Online Concealed Correlation and Bounded Rationality*

Gilad Bavly and Abraham Neyman

The Hebrew University, Jerusalem†

October 15, 2005

Abstract

Correlation of players' actions may evolve in the common course of play of a repeated game with perfect monitoring ("online correlation"), and 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.

*This research was supported in part by Israeli Science Foundation grants 382/98 and 263/03, and by the Zvi Hermann Shapira Research Fund. Helpful discussions and comments by Elchanan Ben-Porath, Olivier Gossner, Penelope Hernandez, Rann Smorodinsky, and Eilon Solan are gratefully acknowledged.

†Institute of Mathematics and Center for the Study of Rationality and Interactive Decision Theory, Hebrew University, 91904 Jerusalem, Israel. gilbav@math.huji.ac.il (G. Bavly), aneyman@math.huji.ac.il, www.ratio.huji.ac.il/neyman (A. Neyman).

The result is illustrated in two models, each capturing a different aspect of bounded rationality. In the first, players use bounded recall strategies. In the second, players use strategies that are implementable by finite automata.

1 Introduction

1.1 The question of correlation

Let $A_1, \ldots, A_n$ be a tuple of sets, and $A = \times_{i=1}^n A_i$. Consider a sequence $a_1, a_2, \ldots$, where $a_t \in A$ for $t = 1, 2, \ldots$, and let $x_t^i : A^{t-1} \to A^i$ be some generating function that determines the i-th coordinate at time t, a_t^i, as a function of the whole history $a_1, \ldots, a_{t-1}$. A sequence of such functions, $x^i = x_1^i, x_2^i \ldots$, thus determines the i-th coordinate at every time t. We refer to such a sequence as i's generating rule. Let $A^* = \{\emptyset\} \cup A \cup A^2 \cup A^3 \ldots$ denote all finite strings of elements in A. Then x^i corresponds to a function $x^i : A^* \to A_i$. A tuple $x = (x^1, \ldots, x^n)$ of generating rules determines the whole sequence $a_1, a_2, \ldots$.

A game-theoretic interpretation is as follows. $N = \{1, \ldots, n\}$ is the set of players, and A_i is player i's set of actions in a normal-form game G. A sequence $a_1, a_2, \ldots$ ($a_t \in A$) is a *play* of the infinitely repeated game, where the game G (the *stage game*) is played at every stage t. Let $A^\infty = \times_{t=1}^\infty A$ denote the set of all possible plays. A generating rule x^i is interpreted as player i's strategy in the repeated game. Thus, a tuple $x = (x^1, \ldots, x^n)$ of strategies induces a play.

This paper studies uncertainties about the realized sequence, which result from uncertainties about the generating rules. The main issue is the relation between the uncertainties about each of the coordinates.

We start with a few definitions. A rule x^i has *bounded recall* if there exists an integer m, s.t. x^i depends only on the last m elements in the sequence, i.e., for every t and every sequence $a_1, \ldots, a_{t-1}$, $x^i(a_1, \ldots, a_{t-1}) = x^i(a_{t-m}, \ldots, a_{t-1})$. We then say that x^i is an *m-recall* rule.

A *behavioral* rule μ^i is a function $\mu^i : A^* \to \Delta(A_i)$, where for a set X,

$\Delta(X)$ stands for the probability distributions over X. Namely, μ^i specifies, for every possible history, a probability distribution over the forthcoming i-th coordinate. A tuple $\mu = (\mu^1, \ldots, \mu^n)$ of behavioral rules defines the conditional probability over the next tuple $a_t = (a_t^1, \ldots, a_t^n)$, given that the history was $h = (a_1, \ldots, a_{t-1})$, as $\mu(h)[a_t] = \times_{i=1}^n \mu^i(h)[a_t^i]$. Thus, μ induces a probability distribution over A^∞. In game-theoretic terms, behavioral rules correspond to behavioral strategies in a repeated game.

Another probabilistic extension is mixed rules: A *mixed rule* σ^i of player i is a probabilistic mixture of a set of deterministic (*pure*) rules of player i. We may also consider mixtures of behavioral rules. In any case, a tuple $\sigma = (\sigma^1, \ldots, \sigma^n)$ of mixed rules defines the product probability over the tuples of pure rules, and thus induces (via this product probability distribution) a probability distribution over A^∞. In game-theoretic terms, mixed rules correspond to mixed strategies.

This paper studies distributions on A^∞ that arise from tuples of independent rules $(\sigma^1, \ldots, \sigma^n)$; i.e., $\sigma^1, \ldots, \sigma^n$ are independent as random variables ranging over their respective sets of strategies. We shall use the term *profile* for such a tuple of independent rules. Indeed, the classical and common analysis of multistage interactions applying Nash equilibria assumes indirectly that the "dynamics" is induced by independent strategies. Moreover, in population equilibrium analysis it is common to interpret the mixture of strategies not as randomization but as the proportions of different types. This interpretation naturally leads to the assumption of independent behavior (compare [21, Section 3.2.2]).

A profile of mixed rules can be an expression of the uncertainty of an observer, at stage 0, about the rules that are actually chosen (and since the rules are independent, this uncertainty can be broken down into a list of independent uncertainties). At any stage, 0 or later, the uncertainty about the rules induces an uncertainty about the forthcoming tuple of actions as well (we borrow the term *action* from the game-theoretic interpretation, and use it here in the general context).

Now fix a profile of mixed rules $\sigma = (\sigma^1, \ldots, \sigma^n)$. These rules are independent; therefore the random variables $a_1^1, \ldots, a_1^n$ are independent. But for $t > 1$, $a_t^1, \ldots, a_t^n$ may certainly be correlated.

Example 1.1. Let $n \geq 2$ and $A_1 = A_2 = \{\alpha, \beta\}$. Let σ^1 choose at the first stage either α or β with equal probability. At any later stage, let it choose the action that it previously chose. Thus, σ^1 is actually a mixture of two pure rules: one always chooses α and the other always chooses β. Let σ^2 imitate the last choice of 1; i.e., at any stage $t \geq 2$ choose α if $a_{t-1}^1 = \alpha$, and choose β if $a_{t-1}^1 = \beta$. Then at any stage $t \geq 2$, the probability distribution over (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).

On the other hand, in this example the conditional probability distribution of (a_t^1, a_t^2), given the history $(a_1, \ldots, a_{t-1})$, is uncorrelated. Either the history is all α's (at least from stage 2 on), in which case we know for sure that the actual rule generating coordinate 1 (*1's rule*) is "always choose α," and then the forthcoming tuple is going to be (α, α) with probability 1; or the history is all β's, and then again we know the rule, and the forthcoming tuple will be (β, β) with probability 1. As a matter of fact, a covert implication of Kuhn's theorem about extensive-form games [14] is that this is always the case; i.e., the conditional probability over $a_t = (a_t^1, \ldots, a_t^n)$, given the history, is always a product distribution; i.e., it is the product of the marginal conditional probabilities over $a_t^1, \ldots, a_t^n$.

Example 1.1 exhibits the kind of issues at the heart of this paper. We start with a profile of mixed rules. Given some information about the realized sequence, we update our probability distribution over the rules. This, in turn, induces the probability distribution over the forthcoming tuple of actions in the sequence. We are especially interested in the question of correlation here: while we explicitly assumed that the prior probability over the rules is a product distribution, we ask whether the posterior probability over the rules, or over the forthcoming tuple, is still a product distribution.

This subject is reminiscent of important issues in statistics and probability concerning updated beliefs over the realized sequence. In this paper we focus our attention on the rules that generate the sequence (as in studies of

Hidden Markov Chains (see, e.g., [22]), Armed Bandits (see, e.g., [5]), and some topics in Time Series (see, e.g., [12])), rather than on the possible sequences themselves. The distinction between two levels, i.e., the sequence on the one hand, and the higher level of the rules that generate the sequence on the other, enables us to formulate questions such as in the previous paragraph.

As stated earlier, when you update according to the whole history $a_1, \ldots, a_{t-1}$, the posterior probability over $a_t = (a_t^1, \ldots, a_t^n)$ is a product distribution. Now consider a "forecaster" who has observed the play in the last k stages $a_{t-k}, \ldots, a_{t-1}$, and computes the posterior over a_t as the conditional probability given his observation. Assume that the rules of, say, players 1 and 2, are mixtures of m-recall rules. What kind of a probability distribution over $A_1 \times A_2$ can we get as a posterior? Clearly, for every product distribution $D = D^1 \times D^2$ there exist behavioral rules that will give D as the posterior: for each i, take the stationary (i.e., 0-recall) rule that chooses the forthcoming i-th coordinate according to the marginal distribution D^i, regardless of the history. But can non-product distributions be generated? Does it depend on the relation between k and m? on the presence of other players?

When a correlated (i.e., non-product) distribution is thus concealed from such a k-recall observer, we shall call it *k-recall concealed correlation.* More generally, we say that a correlation is *epistemically concealed* with respect to a piece of information h of an observer, if the conditional joint distribution of the forthcoming pair of actions, given the information h, is correlated.

