suppose Forest has only two actions, w.l.o.g. $A_4 = \{0, 1\}$. Let $p : \mathbb{N} \to \mathbb{N}$ be some integer-valued function, s.t. $2^p \gg K$, but still $p \ll K$ (recall that K is the size of a block). Let $\bar{\xi} = \xi_1, \ldots, \xi_p$ ($\xi_i \in \{0, 1\}$) be some fixed sequence. To designate beginnings of blocks, Forest plays this sequence $\bar{\xi}$ before the beginning of every block, and Rowena plays some arbitrary actions in the corresponding stages. Thus, instead of just one stage being sacrificed before every K-block, p stages are sacrificed. Since $p \ll K$, the effect on the average payoff is negligible; i.e., it approaches 0 as $n \to \infty$.

There is, however, a further sacrifice: if $\bar{\xi}$ is to designate the beginnings of blocks correctly, Forest must avoid playing this sequence at any other time. Recall that Forest's original strategy consisted of trying at every K-slot all possible $\{0, 1\}^K$ blocks consecutively, until hitting upon the right one. In Forest's modified strategy, he will not try any block that contains $\bar{\xi}$. If, in addition, no prefix of $\bar{\xi}$ equals a suffix of $\bar{\xi}$ (for example, choose $\bar{\xi}$, so that the first half of it is $1, \ldots, 1$, and the second half, $0, \ldots, 0$), then Forest will play $\bar{\xi}$ only between the blocks.

What is the second sacrifice's effect on the payoff? As we saw in Section 4.2, the sequence played by Forest is distributed almost like a random (i.i.d., uniformly distributed) sequence. Since the probability that a random K-block contains $\bar{\xi}$ is less than $K \cdot 2^{-p}$, we again get, by the choice of p, that the effect on the payoff is negligible.

To prove the result, we begin by assuming that the payoff of Matt (player 3) is independent of Forest's actions, and prove the "moreover" part of the theorem; more precisely, we show that if $\log m_4 \ll m_3$, then

$$\liminf_{n\to\infty} \underline{v}^3(\vec{m}) \geq \min_{y\in\Delta(A_1\times A_2)} \max_{x\in A_3} r^3(y, x).$$

Let us grant, in advance, perfect correlation of the actions of Rowena and Colin. Furthermore, let Rowena, Colin, and Forest be united into one player with a set of actions $A_1 \times A_2 \times A_4$, whose recall capacity is m_4. As m_4 is subexponential in m_3, then, by Theorem 1 in Lehrer (1988), in the resulting repeated two-player game Matt can guarantee an expected payoff approaching his i.r.p. in the two-player stage game. This equals $\min_{y\in\Delta(A_1\times A_2)} \max_{x\in A_3} r^3(y, x)$; hence the result.

We proceed to prove the first part of the theorem. First, as noted, Forest's set of actions, A_4, need only be as large as Rowena's, A_1, in order for him to instruct her correctly. If, however, it is as large as Colin's, then Rowena and Colin can switch roles: Rowena will play a long random cycle, and Colin will choose his "instruction dictionary," and be instructed by Forest on how to match Rowena's actions.

The strategies follow an outline similar to the scheme in Section 4.1, but for this general stage game we need to make some adjustments. Let $\bar{y} \in \Delta(A_1 \times A_2)$ be a correlated distribution that achieves the minmax in correlated actions (when Forest plays according to z), i.e., $\bar{y} \in \arg\min_{y\in\Delta(A_1\times A_2)}[\max_{x\in A_3} r^3(y, x, z)]$. Let $\bar{y}^2$ denote the marginal distribution on A_2, induced by $\bar{y}$. For $x \in A_2$, let $(\bar{y}^1|x)$ denote the conditional probability distribution on A_1, induced by $\bar{y}$, given that the outcome in A_2 is x.

Rowena chooses a random 1–1 function $f : (A_4)^K \to (A_1)^K$ to use as her instruction dictionary, and plays according to Forest's play and f, as in Section 4.1. Again, Colin chooses a random L-periodic sequence $x_1, x_2, \ldots$ ($x_i \in A_2$) that satisfies (4.7), but here each x_i is (almost independently) distributed according to $\bar{y}^2$.

Forest randomly chooses a function $g : (A_2)^K \to (A_1)^K$, $g(\bar{x} = x_1, \ldots, x_K) = (g_1(\bar{x}), \ldots, g_K(\bar{x}))$, where $g_i(\bar{x})$ is distributed according to $(\bar{y}^1|x_i)$, independently for every $\bar{x} \in (A_2)^K$ and for every i. His goal is to instruct Rowena to play the block $g(\bar{x})$ whenever Colin plays a block $\bar{x}$. As in Section 4.1, Forest tries the blocks consecutively, until he hits upon the right one. Here, the "right block" means the one that made Rowena play $g_1(\bar{x}), \ldots, g_K(\bar{x})$, when Colin played $\bar{x} = x_1, \ldots, x_K$ (here Forest's strategy is not pure but mixed according to his random choice of g). Therefore, the play of Rowena and Forest will eventually enter an L-cycle, in which Rowena plays blocks that are the function g of Colin's blocks, and Forest plays blocks that are f^{-1} of Rowena's next blocks.

Again, by Lemma 4.1, the sequence $x_1, \ldots, x_L$ played by Colin along a cycle is almost i.i.d. And again, since the probability of a K-block being repeated twice along an L-cycle is negligible, we conclude that Rowena's actions along the cycle are almost mutually independent, by the construction of g, and that Forest's sequence is almost independent of Rowena's (and of Colin's), by the random choice of f (and note that z is indeed the distribution of an action of Forest[11]). Therefore, Matt is practically faced with an L-cycle where, independently at every stage, the action pair of Rowena and Colin is distributed according to $\bar{y}$, and Forest's actions convey no information. The appropriate rephrasing of Claim 4.2 completes the proof.

6. Proof of Theorem 2.2

Let D^{-k} be a distribution on A_{-k} with marginals D^j on A_j and assume that $H(D^{-k}) \geq \sum_{j<k} H(D^j)$. W.l.o.g. we assume that $H(D^j) > 0$ for every player j. Otherwise, if $H(D^j) = 0$ then there is an action $a^j \in A_j$ such that $D^j(a^j) = 1$, and by player j playing repeatedly the action a^j we reduce the problem to one of a game with a set of players $N \setminus \{j\}$.

Let $J = \{1, \ldots, k-1\}$ and $J_+ = \{k+1, \ldots, |N|\}$. If $J_+ = \emptyset$, then $\sum_{j \in J} H(D^j) \geq H(D^J) = H(D^{-k}) \geq \sum_{j \in J} H(D^j)$, and therefore $H(D^J) = \sum_{j \in J} H(D^j)$; hence, the distribution D^J is a product distribution.