The first result states a case where this is "almost impossible": for a subset $J \subset N$ of the coordinates, if we assume that for every $j \in J$ the generating rule σ^j is a mixture of m-recall behavioral rules, then there exists a finite number k, s.t. at almost every stage t, after observing the last k stages $a_{t-k}, \ldots, a_{t-1}$, the posterior distribution over the forthcoming tuple of the coordinates in J, $(a_t^j)_{j \in J}$, is "almost a product distribution" (a precise formulation of the result is given in Section 3). Moreover, this remains true when the posterior is derived without taking the "time" t into account (hence, we may view the posterior as a k-recall behavioral rule of the forecaster, who makes a prediction based solely on $a_{t-k}, \ldots, a_{t-1}$). The required k in this result is exponential[1] in m. This result may be viewed as an asymptotic version of (one aspect of) Kuhn's theorem, as the recall capacity m increases. Note that both allowing for subsets of N and allowing for behavioral rules only serves to strengthen the result.

Reviewing the above result, it may seem that k is unnecessarily large. It is perhaps intuitive to assume that $k = m$ is actually enough for the posterior to be precisely a product distribution. After all, in this case the forecaster knows the actions taken at the last m stages $a_{t-m}, \ldots, a_{t-1}$, which comprises all the information that the rules rely on. For each $j \in J$ this gives a posterior probability over j's rule and over the forthcoming action of j. One might expect these posteriors to combine independently across the members of J (remembering that the initial mixed rules are independent).

In fact, this is not so; in Section 4.3 we give an example with $n = 2$ where we use 1-recall mixed rules. Then, observing the last stage gives a posterior that is not a product distribution. Yet, one may ask about the asymptotic relation between m and k required for a product or an "almost product" distribution to be the k-recall prediction. Does it require an exponential k? A priori, it may seem that $k = m$ is enough.

1.2 Baffling a participant

Furthermore, suppose that the forecaster who makes the prediction $a_t^p \in \Delta(A)$ of the forthcoming tuple a_t (i.e., the posterior, based on his observation of the play in the last k stages) can also, along the stages, produce "notes" for himself, of the following form: at every stage t he chooses a note b_t out of a fixed finite set of notes B. Both the choice of a_t^p and of b_t may rely on the last k stages of the sequence $a_{t-k}, \ldots, a_{t-1}$ and on the last k notes $b_{t-k}, \ldots, b_{t-1}$ as well.

The main result, Theorem 7.1, states that when $J \subsetneqq N$ then even for

[1] I.e., there exists a number C s.t. $k \geq 2^{Cm}$

such a forecaster, and even for k larger than m, but subexponential[2] in m, there exists a setting for which the prediction is correlated. Moreover, this can be achieved by mixtures of deterministic rules. The proof makes use of a rule σ^i for $i \in N \setminus J$, which is still a bounded recall rule, but with recall larger than k (and it also requires that the set A_i be sufficiently large). The asymptotics when $J = N$ is still an open question (see also Section 1.4).

This theorem has a game-theoretic corollary as follows. Call one player "stronger" than another if his recall is larger. Then, if we think of our game-theoretic interpretation, and think of the forecaster as an additional player in the game, we get the following corollary: a strong player can orchestrate a correlation of weak players' actions, where the correlation is concealed from an opponent of intermediate strength.

The key point in the proof of the theorem is that, on the one hand, the strong player's sequence of actions will guide the weak players in correlating their actions. On the other hand, although the opponent can recall all the actions that these weak players recall (and more), these "signals" (i.e., the strong player's actions) will be unintelligible to him.

Whenever the strong player has sufficiently many actions, then any distribution over $\times_{j \in J} A^j$ can be concealed (i.e., be approximated as the prediction of the opponent). If he has, say, only two actions, then he cannot convey a lot of information through his actions, in which case there are some distributions that can be concealed, while as for the rest we cannot tell at present. Section 8.2 deals with the quantification of conditions that enable the concealment of a distribution.

1.3 Applications

In games, the concealment of correlation can be reflected by the payoffs and the individually rational payoffs of the players. For example, consider the normal-form game in Figure 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

[2] I.e., $\frac{\log k}{m} \to 0$

or West). The numbers in the matrix are the payoffs of Matt. The payoffs of Rowena and Colin are not specified.

	L	R
T	-1	0
B	0	0

E

	L	R
T	0	0
B	0	-1

W

Figure 1: Three-player matching pennies

For every pair of mixed actions, $(x, 1-x)$ for Rowena, and $(y, 1-y)$ for Colin, either $xy \leq \frac{1}{4}$ or $(1-x)(1-y) \leq \frac{1}{4}$. 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 $-\frac{1}{4}$. Therefore the individually rational payoff (henceforth i.r.p.) of Matt in this game is $-\frac{1}{4}$.

If a correlation device is available to Rowena and Colin, they can play with probability $\frac{1}{2}$ the action pair (T, L), and with probability $\frac{1}{2}$ the pair (B, R). If the outcome of the correlation device is concealed from Matt, his best response to this correlated play yields him an expected payoff of $-\frac{1}{2}$. Indeed, $-\frac{1}{2}$ is Matt's *minmax in correlated actions*; i.e., it is the lowest level of expected payoff for Matt that Rowena and Colin can guarantee, if their randomization is allowed to be correlated. That is, $-\frac{1}{2} = \min_{y \in \Delta(A_1 \times A_2)} \max_{x \in A_3} r^3(y, x)$. On the other hand, if the correlation signal is public, Matt's i.r.p. remains $-\frac{1}{4} = \min_{y \in \Delta(A_1) \times \Delta(A_2)} \max_{x \in A_3} r^3(y, x)$.

Let us add a fourth player, Forest, to this game. Forest has two actions, but his actions have no influence on the payoff. Therefore, Matt's i.r.p. remains $-\frac{1}{4}$. Consider the repetition of this game. As stated earlier, the epistemic concealment of correlation is impossible in the classical model of repeated games (with standard signalling), with or without public signals (compare, e.g., [24, p. 181]). As a corollary, a player's i.r.p. in the repeated game is equal to his i.r.p. in the stage game. The possibility of concealed correlation naturally emerges however in the variant of the repeated game, in which Matt does not observe the actions of the other players, and at least either Rowena or Colin observes the action of the other, or more generally, in a repeated game with signals.[3]

This paper applies to repeated games with perfect monitoring, without additional signals, when players use simple strategies. Players may tend to use relatively simple strategies when it is difficult (or costly) to implement complicated patterns of behavior, or to process large amounts of data (such as the full history at a given stage of a long repeated game). Here we assume that each player has an individual upper bound for the complexity of the strategies he may use.

In Section 5 we discuss the infinite repetition of the three-player matching pennies game, when the recall of the players is bounded (which is one model of bounded complexity). We show how Forest, Rowena and Colin can bring Matt's payoff down to $-\frac{1}{2}$. This is an example of what we call *online concealed correlation*, that is, when players achieve some correlation between their actions in the common course of play (i.e., by the *public* sequence of actions taken) and thus bring their opponent's i.r.p. down below his i.r.p. in the stage game.

This is achieved when Forest is relatively strong while Rowena and Colin are relatively weak, compared to Matt, and it is an instance of the main result already described. In Section 6 we discuss the same game, but with a different model of bounded complexity, that of finite automata. In Section 7 we give the main result.

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 online concealed correlation.

A "folk theorem" is a characterization of the set of equilibrium payoffs in the infinitely repeated game, by means of the data of the stage

[3]The quantification, as a function of the signalling structure, of concealed correlation in repeated games with signals (see [11]) is subtle, however.

game[4]. For the case of two players, there are folk theorems related to various models of bounded complexity in repeated games (e.g., Ben-Porath (1993), Lehrer(1988), and Neyman(1998)). For the case of more than two players, however, folk theorems are lacking. Our main result enables the derivation of Theorem 9.1, which is a folk theorem for a large class of n-player repeated games in which there is a mechanism designer who sends a public signal at each stage, and the players' complexity is bounded.

The theorem states that when every player's recall capacity is subexponential in the capacity of the others, then the set of equilibrium payoffs converges to the set of feasible payoffs that give each player at least his minmax in correlated actions (in the stage game). For example, consider the three-player matching pennies stage game, and assume that Rowena and Colin each get half of what Matt loses. Then the the limit set of equilibrium payoffs consists of all feasible payoffs that give Rowena and Colin at least 0, and give Matt at least $-\frac{1}{2}$, i.e., it consists of the interval between $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, -\frac{1}{2})$.

1.4 Further results and questions

It seems desirable to have a notion of concealment that is independent of payoff and also of epistemology, but rather depends merely upon the strategies at the players' disposal. Such a notion—of "concealed distribution"—is defined in Section 8.1. Section 8.2 contains substantial extensions of the main result.