The result in this special case appears implicitly in the proof of Lehrer (1994, Theorem 2); see also Neyman (1997, Proposition 6).

Assume that J_+ is nonempty. The strategies of players $N \setminus \{k\}$ (or simply $-k$) will be designed so that their play $(a_t^{-k})_{t \geq 1}$ will eventually enter a periodic play. The length of the period will be a product of two integers dL with $L = m_1/2$, $m_k \ll dL \ll m_{k+1}$, and $\log d \ll m_1$. For example, as (by assumption) m_k is subexponential in m_1 and $m_k \ll m_{k+1}$, we can select $m_k \ll \bar{m}_{k+1} \leq m_{k+1}$ that is subexponential in m_1, and let d be the largest integer that is $\leq \sqrt{\bar{m}_{k+1} m_k}/m_1$.

The play will be partitioned into superblocks of length dL. The ℓ-th superblock consists of the play in stages $(\ell-1)dL + 1, \ldots, \ell dL$, and each superblock is partitioned into d blocks; the i-th block (where $1 \leq i \leq d$) of the ℓ-th superblock consists of the play in stages $(\ell-1)dL + (i-1)L + 1, \ldots, (\ell-1)dL + iL$.

The condition $dL \ll m_{k+1}$ guarantees that (for sufficiently large n) player $j > k$ recalls many of the recent completed superblocks. The condition $L = m_1/2$ guarantees that each player $j < k$ can recall the play of the last completed block.

The action choices of the strategy of a player $j \neq k$ will not rely on the past actions of player k, and those of a strategy of a player $j < k$ will depend only on the past actions of the players in J_+.

We choose a sequence $\varepsilon = (\varepsilon(n))_{n=1}^{\infty}$ such that $\frac{\log dL}{L} \ll \varepsilon^2$ (where $\varepsilon^2 = (\varepsilon^2(n))_{n=1}^{\infty}$) and $0 < \varepsilon(n) \to_{n\to\infty} 0$, and let m be the largest integer that is $\leq L/(1+2\varepsilon)$. For a sequence $x = (x_1, \ldots, x_L)$ we denote by x^* the (ordered) vector of the first m elements x_s of the sequence where s is not an integer multiple of $[\varepsilon m/2]$.

The condition that $\varepsilon(n) \to_{n\to\infty} 0$ guarantees that the play in the last $L - m$ stages of a block has a negligible (as $n \to \infty$) impact on the empirical distribution of the block play.

The relation $\frac{\log dL}{L} \ll \varepsilon$ (which follows from the condition $\frac{\log dL}{L} \ll \varepsilon^2$ and $\varepsilon(n) \to_{n\to\infty} 0$) together with the choice of m, imply that $\varepsilon m \gg \log d$; thus $\varepsilon m - \log d \to_{n\to\infty} \infty$, and therefore $d 2^{-m\varepsilon} \to_{n\to\infty} 0$.

The relation $\varepsilon(n) m(n) \to_{n\to\infty} \infty$ (which follows from $\varepsilon m - \log d \to_{n\to\infty} \infty$) guarantees that the play in the stages that are integer multiples of $[\varepsilon m/2]$ has a negligible (as $n \to \infty$) impact on the empirical distribution of the block play.

The role of stages $m + [\varepsilon m] + 1, \ldots, m + [\varepsilon m] + \lceil \log d \rceil$ of a block is to enable player $k+1$ to signal the index $1 \leq i \leq d$ of a block within the superblock. The relation $\frac{\log dL}{L} \ll \varepsilon$ guarantees that $L - m - [\varepsilon m] \gg$ than the number $\lceil \log d \rceil$ of stages that are needed for this.

The role of stages $m + [\varepsilon m] + \lceil \log d \rceil + 1, \ldots, L$ is to (enable player $k+1$ to) signal the end of the block by a sequence of $L - m - [\varepsilon m] - \lceil \log d \rceil - 1$ repetitions of a fixed action a followed by an action $b \neq a$. The relation $\frac{\log dL}{L} \ll \varepsilon$ guarantees that for sufficiently large n this number of repetitions of the action a is larger than $[\varepsilon m/2] + \lceil \log d \rceil$; hence, if the play (of player $k+1$) in stages $t \leq m + [\varepsilon m]$ (of a block) that are integer multiples of $[\varepsilon m/2]$ is the action b, then the specified sequence $(a, \ldots, a, b)$ will appear only at an end of a block.

Let m^* be the smallest integer such that there are m positive integers that are $\leq m^*$ and are not integer multiples of $[\varepsilon m/2]$. The relation $\frac{\log dL}{L} \ll \varepsilon^2$ guarantees that $m^* - m \ll \varepsilon m$. Therefore, the set S_+ of all integers $m^* + 1 \leq s \leq m + [\varepsilon m]$ that are not integer multiples of $[\varepsilon m/2]$ has, for sufficiently large n, more than $\varepsilon m/2$ elements. Stages $s \in S_+$ of a block will be used by player $k+1$ for additional signaling. As $|A_{k+1}| \geq 2$, we deduce that for sufficiently large n, $|A_{k+1}^{S_+}| \geq 2^{m\varepsilon/2}$.

Let Q be a distribution on A_{-k} such that for every $a \in A_{-k}$, $mQ(a)$ is an integer and $|Q(a) - D^{-k}(a)| < 1/m$. For every subset $C \subset N \setminus \{k\}$ we denote by Q^C the marginal of Q on $A_C = \bigtimes_{j \in C} A_j$. It follows that

$$H(D^C) - O\left(\frac{\log m}{m}\right) \leq H(Q^C) \leq H(D^C) + O\left(\frac{\log m}{m}\right). \tag{6.1}$$

[11] Actually, we could have replaced z in the theorem by any distribution ζ, provided that $H(\zeta) \geq \log_2 |A_1|$, where H signifies the entropy.

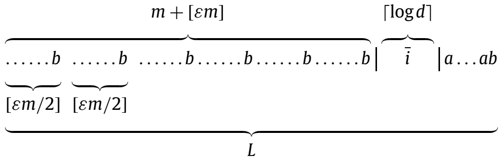

Fig. 2. Block play of player $k+1$.

For a positive integer m and a probability distribution q on a finite set B, $\mathbb{T}^m(q)$ denotes all the m-length sequences $b \in B^m$ with empirical distribution q. In the sequel we use the following estimate (see, for instance, Cover and Thomas, 1991, Theorem 12.1.13, p. 282) of the number of elements in $\mathbb{T}^m(q)$.

$$\frac{2^{mH(q)}}{(m+1)^{|B|}} \le \left|\mathbb{T}^m(q)\right| \le 2^{mH(q)} \quad \text{if } \mathbb{T}^m(q) \neq \emptyset. \tag{6.2}$$

Let Y_j be the set of all elements $x \in A_j^L$ such that $x^* \in \mathbb{T}^m(Q^j)$ and let Y_J be the set of all elements $x \in A_J^L$ such that $x^* \in \mathbb{T}^m(Q^J)$.