As mentioned, it remains an open question whether online concealed correlation by weak players can be achieved without the help of stronger players, in the models discussed so far. In other models, where strategies are implemented by polynomial-time Turing machines, concealed correlation achieved by weak players without help is exhibited in the three-player model of repeated games with standard signalling in Gossner (2000), and in the n-player

[4] In part of the literature the term "folk theorem" is used somewhat differently, referring also to theorems that do not give an exact characterization.

model with *public signals* in Gossner (1999).[5]

2 The Model

From here on we shall stick to the language of games, and forsake the terminology of abstract sequences we started with, since the payoffs of games provide an additional perspective on the concealment of correlation. In doing so there is no loss of generality, and the translation of results back into the sequences terminology is straightforward.

An infinitely repeated game serves also as a convenient paradigm for a long finitely repeated game. 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 Section 10).

2.1 Infinitely repeated games

Let $G = (N, A, r)$ be an n-player game in strategic form, $A = (A_i)_{i \in N}, r = (r^i)_{i \in N}$, where $N = \{1, \ldots, n\}$ is the set of players, A_i is the set of actions of player i, and r^i is the payoff function of i, $r^i : A \to \mathbb{R}$.

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

The game G_∞ denotes the infinite repetition of G. 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 $a^i_t \in A_i$ (superscripts denote players; subscripts, the stage).

A *play* of G_∞ is an infinite sequence $a_1, a_2, \ldots (a_t \in A)$, where $a_t = (a^1_t, \ldots, a^n_t)$. In our context, the payoff for i in G_∞ is the "limit" of his average payoff along the play,[6] namely "$\lim$"$_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r^i(a_t)$. However,

[5] Assuming the widely accepted cryptology assumption of the existence of "trapdoor" functions.

[6] 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

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, in the context of bounded recall strategies, or finite automata, such a limit does exist.

Unless we explicitly impose restrictions, we assume that the information available to player i at stage t consists of all past actions $a_1, \ldots, a_{t-1}$ of all the players (the "history"). Now, A^{t-1} is the set of all histories that may be played at the first $t-1$ stages (where A^0 stands for $\{\emptyset\}$). $A^* = \cup_{t=1}^{\infty} A^{t-1}$ is the set of all possible histories, of any length. Hence, a pure strategy of i in G_∞ is $\sigma^i : A^* \to A_i$. That is, 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. In the context of repeated games, the term "action" is typically used for a pure strategy in the stage game, while "strategy" refers to a strategy in the repeated game.

2.2 Bounded recall strategies

A *bounded recall* (BR) strategy for player i in G_∞ assumes that i's play at any given stage relies only on the last m_i actions played. Formally, a (pure) m-recall strategy for i is $\sigma^i = (e^i, f^i)$, where $f^i : A^m \to A_i$ and $e^i \in A^m$. f^i determines i's next action for any sequence $h \in A^m$, where $h = (a_{t-m}, \ldots, a_{t-1})$ are the last m actions played (h is player i's "memory"). The role of e^i is simply to pad i'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 σ^i, after any history of length $t-1$, is:

$$\sigma^i(a_1, \ldots, a_{t-1}) = \begin{cases} f^i(a_{t-m}, \ldots, a_{t-1}) & \text{if } t > m \\ f^i(e_t, \ldots, e_m, a_1, \ldots, a_{t-1}) & \text{if } t \leq m\,. \end{cases}$$

(Note that a player does not know the time t. Also note that we allow him to rely on the past actions of all players, including himself.)

called *discounted*.

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 $\overrightarrow{m} = (m_1, \ldots, m_n)$, the game $G(\overrightarrow{m})$ is defined as the infinite repetition of G, but where the strategies of player i are his m_i-recall strategies. I.e.,

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

where $\bar{r}$ is defined for a tuple $\phi = (\phi_1, \ldots, \phi_n) \in \times_{j=1}^n BR^j(m_j)$ 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 resulting from the strategy profile ϕ.

Here we can write "lim" without reservations, since any play resulting from BR 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, one of them is bound eventually to appear twice during the play, say at stages t_1, t_2. Then, every player takes at t_2 the same action he took at t_1, and in this way 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 $G(\overrightarrow{m})$, i.e.,

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

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

Throughout, we assume that the players are ordered according to their recall capacities, i.e., $m_1 \leq m_2 \leq \ldots \leq m_n$.

3 A Preliminary Result

Here we state the first result, given without proof. It says that correlation cannot, in a sense, be concealed from a really strong opponent. Thus, it serves as a contrast to the main result, which gives settings in which concealed correlation is feasible.

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. Let $D(x \mid y)$ denote the conditional probability distribution over the random variable x given y.

Theorem 3.1. *Let $J \subset N$. There exists C, s.t. for $k \geq \exp Cm$ there is a map $\mu : A^k \to \times_{j \in J} \Delta(A_j)$, s.t. if for every $j \in J$, σ^j is a mixture of m-recall behavioral strategies, then for sufficiently large T we have*

$$\frac{1}{T} \sum_{t=k+1}^{T} E \left\| D(a_t^J \mid a_1, \ldots, a_{t-1}) - \mu(a_{t-k}, \ldots, a_{t-1}) \right\| \leq \varepsilon(m)$$

where $\varepsilon(m) \to_{m \to \infty} 0$.

4 Examples

In this section we discuss the feasibility of concealed correlation, both from an epistemic and a strategic point of view, and illustrate with a few simple examples. This section is independent of Sections 5–10.

4.1 Preliminaries

Consider the infinite repetition of a stage game $G = (N, A, r)$. For any element $(a_1, \ldots, a_t)$ in A^* ($A^* = A^0 \cup A \cup A^2 \cup A^3 \ldots$), a pure strategy x^i of player i determines his next action. Thus, a profile $(x^1, \ldots, x^n)$ of pure strategies induces a chain on A^* that consists of the histories at every stage.

If $(x^1, \ldots, x^n)$ is a profile of m-recall pure strategies, then, at every stage, the last m actions completely determine each player's next action. Hence, we get a deterministic Markov chain, with state space A^m; a suffix of history, consisting of the last m actions, $(a_{t-m}, \ldots, a_{t-1})$, determines the next suffix, $(a_{t-m+1}, \ldots, a_t)$.

Thus, $(x^1, \ldots, x^n)$ induces a cycle $(\alpha_1, \ldots, \alpha_L)$ on the state space A^m. If we define the payoff of a state $\alpha_i = (a_1, \ldots, a_m)$ by $R(\alpha_i) = r(a_m)$, then the limiting average payoff is just the average of $R(\alpha_i)$ over this cycle.

More generally, let k be any positive integer, and for any $\alpha=(\xi_1,\ldots,\xi_k)\in A^k$, define $R(\alpha)=r(\xi_k)$. Let $\mathfrak{h}=a_1,a_2,\ldots$ be a periodic play. The limiting frequency of the string α appearing along $\mathfrak{h}$ is

$$D(\mathfrak{h},\alpha)=\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{I}((a_{t+1},\ldots,a_{t+k})=\alpha)$$

where $\mathbb{I}$ denotes the indicator function. Then, the limiting average payoff along $\mathfrak{h}$ is $\bar{r}(\mathfrak{h}) = \sum_{\alpha\in A^k} D(\mathfrak{h},\alpha)R(\alpha)$. Again, this amounts to averaging over the irreducible period of the play,[7] and for every α not appearing in this period, $D(\mathfrak{h},\alpha)=0$.

For a profile $\sigma = (\sigma^1,\ldots,\sigma^n)$ of mixed strategies, which are mixtures of pure m-recall strategies, the resulting dynamics over A^m is a mixture of deterministic Markov chains. Thus, σ induces a probability distribution over cycles on the state space A^m.

The expected limiting average payoff $\bar{r}(\sigma)$ may be computed in two different ways. One way is to compute the limiting average payoff over each deterministic periodic play $\mathfrak{h}$ induced by a profile of m-recall pure strategies $(x^1,\ldots,x^n)$, and then average over these plays according to the probability $\Pr_\sigma(\mathfrak{h})$ induced by the product probability over the pure strategy profiles, which is induced by the mixed strategy profile σ; namely $\bar{r}(\sigma) = \sum_{\mathfrak{h}}\Pr_\sigma(\mathfrak{h})\cdot\bar{r}(\mathfrak{h}) = \sum_{\mathfrak{h}}\Pr_\sigma(\mathfrak{h})\sum_{\alpha\in A^k} D(\mathfrak{h},\alpha)R(\alpha)$. Note that the length of the irreducible period is not a constant over these plays; different profiles of pure m-recall strategies induce different cycle lengths.

An alternative computation (which is more illuminating for the construction in the sequel) is accomplished by reversing the order of summation: for each element α of A^k, compute its average limiting frequency according to the probability over plays (averaging "over time and space")

$$D_\sigma(\alpha)=\sum_{\mathfrak{h}}\Pr_\sigma(\mathfrak{h})D(\mathfrak{h},\alpha) \tag{4.1}$$

and then $\bar{r}(\sigma) = \sum_{\alpha\in A^k} R(\alpha)D_\sigma(\alpha)$.