The size of Y_j. As $\mathbb{T}^m(Q^j)$ is nonempty, (6.2) implies that

$$|A_j|^{L-m} \frac{2^{mH(Q^j)}}{(m+1)^{|A_j|}} \le |Y_j| \le |A_j|^{L-m} 2^{mH(Q^j)}. \tag{6.3}$$

The size of Y_J. Similarly,

$$|A_J|^{L-m} \frac{2^{mH(Q^J)}}{(m+1)^{|A_J|}} \le |Y_J| \le |A_J|^{L-m} 2^{mH(Q^J)}. \tag{6.4}$$

We impose additional properties on the play of player $k+1$. These additional properties allow player $k+1$ to signal the end of a block and its index.

Recall that $H(D^{k+1}) > 0$ and therefore $|A_{k+1}| \ge 2$. Let a and b be two distinct actions in A_{k+1}. Let $\{\bar{i} : 1 \le i \le d\}$ be a set of d distinct elements of $\{a, b\}^{\lceil \log_2 d \rceil}$. Let $\vec{a}$ be a sequence of $L - m - [\varepsilon m] - \lceil \log_2 d \rceil - 1$ repeated actions a. Let $X_i(k+1)$, $1 \le i \le d$, be all elements x in A_{k+1}^L of the form $x = (\ldots, \bar{i}, \vec{a}, b)$ such that $x^* \in \mathbb{T}^m(Q^{k+1})$ and $x_s = b$ for every stage $1 \le s \le m + [\varepsilon m]$ that is an integer multiple of $[\varepsilon m/2]$ (see Fig. 2). Note that for sufficiently large m, the string $(\vec{a}, b)$ appears in a sequence $x \in X_i(k+1)$ only once (at the end).

The size of $X_i(k+1)$. Recall that $|A_{k+1}| \ge 2$ and that for m sufficiently large, $|S_+| \ge \varepsilon m/2$. Therefore, for m sufficiently large,

$$\left|X_i(k+1)\right| \ge \frac{2^{m(H(Q^{k+1})+\varepsilon/2)}}{(m+1)^{|A_{k+1}|}}. \tag{6.5}$$

The sets $X_i(k+1)$, $1 \le i \le d$, are disjoint and we set $X(k+1) = \bigcup_i X_i(k+1)$. For every $k+1 \in C \subset N$ let $\gamma : A_C^L \to \{0, 1, \ldots, d\}$ be the surjective map defined by $\gamma(x) = i \ge 1$ if $x^{k+1} \in X_i(k+1)$, and $\gamma(x) = 0$ otherwise.

Let $C \subset N \setminus \{k\}$ be a set of players with $k+1 \in C$. A play $(z_1, \ldots, z_d)$ of the set C of players, where $z_i \in A_C^L$, is said to be *C-correct*, if, for every $1 < i \le d$, the empirical distribution of z_i^* equals Q^C, and for every $1 \le i \le d$, $z_i^{k+1} \in X_i(k+1)$. A *correct* play is an $N \setminus \{k\}$-correct play.

The strategy profile $\sigma^{-k} = (\sigma^j)_{j \neq k}$ aims at generating a distribution over plays, such that with probability close to 1, it eventually repeats a correct play, and such that all correct plays are equally likely in the eventually repeated superblock play.

This aim is achieved by the description below of the strategy profile $\sigma^{J_+} = (\sigma^j)_{j \in J_+}$ of the set of players J_+ and of the strategy σ^j of each player $j < k$.

The strategy profile $\sigma^{J_+} = (\sigma^j)_{j \in J_+}$ tries, in a random order, all the J_+-correct plays in a superblock, until it observes that the play in a superblock is correct. Thereafter, it repeats its play in the correct superblock at each of the following superblocks.

A $(k+1)$-correct play marks the end of each block so that the recall of each player $j < k$ enables him to recall the play x of the last completed block play. The strategy of player $j < k$ selects for each $0 \le i \le d$ a dictionary $f_i^j : A_{J_+}^L \to Y_j$, where each list $f^j = (f_i^j)_{1 \le i \le d}$ is equally likely, and following a play x of players $-k$ in a block, it plays $f_{\gamma(x)}^j(x)$ in the following block (see Fig. 3). Let $f_i = (f_i^j)_{j<k}$ denote the function $f_i : A_{J_+}^L \to Z := \prod_{j<k} Y_j$ where $f_i(x) = (f_i^j(x))_{j<k}$. For completeness of the definition of the strategy, we also have to define its play in the first block, as well as its play when it does not recognize the last completed block. The strategy of player $j < k$ plays in the first block a random play in A_j^L, all equally likely. In all other cases it plays a fixed action.

The random order of "trials" of σ^{J_+} and the uniform randomness of the dictionaries f_j^i imply that the distribution defined by the strategy profile σ^{-k} on plays of the repeated game is such that (conditional to the infinite play entering a cycle of a correct play) all correct plays are equally likely.

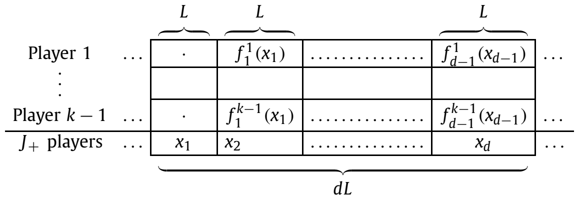

Fig. 3. Correct superblock play.

The trials of all the J_+-correct plays in a superblock is straightforward in the case where J_+ contains a single element, namely, $J_+ = \{k+1\}$. In this case, player $k+1$ selects a random order $\mathcal{O}$ of all $(k+1)$-correct plays. As the recall m_{k+1} of player $k+1$ is greater than $2dL$, player $k+1$ has the capacity to recall at every stage the entire last completed play of a superblock. If the last completed superblock is correct, he repeats his $(k+1)$-correct play of the previous superblock. Otherwise, he tries in the following superblock to play the next (according to the order $\mathcal{O}$) $(k+1)$-correct play.

We now describe the random trials in the case that $J_+ \supsetneq \{k+1\}$. Set $X_0(k+1) = A^L_{k+1} \setminus \bigcup_{i=1}^d X_i(k+1)$. Let w be the smallest integer w that is an integer multiple of d and such that $|X_0(k+1)|^w$ is larger than the number of J_+-correct plays. Note that $w = O(d)$ as $n \to \infty$, and therefore each one of the players $j > k$ recalls the play of the last $wL + dL$ stages. Let the set J_+ of players agree on a dictionary that maps $X_0(k+1)^w$ onto the set of J_+-correct plays. Player $k+1$ selects a random order of all J_+-correct plays, and in wL consecutive stages (namely, w blocks) player $k+1$ plays an element of $X_0(k+1)^w$ to signal via the agreed-upon dictionary the J_+-correct play of the next trial.