[7] If we define $R(\alpha)$ as $r(\xi_1)$ or $r(\xi_2)$ or $r(\xi_3)\ldots$ (or any convex combination thereof), still $\bar{r}(\mathfrak{h}) = \sum_{\alpha\in A^k} D(\mathfrak{h},\alpha)R(\alpha)$.

4.2 What a player knows

4.2.1 Example

Consider again the infinite repetition of the three-player matching pennies game (Figure 1). For notational simplicity, let the sets of pure actions of Rowena, Colin, and Matt (A_1, A_2, and A_3, respectively) be $\{0, 1\}$, instead of {T,B}, {L,R}, and {E,W}. Therefore, the payoff of Matt at the stage game is

$$r^3(a^1, a^2, a^3) = \begin{cases} -1 & \text{if } \; a^1 = a^2 = a^3 \\ 0 & \text{otherwise.} \end{cases}$$

Consider the following simple attempt at correlation by Rowena and Colin (compare the even simpler Example 1.1). Rowena plays a mixture of four pure strategies $\phi_1, \ldots, \phi_4$, where ϕ_1 plays 0 at all stages, ϕ_2 plays 1 at all stages, ϕ_3 starts by playing 0 and continues by alternating between 1 and 0, and ϕ_4 starts by playing 1 and continues by alternating. Colin, at any stage but the first two, looks two stages back at what Rowena played then, and plays the same. Therefore, whether Rowena alternates or not, Colin's actions will match hers (Colin plays arbitrarily at the first two stages).

Let $(a_t)_{t\geq 1}$ be the actual play of this game. Consider the play of Rowena and Colin; with probability $\frac{1}{2}$ Rowena alternates, and then the cycle will consist of an alternating play of 0 and 1. With probability $\frac{1}{4}$ they always play 0, and with probability $\frac{1}{4}$ they always play 1.

For $t > 2$, Rowena's and Colin's actions, a^1_t and a^2_t, are certainly correlated —indeed, they always take the same value. In the extreme case, where Matt's recall capacity, m_3, is 0, the above strategies provide another simple example of correlation successfully concealed when the opposition is weak: since Matt remembers nothing, at any stage $t > 2$, the probability he ascribes to (a^1_t, a^2_t) is the prior probability we ascribed to it before the game started (see (4.1)). That is, the action pair $(1, 1)$ is played with probability $\frac{1}{2}$, and the action pair $(0, 0)$ is played with probability $\frac{1}{2}$. Note that this would still be the distribution ascribed by Matt, even if we allowed his play to be time-dependent, i.e., by letting him use a (0-memory) Markovian strategy, which means playing a deterministic (nonreactive) sequence.

Indeed, the successful concealment of correlation by Rowena and Colin is rewarded here: Matt's expected payoff, whatever 0-recall, either time-dependent or time-independent, strategy he chooses will be $-\frac{1}{2}$. Note that σ^2, Colin's strategy, can be implemented as a pure 2-recall strategy, and σ^1, Rowena's strategy, is a mixture of four pure 1-recall strategies

However, if Matt's recall is at least as large as Rowena's and Colin's (say $m_3 = 2$), and assuming that Matt knows the strategies σ^1 and σ^2 (as we would assume in studying equilibrium, or when computing a best response or the i.r.p.), he can tell, at every stage, what the forthcoming actions of Rowena and Colin are, and achieve a payoff of 0.

We remark that the memory content of a recall-bounded player indeed tells us less about his forthcoming action than the full history does. For example, suppose Rowena played $(0, 1, 0, 1, 0)$ at the first five stages. Knowing that, we know for sure that her chosen pure strategy is ϕ_3, and her forthcoming action is 1. In contrast, although her recall capacity is 1, if we, like her, know only the last stage (not even knowing the time t), then her strategy might be ϕ_1, ϕ_3, or ϕ_4, and we would ascribe a positive probability to her forthcoming action being 0.

4.2.2 Posterior probabilities

Generally speaking, and dropping, for the moment, the assumption of bounded recall, let the players' profile of mixed strategies be $\sigma = (\sigma^1, \ldots, \sigma^n)$. We want to compute the probability distribution[8] $\Pr_\sigma(a_t^j \mid a_1, \ldots, a_{t-1})$ over player j's next action $a_t^j \in A_j$, given a history $h = (a_1, \ldots, a_{t-1})$ of the play.

σ^j is a mixture of pure strategies, and Σ^j is the set of player j's pure strategies. Given a measurable subset Φ of Σ^j, $\sigma^j[\Phi]$ is the probability that the actually chosen pure strategy ϕ is in Φ. Given the history h, we get a revised posterior "mixed strategy" $(\sigma^j \mid h)$, which consists of discarding all the pure strategies ϕ, s.t. ϕ is incompatible with the play h; i.e., if we denote $C^j(h) = \{\phi \in \Sigma^j : \phi \text{ is compatible with } h\}$, then

$$(\sigma^j \,|\, h)[\Phi] = \sigma^j[\Phi \cap C^j(h)] \,/\, \sigma^j[C^j(h)].$$

[8]In the sequel we write simply "Pr" instead of "$\Pr_\sigma$," when there is no ambiguity.

$(\sigma^j|h)$ is a probability distribution over the pure strategies Σ^j, and thus h induces, through $(\sigma^j|h)$, a distribution over j's forthcoming action, denoted $(\sigma^j|h)(h)$. Actually, $\times_{h\in A^*}[(\sigma^j|h)(h)]$ is the behavioral representation of σ^j.

Getting back to bounded recall strategies, suppose we are outside observers who know neither the time t nor the full history $a_1,\ldots,a_{t-1}$, but only that the play of the last m stages is $\alpha = a_{t-m},\ldots,a_{t-1}$. Let $\phi\in\Sigma^j$ be a pure strategy of player j. Assuming we know the profile of mixed strategies $\sigma=(\sigma^1,\ldots,\sigma^n)$, the prior probability of ϕ is $\sigma^j[\phi]$. Suppose $D_\sigma(\alpha)>0$, where $D_\sigma(\alpha)$ is the average limiting frequency of α when the strategy profile is σ. Given that we observed this suffix of history α, the posterior probability we ascribe to ϕ, denoted $(\sigma^j\,|\,\alpha)[\phi]$, will be, as in Bayes' rule, $\sigma^j[\phi]\cdot D_{\sigma^{-j},\phi}(\alpha)/D_\sigma(\alpha)$, where $D_{\sigma^{-j},\phi}(\alpha)$ is the average limiting frequency when the strategy profile is (σ^{-j},ϕ). Now, by (4.1),

$$D_\sigma(\alpha)=\sum_{\mathfrak{h}}\left(\sum_{\psi\in\Sigma^j}\sigma^j[\psi]\cdot \Pr_{\sigma^{-j},\psi}(\mathfrak{h})\right)D(\mathfrak{h},\alpha)$$

$$=\sum_{\psi\in\Sigma^j}\sigma^j[\psi]\sum_{\mathfrak{h}}\Pr_{\sigma^{-j},\psi}(\mathfrak{h})D(\mathfrak{h},\alpha)=\sum_{\psi\in\Sigma^j}\sigma^j[\psi]\cdot D_{\sigma^{-j},\psi}(\alpha)\,.$$

Hence,

$$(\sigma^j\,|\,\alpha)[\phi]=\frac{\sigma^j[\phi]\cdot D_{\sigma^{-j},\phi}(\alpha)}{\sum_{\psi\in\Sigma^j}\sigma^j[\psi]\cdot D_{\sigma^{-j},\psi}(\alpha)}\quad. \tag{4.2}$$

Thus, the revision takes into account not just the compatibility or incompatibility of a pure strategy, but also the average limiting frequency it induces on α.

In the example in Section 4.2.1, if α contains two consecutive 0's played by Rowena, then ϕ, Rowena's chosen pure strategy, must be ϕ_1. If she played two consecutive 1's, then ϕ must be ϕ_2. If she played 0 and 1, no matter in what order, then ϕ is either ϕ_3 or ϕ_4, and by (4.2) we ascribe probability $\frac{1}{2}$ to each. But note that even in this case, we know for certain the forthcoming action of Rowena.

Similarly, let $x=(x^1,\ldots,x^n)$ be a profile of pure BR strategies. Denote $\sigma[x]=\times_{j=1}^n\sigma^j[x^j]$, and $\Sigma=\times_{j=1}^n\Sigma^j$. Given α, the probability of x is

$$(\sigma\,|\,\alpha)[x]=\frac{\sigma[x]\cdot D_x(\alpha)}{D_\sigma(\alpha)}=\frac{\sigma[x]\cdot D_x(\alpha)}{\sum_{y\in\Sigma}\sigma[y]\cdot D_y(\alpha)}\quad. \tag{4.3}$$

$(\sigma \,|\, \alpha)$ is not necessarily equal to $\times_{j=1}^{n}(\sigma^j \,|\, \alpha)$. In other words, the probabilities $(\sigma^1 \,|\, \alpha), \ldots, (\sigma^n \,|\, \alpha)$ may be correlated. This runs counter to conditioning on the full history h, where Kuhn's theorem would imply that the conditional probability $(\sigma \mid h)$ is the product of the marginal conditional probabilities $(\sigma^j \,|\, h)$.

Moreover, as we shall see in Section 4.3, this correlation holds when all players use mixtures of m-recall strategies (recall that $\alpha = a_{t-m}, \ldots, a_{t-1}$). In this case, α induces a probability distribution $(\sigma^j \,|\, \alpha)(\alpha)$ over j's forthcoming action, which we denote $\Pr(a_t^j \mid a_{t-m}, \ldots, a_{t-1})$. We shall furthermore see that these probabilities are correlated too. Thus, the correlation of actions is epistemically self-concealed: the forthcoming tuple of actions is correlated, even when we condition on all the information that the players themselves possess at this stage.

4.2.3 Best response

This brings us back to the discussion at the end of Section 1.1 regarding concealed correlation by weak players. Suppose our outside observer joins the game (but his own actions are still not observed by the other players) and wishes to compute his best response at some stage against the forthcoming tuple of actions of the other players, based on his observation of the last m stages $\alpha = (a_{t-m}, \ldots, a_{t-1})$. Assume that the best response is concerned only with players whose recall is not larger than m (i.e., the actions of stronger players do not affect his payoff). For simplicity, let us assume that all the other players' recall is not larger than m, and let us call our observer who has joined the game player $(n{+}1)$. As we said, the probability he should ascribe to the forthcoming tuple of actions is not given by the prior σ, but by the posterior $(\sigma \,|\, \alpha)$. As we asserted, this posterior probability, $\Pr(a_t^1, \ldots, a_t^n \,|\, \alpha)$, may not be a *product probability*; i.e., it may not be the product of the marginal probabilities $\Pr(a_t^j \,|\, \alpha)$; i.e., it may not belong to $\times_{j=1}^{n}\Delta(A_j)$. Therefore, the best response against this probability distribution is not guaranteed to yield player $(n+1)$ an expected payoff as large as his i.r.p. in the stage game, which is $\min_{y \in \times_{j=1}^{n}\Delta(A_j)} \max_{x \in A_{n+1}} r^{n+1}(y, x)$.

Computing player $(n+1)$'s best-response m-recall-strategy in a "normal" game, where his actions are observed by the other players, is obviously more complex. Then, the actions of players $1, \ldots, n$ depend on the play of all players, including player $(n+1)$. Therefore, the average limiting frequency of α, and consequently the computation of $(\sigma^j \mid \alpha)$, depend on the strategy profile of all the players $\sigma = (\sigma^1, \ldots, \sigma^n, \sigma^{n+1})$. In addition, the action taken by player $(n+1)$ at the forthcoming stage should be judged not only by the payoff at the forthcoming stage, but also by the action's effect on the future play of the other players. Moreover, even in the variation of the model in which the actions of player $(n+1)$ are not observable, the above procedure does not necessarily define an m-recall best-response for player $(n+1)$, the reason being that the action he takes now could be used by *himself* at later stages ("signal to himself"; see also Section 4.4).

4.3 Epistemically concealed correlation

Here we give our example that the probabilities over the pure strategy sets of the players, $(\sigma^1 \mid \alpha), \ldots, (\sigma^n \mid \alpha)$, may be correlated (where $\alpha = a_{t-m}, \ldots, a_{t-1}$). Furthermore, the probabilities $(\sigma^1 \mid \alpha)(\alpha), \ldots, (\sigma^n \mid \alpha)(\alpha)$ over the forthcoming actions of the players are correlated too. That is, we show that even when the recall capacity of all players is no more than m, still

$$\Pr(a_t^1, \ldots, a_t^n \mid a_{t-m}, \ldots, a_{t-1}) \neq \times_{j=1}^n \Pr(a_t^j \mid a_{t-m}, \ldots, a_{t-1})\,.$$

Consider a game of two players, $m_1 = m_2 = 1$ and $A_1 = A_2 = \{0, 1\}$. Each player's mixed strategy is: with probability $\frac{1}{2}$, always imitate the last action of the other player, and with probability $\frac{1}{2}$ always play the opposite action. The initial memory is the tuple $(0, 0)$.

This profile $\sigma = (\sigma^1, \sigma^2)$ of mixed strategies induces four possible plays, each play corresponding to one profile y of pure strategies whose probability is $\sigma[y] = \frac{1}{4}$ (periodic plays are indicated in bold):

(A) When both players imitate each other's actions:

$$\begin{array}{ccc} \mathbf{0} & 0 & \dots \\ \hline \mathbf{0} & 0 & \dots \end{array}$$

(B) When 1 imitates and 2 does the opposite:

$$\begin{array}{cccccc} \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & 0 & \dots \\ \hline \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & \dots \end{array}$$

(C) When 2 imitates and 1 does the opposite:

$$\begin{array}{cccccc} \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & 1 & \dots \\ \hline \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & 0 & \dots \end{array}$$

(D) When both do the opposite:

$$\begin{array}{cccc} \mathbf{1} & \mathbf{0} & 1 & \dots \\ \hline \mathbf{1} & \mathbf{0} & 1 & \dots \end{array}$$

Now let $m=1$, and suppose we observe, at some (unknown) point in time t, that the last actions taken by the players were, say, the pair $a_{t-1}=(0,0)$. We compute the probability distribution over the forthcoming pair of actions $a_t=(a_t^1,a_t^2)$, given the observation $\alpha=(0,0)$. This is achieved through the computation of the probability distribution over pure strategy profiles, given in (4.3).

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

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

and

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

Similarly,

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

The action pair a_t, following the action pair $a_{t-1} = \alpha = (0,0)$, is equal to $(0,0)$ iff the play is A; therefore $\Pr(a_t = (0,0) \mid \alpha) = (\sigma \mid \alpha)[x_A] = \frac{1}{2}$. Similarly, $\Pr(a_t = (0,1) \mid \alpha) = (\sigma \mid \alpha)[x_B] = \frac{1}{8}$; $\Pr(a_t = (1,0) \mid \alpha) = (\sigma \mid \alpha)[x_C] = \frac{1}{8}$; $\Pr(a_t = (1,1) \mid \alpha) = (\sigma \mid \alpha)[x_D] = \frac{1}{4}$.

Thus we have computed the probability distribution $\Pr(a_t \mid \alpha)$, and the result is not a product probability distribution.

4.4 Discussion

If each player's mixed strategy is known, then the uncertainty about the players' chosen pure strategies, before the game starts, can be broken down into a list of independent uncertainties. During the play, if we know the full history of past actions, the uncertainty about the chosen pure strategies, as well as the forthcoming actions of the players, can again be broken down into a list of independent uncertainties.

In Section 4.3 we saw that for a profile of m-recall strategies the uncertainties, conditional on the players' information, need not be independent. This still does not resolve the *strategic* question of online concealed correlation by weak players. Indeed, there are a few things we should note.

First, in Section 4.3 there is no player from whom correlation is supposed to be concealed. If we add a third player there, endowed with a recall capacity of, say, $m_3 = 1$, we may consider again his uncertainties during the play. This third player has more information at his disposal than just the other players' past actions, having also his own actions to rely on. Therefore, the epistemic analysis may have to be modified, since he can choose his strategy in such a way that his own past actions will, at least at some stages, convey additional information about the other players' chosen strategies (i.e., he can signal to himself).

Second, the notion of online concealed correlation is quantified by the i.r.p., i.e., in terms of expected payoff. Suppose we again add a third player, and now consider the three-player matching pennies game, where player 1 is Rowena, 2 is Colin, and 3 is Matt (and $A_1 = A_2 = A_3 = \{0,1\}$). What is the best expected payoff Matt can achieve against the strategies σ^1 and σ^2 of

Rowena and Colin described in Section 4.3? Is it lower than $-\frac{1}{4}$? Of course, in answering this, we have to remember the first point, namely, that Matt may signal to himself through his own actions.

But let us consider another commonly used model for bounded recall, which we shall call *exact bounded recall* (EBR), that allows a player to rely only on the past actions of other players. In this model, too, the strategies σ^1 and σ^2 considered above are not sufficient to bring Matt's expected payoff to below $-\frac{1}{4}$; for Matt can secure an expected payoff of $-\frac{1}{4}$ simply by playing 1 at every stage (this is a 0-recall strategy. As it turns out, there is no 1-EBR strategy that does strictly better than that against σ^1, σ^2).