We turn now to the proof that with probability close to 1 there is a play $(x_1, \ldots, x_d) \in A^{dL}_{J_+}$ (where $x_i \in A^L_{J_+}$) such that $(x_1, y_1, \ldots, x_d, y_d)$, where $y_{i+1} = f_i(x_i)$ for $i < d$, is correct. Let $Y = Y_J$, X_1 is the set of all $x \in A^L_{J_+}$ such that $\gamma(x^{k+1}) = 1$, and for $1 < i \le d$ and $y \in Y$, X_i is the set of all $x \in A^L_{J_+}$ such that $x^{k+1} \in X_i(k+1)$, and $X_i(y)$ is the set of all $x \in X_i$ such that $(x^*, y^*) \in \mathbb{T}^m(Q)$.

Size of $X_i(y)$.

$$\left|X_i(y)\right| \ge 2^{m(H(Q)-H(Q^J)+\varepsilon/2-O(\frac{\log m}{m}))}. \tag{6.6}$$

As $\frac{\log m}{m} \ll \varepsilon$, $\varepsilon/2 - O(\frac{\log m}{m}) \ge \varepsilon/3$ for sufficiently large m.

We use the following auxiliary concept. An element $x_1 \in X_1$ is f-correct if $f_1(x_1) \in Y$. By induction on $1 < i < d$, we say that an element $(x_1, \ldots, x_i) \in \prod_{1 \le i' \le i} X_{i'}$ is f-correct if $(x_1, \ldots, x_{i-1})$ is f-correct, $x_i \in X_i(f_{i-1}(x_{i-1}))$ and $f_i(x_i) \in Y$.

By inequalities (6.3) and (6.4), $|Y|/|Z| = 2^{m(H(Q^J)-\sum_{j<k} H(Q^j)+O(\frac{\log m}{m}))}$ as $m \to \infty$. Therefore, for sufficiently large m, using inequality (6.6), the conditional probability, given that $(x_1, \ldots, x_{i-1})$ is f-correct, that there is no $x_i \in X_i(y_i)$, where $y_i = f_i(x_{i-1})$, such that $(x_1, \ldots, x_i)$ is f-correct, is

$$\begin{aligned} &\le \left(1 - |Y|/|Z|\right)^{|X_i(y_i)|} \le e^{-2^{m(H(Q)+\varepsilon/2-H(Q^J)+H(Q^J)-\sum_{j<k} H(Q^j)+O(\frac{\log m}{m}))}} \\ &\le e^{-2^{\varepsilon m/3}} \le 2^{-\varepsilon m} \quad \text{for } m \text{ sufficiently large.} \end{aligned}$$

Therefore, the probability that there is no $(x_1, \ldots, x_d) \in A^{dL}_{J_+}$ such that $(x_1, y_1, \ldots, x_d, y_d)$, where $y_{i+1} = f_i(x_i)$ for $i < d$, is correct, is $\le d2^{-\varepsilon m} \to_{n\to\infty} 0$.

Recall that d, m, and ε are functions of n, with $\frac{\log m}{m} \ll \varepsilon \ll 1 \ll d$. Let $\varepsilon_1 = 2d2^{-\varepsilon m}$. Fix a sufficiently large T_n such that for all sufficiently large n and any $T \ge T_n$ that is an integer multiple of dL,

$$P_{\sigma^{-k}}\left(\left(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}\right) \text{ is a correct play}\right) > 1 - \varepsilon_1.$$

Note that if the play $(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL})$ is correct, its empirical distribution $e(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL})$ is close to Q; explicitly, $\|e(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}) - Q\| \le \frac{1}{d}\sum_{i=1}^d \|e(a^{-k}_{T+(i-1)L+1}, \ldots, a^{-k}_{T+iL}) - Q\| \le \frac{2}{d} + 2\frac{L-m}{m} \le \frac{2}{d} + 5\varepsilon \ll 1$. As $\|Q - D^{-k}\| \le O(\frac{\log m}{m}) \ll \varepsilon$, we conclude that if the play $(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL})$ is correct, then $\|e(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}) - D^{-k}\| \le \frac{2}{d} + 6\varepsilon \ll 1$ for n sufficiently large. Therefore, for n sufficiently large, $P_{\sigma^{-k}}(\|e(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}) - D^{-k}\| > \frac{2}{d} + 6\varepsilon) < \varepsilon_1$ and therefore

$$E_{\sigma^{-k}}\left(e\left(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}\right)\right) \to_{n\to\infty} D^{-k}. \tag{6.7}$$

By the concavity and continuity of the entropy (as a function of the distribution), (6.7) implies that

$$\limsup_{n\to\infty} \frac{1}{dL} H\left(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}\right) \le H\left(D^{-k}\right).$$

The number of correct plays is $\ge 2^{dmH(Q)}/(m+1)^{d|A_{-k}|}$, and all correct plays are equally likely. Therefore $H(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}) \ge (1-\varepsilon_1)mdH(Q) - d|A_{-k}|\log(m+1)$. Using the continuity of the entropy function, $H(Q) \to_{n\to\infty} H(D^{-k})$, and therefore, as $\frac{(1-\varepsilon_1)mdH(Q)-d|A_{-k}|\log(m+1)}{dL} \to_{n\to\infty} H(D^{-k})$, we deduce that $\liminf_{n\to\infty} \frac{1}{dL} H(a^{-k}_{T+1}, \ldots, a^{-k}_{T+dL}) \ge H(D^{-k})$. We conclude that

$$\frac{1}{dL}H\big(a^{-k}_{T+1},\dots,a^{-k}_{T+dL}\big)\to_{n\to\infty}H\big(D^{-k}\big). \tag{6.8}$$

Let σ^k be a pure strategy of the opponent. The number of strategies $(\sigma^k \mid h)$, where h ranges over all plays $(z_1,\dots,z_T)\in A^T$, is bounded by $|A|^{m_k}$.

Following Neyman (2008), for every $q\in\Delta(A_{-k})$ and $\alpha>0$ we define $v(q,\alpha)$ as the maximum of $E_Q r^k(a)$, where the maximum ranges over all distributions Q on A such that its marginal Q^{-k} on A_{-k} equals q and $H(Q^{-k})+H(Q^k)-H(Q)\le\alpha$. It follows that $v(q,\alpha)\to\max_{a^k\in A_k}r^k(D^{-k},a^k)$ as $(q,\alpha)\to(D^{-k},0)$. Therefore, as (1) $\frac{m_k\log|A|}{dL}\to_{n\to\infty}0$, (2) $E_{\sigma^{-k}}(e(a^{-k}_{T+1},\dots,a^{-k}_{T+dL}))\to_{n\to\infty}D^{-k}$, and (3) $\frac{1}{dL}H(a^{-k}_{T+1},\dots,a^{-k}_{T+dL})\to_{n\to\infty}H(D^{-k})$, we deduce from Neyman (2008, Proposition 2) that inequality (2.4) holds.