Still, in the EBR model, we can get a simple example of successful online concealed correlation by players whose opponent is as strong as they are. We only need to slightly modify the three-player matching pennies stage game to get a repeated EBR game in which Rowena and Colin bring Matt's expected payoff down to below his i.r.p. in the stage game, by using the same strategies σ^1, σ^2 we have considered. Let $G\varepsilon$, depicted in Figure 2, be the stage game. Let Rowena and Colin use the strategies σ^1, σ^2 (these are

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

Matrix 0

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

Matrix 1

Figure 2: The game $G\varepsilon$

indeed 1-EBR mixed strategies), and let Matt use his best 1-EBR strategy response against these strategies. As we saw in Section 4.3, after observing Rowena and Colin play $(0,0)$ at the last stage, Matt ascribes the following probability to the forthcoming pair of actions:

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

Therefore, for ε small enough, Matt's best response in this case is 1. Note that σ^1 and σ^2 are not contingent on Matt's play. 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. Hence, he should indeed play 1 here.

For the other possible pairs of actions, we can do the computation as in Section 4.3, and get 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 $(1,1)$

Therefore Matt's best responses are 0, 0, and 1, respectively.

This defines σ^3, Matt's best 1-EBR response against σ^1, σ^2. We may now compute his expected payoff $r^3(\sigma^1, \sigma^2, \sigma^3)$. Here we use the first method mentioned in Section 4.1: we first compute the payoff over each deterministic play, and then average over plays. There are four possible plays (the same A, B, C, D we saw in Section 4.3, augmented by Matt's actions), and each has probability $\frac{1}{4}$. The play A yields an average payoff of 0 for Matt, since Matt will always play 1 (at least from the second stage on, since we have not considered his initial memory). 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, $r^3(\sigma^1, \sigma^2, \sigma^3) = (\frac{-1+\varepsilon}{4} + \frac{-1+\varepsilon}{4} + \frac{-1-\varepsilon}{2})/4 = -\frac{1}{4}$; so σ^1 and σ^2 guarantee that Matt's expected payoff in the 1-EBR infinitely repeated game will be no more than $-\frac{1}{4}$, which is indeed smaller than his i.r.p. in the stage game $G\varepsilon$.

Thus, in the EBR model, we have seen an example of online concealed correlation by players not stronger than their opponent (unlike in the main result, this was achieved without the help of a stronger player). In the sequel, we shall return to our previous model of bounded recall.

We shall be interested not only in an example with some numbers as recall capacities, but mainly in the asymptotic behavior, as the recall capacities go to infinity.

5 Online Concealed Correlation by "Weak" Players

Here we give an instance of the main result: an example of online concealed correlation by weak players, with the help of stronger ones.

5.1 The Game and the Strategies

Take the three-player matching pennies stage game depicted in Figure 1, and add a fourth player, Forest, to the game. Forest's recall will be larger 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 min-max 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 capacity $m_i(n)$, and $\forall i \ \lim_{n \to \infty} m_i(n) = \infty$. Still, $m_1 \leq m_2 \leq m_3 \leq m_4$, 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 $m_1 \le m_2 \le m_3 \le m_4$:

$$m_3 \ll m_4 \quad (\text{i.e., } \lim_{n\to\infty} \frac{m_3(n)}{m_4(n)} = 0) \tag{5.1}$$

$$\log m_3 \ll m_1,\, m_2 \qquad (m_3 \text{ is "subexponential" in } m_1, m_2). \tag{5.2}$$

Actually, instead of (5.2), we can settle for the following two weaker requirements:

$$(m_3)^6 \ll 2^{m_2} \tag{5.2a}$$

$$(m_3)^4 \ll (m_1)^4 \cdot 2^{m_1} \ , \tag{5.2b}$$

which are both implied by (5.2).

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 $A_4 = \{0, 1, x\}$.

$$\text{Let } K(n) = \frac{m_1(n)}{2} - 1 \tag{5.3}$$

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

$$L \le m_4 - K \tag{5.4}$$

$$m_3 \ll L \tag{5.5}$$

$$L^6 \ll 2^{m_2} \tag{5.6a}$$

$$L^4 \ll (m_1)^4 \cdot 2^{m_1}\ , \tag{5.6b}$$

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 $x_1, x_2, \ldots$ ($x_i \in \{0,1\}$), where $x_1, \ldots, x_L$ is chosen with the conditional probability of an i.i.d. sequence distributed $\frac{1}{2} - \frac{1}{2}$, given that:

$$\forall s,t \text{ s.t. } 1 \le s < t \le L \ \ \exists\, 0 \le i < m_2 \text{ s.t. } x_{s+i} \ne x_{t+i}\,. \tag{5.7}$$

Colin's strategy, σ^2, is to play the chosen sequence (an alternative description of σ^2: choose at random an L-periodic sequence $x_1, x_2, \ldots$ that obeys (5.7), with uniform probability over all such sequences; play this sequence).

Condition (5.7) 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	$f(y_1 \ldots y_K)$	....
Forest	....	x	$y_1 \ldots y_K$			....

Thus, if Forest's block was $(y_1, \ldots, y_K)$, then the block Rowena plays is $f(y_1, \ldots, y_K)$. Equality (5.3) 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 Current

He finds his own block that is L stages back ((5.4) 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 $f(\bar{y})$ as 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	...	$1\mid\ z_1^1 \dots z_K^1$	$1\mid\ z_1^2 \dots z_K^2$	...	$1\mid\ z_1^c \dots z_K^c$	...
Colin	...	$*\mid\ z_1^1 \dots z_K^1$	$*\mid\ z_1^2 \dots z_K^2$	...	$*\mid\ z_1^c \dots z_K^c$	...
Forest	...	$x\mid f^{-1}(z_1^2 \dots z_K^2)$	$x\mid f^{-1}(z_1^3 \dots z_K^3)$	...	$x\mid f^{-1}(z_1^1 \dots z_K^1)$	...

L

5.2 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 $\leq -\frac{1}{2}+\varepsilon$. In other words, let $\sigma^{-3} = (\sigma^1, \sigma^2, \sigma^4)$. Then,

$$\lim_{n\to\infty} \max_{\tau \in BR^3(m_3)} \bar{r}^3(\tau, \sigma^{-3}) = -\frac{1}{2}\,. \tag{5.8}$$

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. (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 $x_1, \ldots, x_L$ that satisfies (5.7), 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 (5.7) is arbitrarily close to 1. For this we can use the following, more general, claim (see [18, pp. 247-248]):

Lemma 5.1. *Let $l : \mathbb{N} \to \mathbb{N}$, and let $x_1, x_2, \ldots$ 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 $1 \leq i \neq j \leq l(n)$ $\Pr(x_i = x_j) \leq \alpha$, we have*

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

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

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

and by (5.6a), 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 $\Pr(\bar{z} \text{ contains two identical blocks}) \leq$

$\binom{c}{2}2^{-K}$, where $c=\frac{L}{K+1}$ is the number of blocks, and this probability converges to 0, by (5.3) and (5.6b). Now, the sequence $x_1,\ldots,x_L$ 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 $x_1,\ldots,x_L$.

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 (5.8):

Claim 5.2. *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[9] $B=(t_1,\ldots,t_L)$ 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 BR) 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 induced by τ and σ^{-3}. Since the actions $\bar{x}$ are independent throughout B (and independent of $\bar{\alpha}$), $\forall t\in B\ \ E_{\sigma^{-3}}(r^3(a_t)\mid a_{t_1},\ldots,a_{t-1})=-\frac{1}{2}$ ($\bar{x}$ 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., [1, p. 79]) implies that

[9] The proof is an adaptation of the proof in [18, p. 238] of Ben-Porath's results for finite automata [3],[4].

for every $\varepsilon > 0$, there exists $C(\varepsilon) > 0$, s.t.

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

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

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

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

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 (5.5), $|\Theta| \cdot e^{-C(\varepsilon)\cdot L} \to 0$, 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}.

6 Online Concealed Correlation by Finite Automata

Here we demonstrate that the same type of online concealed correlation achieved in the example of Section 5, where the players were restricted to

BR 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 BR 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 $\mathcal{A}$ induces 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 current state of $\mathcal{A}$ at stage t. At stage 1, $z_1=q_1$. At stage $t+1$, $z_{t+1}=g(s_t,a_t^{-i})$, 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 pure strategies induced 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., [18, p. 247]). Or in symbols (and identifying a strategy with its equivalence class), $BR^i(m)\subset\Sigma^i(|A|^m)$.