7. Equilibrium payoffs

A classical "folk theorem" characterizes the set of equilibrium payoffs in the infinitely repeated game by means of the data of the stage game. Other "folk theorems" characterize the asymptotic behavior of the equilibrium payoffs of game models that "approximate" the undiscounted infinitely repeated game model. Such approximations may involve variations on the duration (e.g., a long finitely repeated game), or discounted games with patient players, or repeated games with bounded rationality (e.g., repeated games with finite automata or with bounded recall), or any combination of such approximations. The present section focuses on the folk theorems for repeated games with bounded recall.

Recall that for $\vec{m}=(\vec{m}(n))_{n\in\mathbb{N}}$, where $\vec{m}(n)=(m_i(n))_{i\in N}$, $G_\infty(\vec{m})$ denotes the infinitely repeated game, where player i uses an $m_i(n)$-recall strategy. A bounded-recall folk theorem here is a characterization of the asymptotic behavior of $NE(G_\infty(\vec{m}))$, the set of Nash equilibrium payoffs of the game $G_\infty(\vec{m})$.

The existence of the limit, as $n\to\infty$, of $NE(G_\infty(\vec{m}))$ depends obviously on the asymptotic properties of the sequence $\vec{m}$ of the players' recall. In some cases, where the existence of a limit is not known, information about the asymptotic behavior of $NE(G_\infty(\vec{m}))$ is provided by exhibiting upper and lower bounds for $NE(G_\infty(\vec{m}))$. Obviously, the smaller the upper bound and the larger the lower bound, the more informative such results are.

The asymptotic behavior of $NE(G_\infty(\vec{m}))$ in the case of two-player games ($N=\{1,2\}$), where each player's length of recall is subexponential in the other player's length of recall (i.e., $m_i\gg\log m_j$) and the length of recall $m_i(n)$ goes to infinity as n goes to infinity, is characterized in Lehrer (1988).[12] For more than two players, the issue is wide open; existing theorems (e.g., Lehrer, 1994, Neyman, 1997) give upper and lower bounds for $NE(G_\infty(\vec{m}))$. The asymptotic analysis of this set essentially boils down to the asymptotic analysis of the i.r.p.s of the players, $(\bar{v}^i(\vec{m}))_{i\in N}$. To find these i.r.p.s, we need to account for the possibilities of concealed correlation. Therefore, our main results contribute to the asymptotic analysis of $NE(G_\infty(\vec{m}))$.

Given a stage game $G=(N,A,r)$, we denote by $F=F(G)$ the convex hull of all points $r(a)$, $a\in A$ (then F is the set of feasible payoffs in the infinitely repeated (undiscounted) game), and $u^i=u^i(G)$ denote player i's minmax in correlated actions in G, i.e., $u^i(G)=\min_{y\in\Delta(A_{-i})}\max_{a^i\in A_i}r^i(y,a^i)$ $(=\max_{x^i\in\Delta(A_i)}\min_{b^{-i}\in A_{-i}}r^i(x,b^{-i}))$.

Theorem 7.1. *Let G be a $(k+1)$-player game, where (1) player $k+1$ has sufficiently many stage actions, e.g., $|A_{k+1}|\ge|A_S|$ for every $S\subset\{1,\dots,k\}$ with $|S|\le k-2$, (2) r^{k+1} is constant, and (3) r^i is independent of the actions of player $k+1$, and there exists a vector payoff $v\in F(G)$ such that for every player $1\le i\le k$, $v^i>u^i(G)$. Let $\vec{m}$ be a sequence of recall lengths $m_i(n)$ that satisfy assumptions (A1), (A2), and (A3). Then,*

$$\liminf_{n\to\infty}NE\big(G_\infty(\vec{m})\big)\supset\big\{v\in F(G): v^i\ge u^i(G)\big\}. \tag{7.1}$$

If, in addition, $\log m_{k+1}\ll m_1$, *then*

$$\lim_{n\to\infty}NE\big(G_\infty(\vec{m})\big)=\big\{v\in F(G): v^i\ge u^i(G)\big\}. \tag{7.2}$$

The theorem follows from the following observations. First, for any player $1\le i\le k$ and $D^S\in\Delta(A_S)$, where $S=\{1,\dots,k\}\setminus\{i\}$, if z is the uniform distribution on A_{k+1} and $p\in S$, then $H(D^S\otimes z)=H(D^S)+H(z)\ge H(D^p)+\log|A_{k+1}|\ge H(D^p)+\log|A_{S\setminus\{p\}}|\ge\sum_{j\in S}H(D^j)$. Therefore, by Theorem 2.2, players $S\cup\{k+1\}$ can conceal the distribution $D^S\otimes z$ from player i. Therefore, $\limsup_{n\to\infty}\bar{v}^i(\vec{m})\le u^i(G)$, and the classical pattern of proving a folk theorem – a plan with punishments – yields (7.1). Second, the condition $\log m_{k+1}\ll m_1$ implies that each player i can conceal from the other players any distribution D^i on A_i. Therefore, $\liminf_{n\to\infty}\underline{v}^i(\vec{m})\ge u^i(G)$ which, together with (7.1), implies (7.2).

Interpreting player $k+1$ in Theorem 7.1 as a mechanism designer, rather than a participant, yields a folk theorem for repeated games with bounded recall and a mechanism designer. In Eq. (7.2), the set of equilibrium payoffs of players $1,\dots,k$ converges to the set of correlated equilibrium payoffs in a classical k-player repeated game, without any rationality bounds (this is the "Correlated Folk Theorem").

[12] The analogous (and related) result in the model of repeated games with finite automata appears in Ben-Porath (1986, 1993).

When player $k+1$ is a participant in the game, rather than a mechanism designer, Theorem 2.2 specifies a set of distributions that can be concealed from a player in the repeated game with bounded recall. As a corollary, this narrows the gap between the known upper and lower bounds for the equilibrium payoffs in these games.

When the lengths of the players' recalls satisfy assumption (A1), namely, $m_{i+1}(n) \geq m_i(n) \to_{n\to\infty} \infty$, and $m_k \gg |A|^{m_{k-1}}$, then Theorem 2.3 yields a lower bound on $\liminf \underline{v}^k(\vec{m})$:

$$\liminf \underline{v}^k(\vec{m}) \geq \min_{a\in A_{<k}} \max_{b\in A_k} \min_{c\in A_{>k}} r^k(a,b,c),$$

where $<k$, respectively $>k$, is the set of all players i with $i<k$, respectively with $i>k$.

8. Remarks

8.1. Concealed distributions

A profile of mixed SBR (or finite automata) strategies $\sigma = (\sigma^1,\dots,\sigma^n)$ induces a probability distribution over periodic plays of the repeated game, thereby inducing the average limiting frequency of any action profile $a=(a^1,\dots,a^n)\in A$ (the average "empirical" probability of a). Thus, σ induces the average empirical distribution $D_\sigma \in \Delta(A)$. For $K\subset N$ let $A_K = \bigtimes_{k\in K} A_k$ and let $D_\sigma \mid A_K$ denote the marginal distribution of D_σ on A_K (the average empirical distribution over A_K).

We now define a notion of strategic concealment that is independent of payoff. Let $J\subset N$ be a group of players, let $D^J \in \Delta(A_J)$ be a given distribution over their actions, let player $i\notin J$ be the opponent, and let $\mathcal{C}^i$ be a class of strategies of i. We would say that the strategy tuple σ^{-i} of players $N\setminus i$ conceals the distribution (over the actions of the members of J) D^J against $\mathcal{C}^i$, if the following holds: for every $\sigma^i\in\mathcal{C}^i$ and every action $a^i\in A_i$ whose average empirical distribution is nonnegligible, the average empirical conditional distribution over A_J, given that i plays a^i, is close to D^J.

This notion may be formally and succinctly defined using information-theoretic terminology as follows. We use the notion of the Relative Entropy of a distribution p with respect to a distribution q, $\mathfrak{D}(p\parallel q)$ (see, e.g., Cover and Thomas, 1991), to measure how much p differs from q. (This choice is quite immaterial for our purposes; other notions, e.g., the bounded variation norm, would serve just as well.)

First we define the notion of implementation. Fix a player $i\in N$ and let $J\subset N\setminus\{i\}$.

Definition 8.1. σ^{-i} ε-implements D^J *against* $\mathcal{C}^i$ *iff* $\forall\sigma^i\in\mathcal{C}^i$ $\mathfrak{D}((D_{\sigma^{-i},\sigma^i}\mid A_J)\parallel D^J)\leq\varepsilon$.

Let x^j be the projection from A onto A_j, $x=(x^j)_{j\in N}$, and for a subset $J\subset N$, $x^J=(x^j)_{j\in J}$. Let H_σ denote the entropy operator according to the distribution D_σ. Thus, for example, $H_\sigma(x^J)$ is the entropy of the A_J-valued random variable x^J that is distributed according to $D_\sigma|A_J$, and $H_\sigma(x^J\mid x^i)$ is the conditional entropy of x^J given the A_i-valued random variable x^i where x is distributed according to D_σ.

Definition 8.2. σ^{-i} ε-conceals D^J *against* $\mathcal{C}^i$ *iff* σ^{-i} ε*-implements* D^J *against* $\mathcal{C}^i$ *and* $\forall\sigma^i\in\mathcal{C}^i$ $H_{\sigma^{-i},\sigma^i}(x^J) - H_{\sigma^{-i},\sigma^i}(x^J\mid x^i)\leq\varepsilon$.

The conclusion of Theorem 2.2 can be stated using this (payoff independent) concept of concealment. Explicitly, under the assumptions of Theorem 2.2, there is a profile $\sigma^{-k}=(\sigma^j)_{j\neq k}$ of $m_j(n)$-recall strategies such that for every $\varepsilon>0$, for n sufficiently large the strategy profile σ^{-k} ε-conceals D^{-k} against $BR^k(m_k)$.

8.2. Alternative bounded recall models

The general class $BR^j(m)$ of bounded m-recall strategies does not impose any constraints on its action choices in the first m stages. A classical subclass of $BR^j(m)$ is that of the initialized bounded m-recall strategies $BR^j_*(m)$, which does impose restrictions on the action choices in stages $t=1,\dots,m$. A strategy $\sigma^j\in BR^j_*(m)$ is defined by a pair (e^j,f^j), where $f^j: A^m\to A_j$ and $e^j\in A^m$. The first item, f^j, determines j's action at any stage $t>m$, as a function of $(a_{t-m},\dots,a_{t-1})\in A^m$. (The string $(a_{t-m},\dots,a_{t-1})$ can be viewed as player j's memory before stage t.) The role of e^j is simply to pad j's memory up to length m at the early stages of the game, before there is an actual history of length m.

Therefore, the action prescribed by the strategy $\sigma^j\in BR^j_*(m)$, after any history of length $t-1$, is

$$\sigma^j(a_1,\dots,a_{t-1}) = \begin{cases} f^j(a_{t-m},\dots,a_{t-1}) & \text{if } t>m \\ f^j(e^j_t,\dots,e^j_m,a_1,\dots,a_{t-1}) & \text{if } t\leq m. \end{cases}$$

The main results remain intact if we restrict the minimization to initialized bounded-recall strategies.

Note that a strategy in $BR^j(m)$ (or $BR^j_*(m)$) may allow a player to rely on the past actions of all players, including himself. A bounded recall strategy whose action choice is independent of its own past actions is called an *exact bounded recall* strategy, and the corresponding classes are denoted $EBR^j(m)$ and $EBR^j_*(m)$.

	0	1
0	$-1+\varepsilon$	0
1	0	0

Matrix 0

	0	1
0	0	0
1	0	$-1-\varepsilon$

Matrix 1

Fig. 4. The game $G\varepsilon$.

For a finite set $N = \{1, \dots, k, k+1, \dots\}$ of players with $J_+ \neq \emptyset$, the main results remain intact if in the minimization at least one of the players $j > k$ is allowed to use a strategy $\sigma^j \in \Delta(BR^j_*(m_j))$ and all the other minimizers are restricted to strategies $\sigma^j \in \Delta(EBR^j_*(m_j))$.

A pure time-dependent m-recall strategy of player k is a strategy σ^k of player k such that $\sigma^k(a_1, \dots, a_{t-1})$ is a function of the stage t and the last m action profiles $(a_{t-m}, \dots, a_{t-1})$. The main result remains intact if we allow player k to maximize over all his time-dependent m-recall strategies.

8.3. Concealing correlation without a strong player

Here we give, in the exact bounded recall model, a simple example of a successful online concealed correlation by two players whose opponent is as strong as they are, with no other players in the game.

The stage game $G\varepsilon$, depicted in Fig. 4, is a slight modification of the three-player matching pennies game. For sufficiently small $\varepsilon > 0$, Matt's i.r.p. in the stage game is $> -\frac{1}{4}$. We will see that Matt's i.r.p. in the repeated game where all three players use 1-EBR strategies is smaller than his i.r.p. in the stage game.