Consider the infinite repetition of the same stage game as in Section 5, when players use strategies induced by automata of sizes $(s_i)_{i\in N}$. Let player i's automaton size, s_i, correspond to $|A|^{m_i}$, where m_i is the recall capacity of player i in Section 5. Formally, the s_i's are again functions of a parameter n, and $\forall i\ \lim_{n\to\infty}s_i(n)=\infty$. We require that (analogously to (5.1) and (5.2)):

$$\log s_3 \ll \log s_4 \tag{6.1}$$

$$\log\log s_3 \ll \log s_1,\, \log s_2\,. \tag{6.2}$$

We claim that in this setup, too, as n goes to infinity, Matt's i.r.p. (at best) approaches his minmax in correlated actions, i.e.,

$$\lim_{n\to\infty} \min_{\sigma^{-3}\in\Delta(\times_{j\neq 3}\Sigma^j(s_j))} \max_{\tau\in\Sigma^3(s_3)} \bar{r}^3(\tau,\sigma^{-3}) \leq \min_{y\in\Delta(A_{-3})} \max_{x\in A_3} r^3(x,y) = -\frac{1}{2}. \quad (6.3)$$

Let $s_i = |A|^{m_i}$, for $i = 1, \ldots, 4$. Rowena, Colin, and Forest will use exactly the same strategies as in Section 5. Each of them can implement their strategy by a mixture of automata of sizes $s_i = |A|^{m_i}$, as stated above. So we only need to verify again that Matt can do no better than $-\frac{1}{2}+\varepsilon$ against this play, this time with a strategy in $\Sigma^3(s_3)$.

Reviewing the analysis of Matt's payoff, we see that everything stays just the same, up to near the end of the Proof of Claim 5.2, where we finally utilize the fact that Matt is restricted to an m_3-recall strategy. There, as Matt reaches the beginning of the cycle B, his play is determined by the m_3-recall strategy he chose, plus whatever memory he may possess at that point. For an automaton, the state replaces the memory: Matt's play from the beginning of the cycle B is determined by the automaton he chose at the beginning of the game and by the current state of the automaton at that point. The number of possible states is $s_3 = |A|^{m_3}$, which is equal to the number of possible m_3-length memories, and this completes the proof (indeed, proof of a claim equivalent to Claim 5.2, in the context of finite automata, originally appeared in [3], [4]).

Requirement (6.2) can be relaxed. For s_1, we can use the requirement analogous to (5.2b). Better yet, for s_2 we need only require that

$$\log s_3 \ll s_2 \quad (6.4)$$

because Colin can play a random i.i.d. sequence of length $L = s_2$, using a mixture of automata of size s_2 (see [3], [4]). In the proof we needed $\log s_3 \ll L$; hence, (6.4) is sufficient.

Note that we have asserted only an inequality in (6.3), not an equality (as should have also been noted in the previous case of bounded recall). If Forest is much stronger than Matt, it is conceivable that he may learn Matt's strategy, and help Rowena and Colin not only to conceal the correlation of their actions, but also to predict Matt's play, and thus do even better than the minmax in correlated actions.

However, if Forest's automaton size is subexponential in Matt's (i.e., $\log s_4 \ll s_3$), Matt can make sure that the odds for a correct prediction are arbitrarily small (see [3], [4]), and this also holds for BR, where the recall capacity of the strong player is subexponential in the weak player's capacity (see [15]). Therefore, if this is the case, the inequality can be replaced by an equality.

7 Main Result

So far, we have discussed one specific example, based on the three-player matching pennies stage game. 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 5, 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. However, this 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 $n \to \infty$.

However, there is 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 5.2, 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.

For $\overrightarrow{m} = (m_1, \ldots, m_k)$ denote by $\bar{v}^i(\overrightarrow{m})$ the i.r.p. of player i in the repeated game with recall levels $(m_j)_{j=1}^k$, namely,

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

and denote by $\underline{v}^i(\overrightarrow{m})$ his maxmin level, namely,

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

Theorem 7.1. *Let $G = (N, A, r)$ be a four-player game ($N = \{1, 2, 3, 4\}$), s.t. $|A_4| \geq |A_1|$ (or $|A_4| \geq |A_2|$), and let $m_i : \mathbb{N} \to \mathbb{N}$ $(i = 1, \ldots, 4)$ satisfy*

$\forall i \ \lim_{n\to\infty} m_i(n) = \infty$

$m_3 \ll m_4$

$\log m_3 \ll m_1, m_2$.

Then,

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

where z is the uniform distribution over A_4, and $\overrightarrow{m} = (m_1, \ldots, m_4)$.
In particular, if, in addition, the payoff function r^3 is independent of the actions of player 4, then

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

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

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

(As in Section 6, an analogous theorem may be stated for finite automata.)

Proof. 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

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

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 [15], 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 5.1, 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 5.1. Again, Colin chooses a random L-periodic sequence $x_1,x_2,\ldots$ ($x_i\in A_2$) that satisfies (5.7), 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 5.1, 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 5.1, 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[10]). 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 5.2 completes the proof. □

If we allow Matt's bounded recall strategy to be time-dependent, i.e., let the actions at every stage t depend both on t and on the memory, then the result still holds, since the proof of Claim 5.2 need not be altered. The situation is the same for a time-dependent automaton, e.g., an automaton whose action may depend both on t and on the current state. Moreover, the result holds if, in addition, Matt's strategy is also allowed to be behavioral.

7.1 More players

We may naturally ask, what happens when there are more than four players? For example, suppose there are ten weak players who want their actions to be correlated and concealed from the eleventh player, and a strong twelfth player is interested in helping them out. If the strong player has many possible

[10] 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 (see Section 8.2).

actions (in the stage game), e.g., $|A_{12}| \geq \times_{j=2}^{10} |A_j|$, then the same scheme can still be followed when the above relations hold between the strength levels of the strong, weak, and intermediate players. The first player (playing the role of Colin) chooses a random sequence with his desired distribution, and the other nine players are instructed by the strong player on how to correlate their actions with that player. At every stage, a choice of action by the strong player contains nine instructions (similar to nine bits of information), one for each player (w.l.o.g. $A_{12} = A_2 \times A_3 \times \ldots \times A_{10}$). Alternatively, if there are many strong players, none possessing that large a set of actions, they may divide the instructing tasks among themselves.

In Section 8.2 we see how the requirement on the instructing player's set of actions can be weakened.

8 Generalizations and Extensions

8.1 Concealed distribution

A profile of mixed (BR or finite automata) strategies $\sigma = (\sigma^1, \ldots, \sigma^n)$ induces a probability distribution over periodic plays of the repeated game, thereby inducing the average limiting frequency of any action profile $a = (a^1, \ldots, a^n) \in A$ (the average "empirical" probability of a), as in (4.1). Thus, σ induces the average empirical distribution $D_\sigma \in \Delta(A)$. For $K \subset N$ denote $A_K = \times_{k \in K} A_k$ and let $D_\sigma \,|\, A_K$ 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, $D^J \in \Delta(A_J)$ a given distribution over their actions, player $i \notin J$ the opponent, and $\mathcal{C}^i$ a class of strategies of i. We would say that the strategy tuple σ^{-i} of players $N \backslash i$ 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., [6]), 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.)

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 $\forall \sigma^i \in \mathcal{C}^i \;\; H_{\sigma^{-i},\sigma^i}(x^J) - H_{\sigma^{-i},\sigma^i}(x^J \mid x^i) \leq \varepsilon$.

If $|A_i| \geq |A_J|$ then for any distribution $D^J \in A_J$ there exists a payoff function r, s.t. D^J is the unique distribution over A_J that guarantees that i's payoff in the stage game $G = (N, A, r)$ will not exceed his minmax in correlated actions, $\underline{v}^i(G)$. Then, if σ^{-i} ε-conceals D^J for ε small enough, then $\forall \sigma^i \in \mathcal{C}^i \; \bar{r}^i(\sigma^{-i}, \sigma^i) < \underline{v}^i(G) + \delta$. In particular, if D^J is correlated and ε is small enough, then $\bar{r}^i(\sigma^{-i}, \sigma^i)$ is smaller than i's i.r.p. in the stage game.

8.2 Fewer actions

For simplicity of notation we assume in the discussion below that the set of players is $N = \{1, \ldots, k+1\}$. Let $J = \{1, \ldots, k-1\}$; the scheme in Section 7 implements the concealment of a distribution $D^J \in \Delta(A_J)$ from player k, when player $k+1$ has a large enough set of actions, i.e.,

$$\exists \widetilde{J} \subset J \text{ with } |\widetilde{J}| = k-2, \text{ s.t. } |A_{k+1}| \geq \times_{j \in \widetilde{J}} |A_j| \,.$$

We can relax this requirement; first, a standard information-theoretic argument shows that it is sufficient to require that $\log_2 |A_{k+1}| \geq \sum_{j \in \widetilde{J}} H(D^j)$, where $D^j = D^J \mid A_j$, since we need to dedicate only $H(D^j)$ bits per stage to encode a "typical sequence" distributed according to D^j. This condition is, however, still not tight; consider the case where D^J is a product distribution: the members of J need no instructions at all in order to conceal D^J. Therefore it seems that the right condition here is as follows.

Theorem 8.3. *Let $G = (N, A, r)$ be a $(k{+}1)$-player game and let $m_i : \mathbb{N} \to \mathbb{N}$ $(i = 1, \ldots, k{+}1)$ satisfy*

$\forall i \ \lim_{n\to\infty} m_i(n) = \infty$

$m_k \ll m_{k+1}$

$\log m_k \ll \min_{i=1}^{k-1} m_i$.