The mixed strategies of Rowena and Colin, σ^1 and σ^2, are both the same: a mixture of two pure strategies, s_I and s_O, each with probability $\frac{1}{2}$. The pure strategy s_I always imitates the last action of the other, and at the first stage it plays 0. The pure strategy s_O always plays the action opposite to what the other has played, and at the first stage it plays 1. (σ^1, σ^2) induce four possible plays of Rowena and Colin, each with probability $= \frac{1}{4}$ (periodic plays are indicated in bold):

Play	Pure profile	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	......
A	(s_I, s_I)	$(\mathbf{0}, \mathbf{0})$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	......
B	(s_I, s_O)	$(\mathbf{0}, \mathbf{1})$	$(\mathbf{1}, \mathbf{1})$	$(\mathbf{1}, \mathbf{0})$	$(\mathbf{0}, \mathbf{0})$	$(0,1)$	......
C	(s_O, s_I)	$(\mathbf{1}, \mathbf{0})$	$(\mathbf{1}, \mathbf{1})$	$(\mathbf{0}, \mathbf{1})$	$(\mathbf{0}, \mathbf{0})$	$(1,0)$	......
D	(s_O, s_O)	$(\mathbf{1}, \mathbf{1})$	$(\mathbf{0}, \mathbf{0})$	$(1,1)$	$(0,0)$	$(1,1)$	......

We compute Matt's best 1-EBR strategy response against $\sigma = (\sigma^1, \sigma^2)$, by computing his best action for every contingency (= a pair of actions of Rowena and Colin played at some (unknown) point in time).

Suppose Matt saw $(0,0)$. Let x_A denote the pure strategy profile corresponding to the play A. The frequency of $\alpha = (0,0)$ in this play, $D_{x_A}(\alpha)$, is 1. Similarly, $D_{x_B}(\alpha) = \frac{1}{4}$, $D_{x_C}(\alpha) = \frac{1}{4}$, and $D_{x_D}(\alpha) = \frac{1}{2}$. Hence the average frequency is

$$D_\sigma(\alpha) = \sum_{y=x_A}^{x_D} \sigma[y] \cdot D_y(\alpha) = \frac{1}{4} \sum_{y=x_A}^{x_D} D_y(\alpha) = \frac{1}{4}\left(1 + \frac{1}{4} + \frac{1}{4} + \frac{1}{2}\right) = \frac{1}{2}.$$

Therefore, after seeing α, the probability of the play being A is

$$(\sigma \mid \alpha)[x_A] = \frac{\sigma[x_A] \cdot D_{x_A}(\alpha)}{D_\sigma(\alpha)} = \frac{\frac{1}{4} \cdot 1}{\frac{1}{2}} = \frac{1}{2}.$$

Similarly,

$$(\sigma \mid \alpha)[x_B] = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{8}; \qquad (\sigma \mid \alpha)[x_C] = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{2}} = \frac{1}{8}; \qquad (\sigma \mid \alpha)[x_D] = \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{1}{4}.$$

This induces the following probability over the forthcoming pair of actions (because if the play is A then after $(0,0)$ it will follow $(0,0)$, etc.).

	0	1
0	1/2	1/8
1	1/8	1/4

Therefore,[13] for ε small enough, Matt's best action in this case is 1.

[13] σ^1 and σ^2 are not contingent on Matt's play; and since Matt's EBR strategy cannot rely on his own actions either, the action he will take at this given stage has no effect on future play.

For the other possible pairs of actions, the same computation gives the distributions after observing $(0,1)$, $(1,0)$, and $(1,1)$:

1/2	0
0	1/2

After $(0,1)$

1/2	0
0	1/2

After $(1,0)$

1/2	1/4
1/4	0

After $(1,1)$

and Matt's best actions are 0, 0, and 1, respectively.

Thus we have defined σ^3, Matt's best 1-EBR response against (σ^1,σ^2). Now, the play A will yield an average payoff of 0 for Matt, since he will always play 1 (at least from the second stage on). In the play B, Matt gets $(-1+\varepsilon)$ at stages where Rowena and Colin play $(0,0)$, since he also plays 0 there (since Rowena and Colin played $(1,0)$ at the previous stage), and he gets 0 at the other stages; hence, his average payoff is $\frac{-1+\varepsilon}{4}$. Similarly, the play C also yields an average payoff of $\frac{-1+\varepsilon}{4}$, and D yields $\frac{-1-\varepsilon}{2}$. Therefore, Matt's expected payoff is $\bar{r}^3(\sigma^1,\sigma^2,\sigma^3)=(\frac{-1+\varepsilon}{4}+\frac{-1+\varepsilon}{4}+\frac{-1-\varepsilon}{2})/4=-\frac{1}{4}$; thus (σ^1,σ^2) indeed guaranteed that Matt's expected payoff in the 1-EBR infinitely repeated game will be no more than $-\frac{1}{4}$.

8.4. Online concealed correlation by finite automata

Here we demonstrate that the same type of online concealed correlation achieved in our results, where the players were restricted to SBR strategies, is also achievable in another model in which the players are restricted to strategies induced by finite automata, when analogous relations hold between the strength levels of the players. The adaptation to finite automata is done in a straightforward manner, using the same setup of strategies that was used for the SBR model. It seems conceivable that, for the finite automata model, the same result could perhaps be achieved under weaker assumptions, by some modification of these strategies.

A *finite automaton* for player i (in a repetition of a game $G=(N,A,r)$) is a tuple $\mathcal{A}=\langle S,q_1,f,g\rangle$, where S is a finite *state space*, $q_1\in S$ is the *initial state*, $f:S\to A_i$ is the *action function* prescribing an action to play at any given state, and $g:S\times A_{-i}\to S$ is the *transition function*.

Such an automaton $\mathcal{A}$ defines a strategy in the repeated game as follows. At any stage, the action taken by $\mathcal{A}$ is determined by the *current state* of $\mathcal{A}$, according to the action function f, while the current state along the stages is determined by the transition function g. Thus, let z_t denote the state of $\mathcal{A}$ at stage t. At stage 1, $z_1=q_1$. At stage $t+1$, $z_{t+1}=g(z_t,a_t^{-i})$, where a_t^{-i} denotes the actions taken by the other players at stage t. The action a_t^i taken by the automaton at stage t is $f(z_t)$.

The number of states $|S|$ is called the *size* of the automaton. We denote by $\Sigma^i(s)$ all strategies defined by some automaton of size s. Let $\sigma^i\in BR_*^i(m)$ be an m-recall strategy. Such a strategy is equivalent to a strategy induced by an automaton of size $|A|^m$ (see, e.g., Neyman, 1997, p. 247). Or, to put it in symbols (and identifying a strategy with its equivalence class), $BR_*^i(m)\subset\Sigma^i(|A|^m)$. Similarly, $BR^i(m)\subset\Sigma^i(\sum_{t=0}^m|A|^t)$.