Let $J = \{1, \ldots, k{-}1\}$, and let D^J be a distribution on A_J with marginals D^j on A_j. If

$$\log_2 |A_{k+1}| \geq (\sum_{j\in J} H(D^j)) - H(D^J),$$

then there is a strategy tuple $\sigma^{-k} \in \times_{j\neq k}\Delta(BR^j(m_j))$ such that the tuple σ^{-k} $\varepsilon(n)$-conceals D^J against $BR^k(m_k)$ where $\varepsilon(n) \to 0$ as $n \to \infty$.

The proof of the theorem uses a refined random dictionary of players $1, \ldots,$ $k-1$, together with a more delicate instruction scheme.

The concealment of a distribution on A_{-k} (i.e., where the actions of player $k+1$ do matter) is more delicate. Applying the tools developed in [10] in concert with the scheme used in the present paper enables us to provide sufficient conditions for the concealment of such a distribution.

Theorem 8.4. *Let $G = (N, A, r)$ and m_i $(i = 1, \ldots, k{+}1)$ be as in Theorem 8.3. Let D^{-k} be a distribution on A_{-k} with marginals D^j on A_j. If*

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

then there is a strategy tuple $\sigma^{-k} \in \times_{j\neq k}\Delta(BR^j(m_j))$ such that σ^{-k} $\varepsilon(n)$-conceals D^{-k} against $BR^k(m_k)$ where $\varepsilon(n) \to 0$ as $n \to \infty$.

Note that this is a generalization of Theorem 8.3: The special case in which the distribution D^{-k} is the cartesian product of D^J and the uniform distribution over A_{k+1} $(J = \{1, \ldots, k{-}1\})$ gives Theorem 8.3 as a corollary.

9 Equilibrium Payoffs

A "folk theorem" is a characterization of the set of equilibrium payoffs in the infinitely repeated game, by means of the data of the stage game. For $\overrightarrow{m} = (m_1(n), \ldots, m_k(n))$, let $G_\infty(\overrightarrow{m})$ denote the infinitely repeated game, where player i uses an $m_i(n)$-recall strategy (or analogously an $m_i(n)$-state automaton). A folk theorem here is a characterization of the asymptotics (i.e., the limit as $n \to \infty$) of $NE(G_\infty(\overrightarrow{m}))$, the set of Nash equilibrium payoffs of this game. In some cases, the existence of a limit is not known, and one may study the limsup and liminf of the set.

We confine ourselves here to the case where each player's capacity is subexponential in any other player's, i.e., $m_k \ll \log(m_1)$. The case of two-player games was solved for finite automata in [3], [4], and for bounded recall in [15]. For more than two players, the issue is wide open; existing theorems give upper and lower bounds for $NE(G_\infty(\overrightarrow{m}))$. The asymptotics of this set essentially boils down to the asymptotics of the i.r.p.'s of the players, $\bar{v}^1(\overrightarrow{m}), \ldots, \bar{v}^k(\overrightarrow{m})$. To find these i.r.p's, we need to account for the possibilities of concealed correlation.

Theorem 8.3 gives us, as a corollary, a folk theorem for a large class of games. We may think of player $k+1$ in Theorem 8.3 as a mechanism designer, instead of a participant. This mechanism designer has a large enough set of signals, and sends a *public* signal at every stage. In the presence of such a mechanism designer, each player i's i.r.p. is his minmax in correlated actions in the stage game, denoted u^i. Consequently, the set of equilibrium payoffs converges to the set $F \cap IRc$, where F is the mixture of of feasible payoffs and IRc is the set $\{(x^1, \ldots, x^k) \mid \forall i \; x^i \geq u^i\}$. Thus,

Theorem 9.1. *Let $G^D_\infty(\overrightarrow{m})$ denote a k-player infinitely repeated game, where player i uses an $m_i(n)$-recall strategy, with a mechanism designer sending a public signal at each stage, out of a large enough set of signals. Assume that $\forall i \;\; \log m_i \ll \min_{i=1}^k m_i$ and $\lim_{n\to\infty} m_i(n) = \infty$. Then,*

$$\lim_{n\to\infty} NE(G^D_\infty(\overrightarrow{m})) = F \cap IRc \; .$$

Note that the set $F \cap IRc$ is the set of correlated equilibrium payoffs in a classic k-player repeated game, without any rationality bounds (this is the "Correlated Folk Theorem").

When player $k+1$ is a participant in the game, rather than a mechanism designer, Theorem 8.4 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.

10 Remarks

1. In [16] it is proved, in particular, that if player i is the weakest in the field (i.e, $i=1$), he can still conceal his own chosen pure strategy from the other strong players, when their recall capacity is subexponential in his. However, our example in Section 5 disproves Theorem 1 in [16], due to the possibility of concealed correlation by players weaker than i (when $i \geq 3$).

2. We have only discussed infinitely repeated games where the payoff is the limiting average payoff. Note, however, that in every case we have discussed the play of the relevant players enters a cycle at some point, and the time it takes to enter that cycle does not depend upon the payoff function. Therefore, our discussion and results apply to finitely repeated games, provided the number of repetitions is sufficiently long, and to games with discounted payoffs, provided the discount factor is close enough to 1.

3. We may wonder whether in a classic repeated game, without any rationality bounds, a non-product distribution can be ε-concealed. 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.

References

[1] Alon, N., and Spencer, J., (1992) *The Probabilistic Method*, New York: Wiley-Interscience.

[2] Aumann, R. J., (1981) Survey of repeated games, in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Mannheim, Wien, Zurich: Wissenschaftsverlag, Bibliographisches Institut, 11–42.

[3] Ben-Porath, E., (1986) Repeated games with bounded complexity (in Hebrew), M.Sc. Thesis, Hebrew University, and Stanford University (mimeo).

[4] Ben-Porath, E., (1993) Repeated games with finite automata, *Journal of Economic Theory* **59**, 17–32.

[5] Berry, D. A., and Fristedt, B., (1985) *Bandit Problems: Sequential Allocation of Experiments*, London: Chapman and Hall.

[6] Cover, T. M., and Thomas, J. A., (1991) *Elements of Information Theory*, New York: Wiley-Interscience.

[7] Gilboa, I., and Schmeidler, D., (1994) Infinite histories and steady orbits in repeated games, *Games and Economic Behavior* **6**, 370–399.

[8] Gossner, O., (1999) Repeated games played by cryptographically sophisticated players, DP 99-07, THEMA.

[9] Gossner, O., (2000) Sharing a long secret in a few public words, DP 2000-15, THEMA.

[10] Gossner, O., Hernandez, P., and Neyman, A., (2004) Optimal use of communication resources , DP 377, Center for the Study of Rationality, Hebrew University, Jerusalem.

[11] Gossner, O., and Tomala, T., (2004) Secret correlation in repeated games with imperfect monitoring, DP 2004-21, Cahiers du CEREMADE, Université Paris Dauphine, Paris.

[12] Harvey, A. C., (1993) *Time Series Models*, 2nd edition, New York: Harvester-Wheatsheaf.

[13] Kalai, E., (1990) Bounded rationality and strategic complexity in repeated games, in *Game Theory and Applications*, edited by T. Ichiishi, A. Neyman and Y. Tauman, San Diego: Academic Press, 131–157.

[14] Kuhn, H. W., (1953) Extensive games and the problem of information, in *Contributions to the Theory of Games I*, edited by H. W. Kuhn and A. W. Tucker, Princeton: Princeton University Press, 193–216.

[15] Lehrer, E., (1988) Repeated games with stationary bounded recall strategies, *Journal of Economic Theory* **46**, 130–144.

[16] Lehrer, E., (1994) Finitely many players with bounded recall in infinite repeated games, *Games and Economic Behavior* **7**, 390–405.

[17] Neyman, A., (1985) Bounded complexity justifies cooperation in the finitely repeated prisoner's dilemma, *Economics Letters* **19**, 227–229.

[18] Neyman, A., (1997) Cooperation, repetition, and automata, in *Cooperation: Game-Theoretic Approaches*, edited by S. Hart and A. Mas-Colell, Berlin: Springer, 233–255.

[19] Neyman, A., (1998) Finitely repeated games with finite automata, *Mathematics of Operations Research* **23**, 513–552.

[20] Papadimitriou, C. H., and Yannakakis, M., (1994) On complexity as bounded rationality (extended abstract), STOC-94, 726–733.

[21] Osborne, M. J., and Rubinstein, A., (1994) *A Course in Game Theory*, Cambridge, MA: MIT Press.

[22] Rabiner, L. A., (1989) A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE* **77**, 257–285.

[23] Rubinstein, A., (1986) Finite automata play the repeated prisoner's dilemma, *Journal of Economic Theory* **39**, 83–96.

[24] Sorin, S., (1997) Cooperation through repetition: Complete information, in *Cooperation: Game-Theoretic Approaches*, edited by S. Hart and A. Mas-Colell, Berlin: Springer, 169–198.