The proof of Theorem 2.2 implies the following result in repeated games with finite automata. Let $G=(N,A,r)$ be a finite stage game with $N=\{1,\dots,k,\dots,|N|\}$. Let D^{-k} be a distribution on A_{-k} with $H(D^{-k})\geq\sum_{1\leq j<k}H(D^j)$, and let the automata sizes $s_j(n)$, $j\in N$ and $n\in\mathbb{N}$, satisfy

(A4) $s_{j+1}(n)\geq s_j(n)\to_{n\to\infty}\infty$,
(A5) $\log s_k\ll\log s_{k+1}$, and
(A6) $\log\log s_k\ll\log s_1$.

Then

$$\limsup_{n\to\infty}\min_{\sigma^{-k}\in\times_{j\neq k}\Delta(\Sigma^j(s_j))}\max_{\tau\in\Sigma^k(s_k)}\bar{r}\big(\sigma^{-k},\tau\big)\leq\min_{x\in A_k}r\big(D^{-k},x\big).$$

By setting $m_j=\frac{\log s_j}{\log|A|}$, $BR_*^i(m_j)\subset\Sigma^j(s_j)$; hence, it suffices to prove that

$$\limsup_{n\to\infty}\min_{\sigma^{-k}\in\times_{j\neq k}\Delta(BR_*^j(m_j))}\max_{\tau\in\Sigma^k(s_k)}\bar{r}\big(\sigma^{-k},\tau\big)\leq\min_{x\in A_k}r\big(D^{-k},x\big).$$

In deriving this inequality for the maximization over $\tau\in BR^k(m_k)$, the only property of such a strategy τ that was used is that the number of distinct strategies it defines on subgames is bounded by $|A|^{m_k}$, a property that holds also for any $\tau\in\Sigma^k(s_k)$.

A time-dependent automaton is a machine whose action choices may depend both on t and on the current state. The main results remain intact also when allowing player k to maximize over all his time-dependent automata with s_k states. Moreover, we can allow the maximizer to maximize over all time-dependent automata with mixed actions and mixed transitions.

8.5. Other comments

1. In Lehrer (1994) it is stated implicitly (see Lehrer, 1994, Theorem 1) that in the model of repeated games with bounded recall, if the recall of player k is larger than that of each player j in a subset J of players, then the marginal on A_J of a distribution D^{-k} on A_{-k} that can be concealed from player k is a product distribution. However, our example in Section 4 (or the general result) disproves Theorem 1 in Lehrer (1994), due to the possibility of concealed correlation.

2. We have discussed only infinitely repeated games where the payoff is the limiting average payoff. Note, however, that in the main result the play of the concealing group enters a cycle at some point, and the time it takes to enter that cycle does not depend upon the payoff function. The limiting average payoff equals the average along that cycle, and the discounted payoff converges to this average as the discount factor approaches 1. Therefore, our results apply also to games with discounted payoffs, provided the discount factor is close enough to 1; and likewise to finitely repeated games, provided the number of repetitions is sufficiently long.

3. We may wonder whether in a classical repeated game, without any rationality bounds, a non-product distribution can be ε-concealed (see Definition 8.2). Of course, the question makes sense only when the opponent has more than one available action. As it turns out, when the opponent has at least two actions, the distributions that can be ε-concealed for any $\varepsilon > 0$ are only the product distributions.

4. Apart from Section 8.3, we have not discussed the quantification and the feasibility of concealed correlation against an opponent who is the strongest in the field. Recently Peretz (2013) proved interesting results for this case. For a treatment of this problem in a different model of bounded rationality, where strategies are implemented by polynomial-time Turing machines, see Gossner (1999, 2000). For concealed correlation in the model of repeated games with signals, see Gossner and Tomala (2007).

5. In economic situations where “the rules of the game” forbid direct coordination among players (e.g., anti-cartel regulations, or various types of multistage auctions), the technique exhibited here may be used by the participants as a loophole, enabling them to circumvent these rules. Hence, the supervisor of the game may have to be able to identify and forbid this type of concealed correlation.

6. Additional related results and comments are in Bavly and Neyman (2003).

References

Alon, N., Spencer, J., 1992. The Probabilistic Method. Wiley-Interscience, New York.
Aumann, R.J., 1981. Survey of repeated games. In: Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Wissenschaftsverlag, Bibliographisches Institut, Mannheim, Wien, Zurich, pp. 11–42.
Bavly, G., Neyman, A., 2003. Online concealed correlation by boundedly rational players. DP 336. Center for the Study of Rationality, Hebrew University, Jerusalem.
Ben-Porath, E., 1986. Repeated games with bounded complexity (in Hebrew). M.Sc. Thesis. Hebrew University, and Stanford University (mimeo).
Ben-Porath, E., 1993. Repeated games with finite automata. J. Econ. Theory 59, 17–32.
Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. Wiley-Interscience, New York.
Gilboa, I., Schmeidler, D., 1994. Infinite histories and steady orbits in repeated games. Games Econ. Behav. 6, 370–399.
Gossner, O., 1999. Repeated games played by cryptographically sophisticated players. DP 99-07. THEMA.
Gossner, O., 2000. Sharing a long secret in a few public words. DP 2000-15. THEMA.
Gossner, O., Hernandez, P., Neyman, A., 2006. Optimal use of communication resources. Econometrica 74, 1603–1636.
Gossner, O., Tomala, T., 2007. Secret correlation in repeated games with imperfect monitoring. Math. Operations Res. 32, 413–424.
Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. J. Amer. Statistical Assoc. 58, 13–30.
Kalai, E., 1990. Bounded rationality and strategic complexity in repeated games. In: Ichiishi, T., Neyman, A., Tauman, Y. (Eds.), Game Theory and Applications. Academic Press, San Diego, pp. 131–157.
Kuhn, H.W., 1953. Extensive games and the problem of information. In: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games I. Princeton University Press, Princeton, pp. 193–216.
Lehrer, E., 1988. Repeated games with stationary bounded recall strategies. J. Econ. Theory 46, 130–144.
Lehrer, E., 1994. Finitely many players with bounded recall in infinite repeated games. Games Econ. Behav. 7, 390–405.
Neyman, A., 1985. Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma. Econ. Letters 19, 227–229.
Neyman, A., 1997. Cooperation, repetition, and automata. In: Hart, S., Mas-Colell, A. (Eds.), Cooperation: Game-Theoretic Approaches. Springer, Berlin, pp. 233–255.
Neyman, A., 1998. Finitely repeated games with finite automata. Mathematics Operations Res. 23, 513–552.
Neyman, A., 2008. Learning effectiveness and memory size. DP 476. Center for the Study of Rationality, Hebrew University.
Papadimitriou, C.H., Yannakakis, M., 1994. On complexity as bounded rationality (extended abstract). In: STOC-94, pp. 726–733.
Peretz, R., 2013. Correlation through bounded recall strategies. Int. J. Game Theory 42, 867–890.
Rubinstein, A., 1986. Finite automata play the repeated prisoner's dilemma. J. Econ. Theory 39, 83–96.