Learning to play partially-specified equilibrium

Ehud Lehrer[*]and Eilon Solan[†]

August 23, 2007

August 23, 2007
First draft: June 2007

ABSTRACT:

In a partially-specified correlated equilibrium (PSCE) the players are partially informed of the conditional strategies of the other players, and they best respond to the worst-case possible strategy. We construct a decentralized procedure that converges to PSCE when the monitoring is imperfect. This procedure is based on minimizing conditional regret when players obtain noisy signals that depend on the actions that have been previously played.

Journal of Economic Literature classification numbers: C61, C72, D81, D82, D83

Keywords: partially-specified correlated equilibrium, conditional no-regret, imperfect monitoring, approachability.

[*] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: lehrer@post.tau.ac.il. This author acknowledges the support of the Israel Science Foundation, Grant #762/045.

[†] School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. e-mail: eilons@post.tau.ac.il

1 Introduction

The two main solution concepts of strategic form games, Nash equilibrium (Nash, 1951) and correlated equilibrium (Aumann, 1974, 1987), are based on the implicit assumption that when players take actions they know, or have an accurate belief, regarding the joint strategy played by the other players. This paper deals with a variant of correlated equilibrium, called partially-specified correlated equilibrium (Lehrer, 2005), whereby each player has partial information about the strategies played by the other players.

A correlated strategy is a probability distribution over the joint strategies of all players. A mediator randomly selects a joint strategy according to a correlated strategy, and recommends each player to play her component. In a correlated equilibrium this recommendation, which has no effect on payoffs, is a best response to the conditional joint strategies played by the other players. Thus, a correlated equilibrium is a Nash equilibrium of an extension of the game, whereby prior to taking an action each player receives a private payoff-irrelevant signal.

In a partially-specified correlated equilibrium a mediator recommends a strategy to each player, but the correlated strategy is not fully known to the players. In particular, when a player receives a recommendation she does not know the conditional strategy of the other players. Rather, she receives a partial information about it. Formally, each player is informed of the expectation of some, but not all, random variables defined over the set of all other players' joint strategies. For instance, a player may be informed of the probability that the other players play a joint strategy in a certain set, without knowing the probability of each strategy profile within this set.

Realistically speaking, it is unreasonable to exclude the possibility that players have only partial knowledge about their opponents' behavior, and they are therefore lead to partially-specified equilibrium. Our goal is to study interactions with partial knowledge on the opponents' behavior, and to find a natural decentralized learning procedure that converges to the set of partially-specified correlated equilibria.

There are several known procedures that converge to the set of correlated equilibria. These procedures are based on a full monitoring assumption, or at least on a condition that amounts to observable payoff (see, for instance, Foster and Vohra (1997), Fudenberg and Levine (1999) and Hart and Mas-Colell (2000, 2001)).

We construct a procedure that converges to the set of partially-specified correlated equilibria in the presence of imperfect monitoring. As Hart and Mas-Colell (2000), this procedure is based on regret minimization.

When payoffs are observable a player's regret is the difference between the maximal payoff she could have received had she known at the outset of the game the empirical distributions of actions played by the other players, and her actual payoff. When the player does not observe her payoffs, but rather a noisy signal of these payoffs, she cannot calculate these two quantities, and hence she is unable to calculate her regret. The procedure we introduce is based on a general no-regret theorem for sequential decision problems with imperfect monitoring.

The issue of having no regret in such decision problems has been treated so far only in the sense that the average payoff over the entire history is no less than the worst payoff consistent with the signals obtained. This setup is known as external no-regret.

Consider a decision maker (DM) who chooses an action at every stage, and whose payoff depends on her action and on the realized state of nature. The evolution of the state variable is unknown. DM does not observe the state, but receives a noisy signal that depends on the state of nature and on the action chosen. In such a model of imperfect monitoring, the stage optimal strategy may be mixed. Indeed, consider, for instance, the 'matching pennies' game, where DM chooses a row, T or H (an "action"), nature chooses a column T or H (a "state"), and the payoff to DM is 1 if the two choices match, and 0 otherwise. When DM receives no information about the state, choosing T or H with equal probabilities guarantees her the best of the worst-case scenario payoffs. In this sense the uniform mixed strategy is the unique best response.

A strategy is *external regret-free* if for every finite horizon sufficiently long, the empirical distribution of the actions of DM is optimal assuming that nature chooses a stationary strategy that is consistent with the observed signals. In the example of the 'matching pennies' with no signals, the distribution $[\frac{1}{2}(T), \frac{1}{2}(H)]$ is optimal. Therefore, any strategy that chooses each action half of the times is external regret-free. In particular, the deterministic strategy that plays T in odd stages and H in even stages is external regret-free. However, this strategy is vulnerable to an adversarial nature which chooses H in odd stages and T in even stages (possibly after realizing that DM plays in a learnable pattern). The stationary strategy $[\frac{1}{2}(T), \frac{1}{2}(H)]$, however, is immunized against regrets, because nature cannot hurt DM even if it learns that DM uses it. In other words, the stationary strategy $[\frac{1}{2}(T), \frac{1}{2}(H)]$, guarantees the maxmin payoff against any choice of states.

Suppose that in a general sequential decision problem with imperfect monitoring DM uses a stationary strategy. The empirical frequency of the signals provides only a partial information about the frequency of states. DM can calculate the set of all distributions over states that are consistent with his observations, and may wonder whether he has done his best against this set of distributions. As different individuals might have different attitudes toward a set of plausible distributions, we allow DM to have his subjective response function. This function indicates what actions best respond to any set of the opponents' strategies in DM's subjective view. For example, a pessimistic DM would have a response function that chooses the actions that maximize the worst-case payoff.

Given a response function, a strategy in the sequential decision problem is *conditional regret-free*, if every mixed action that is played frequently is a best response (w.r.t. to the response function) to the set of all strategies that are consistent with the signals observed while playing this action.

Viewed differently, suppose that DM has several agents, each of them can execute a different mixed action. At every stage DM chooses one agent, ask him to play the game and report back the signal he observed. The strategy of DM is conditional no regret if the mixed action that is played by an agent that is chosen often is optimal given the signals that agent reports.

In the general case of imperfect monitoring we prove that if the response function satisfies weak conditions, a conditional regret-free strategy exists. When applied to the worst-case response function described above, we obtain a particularly important conditional regret-free strategy. It ensures that any mixed action which is frequently played, guarantees the maxmin level against an adversary nature restricted to choosing states according to the distributions that are informationally equivalent to the signals received.

This result enables us to define a procedure that converges to the set of partially specified correlated equilibria in an n-player game with imperfect monitoring. Suppose that each player faces a sequential decision problem, in which she perceives the other $n-1$ players as "nature". Analogously to the case of full-monitoring, if each player plays an conditional regret-free strategy w.r.t. the worst-case response function in her sequential decision problem, the play converges to the set of partially specified correlated equilibria. By letting the players play conditional regret-free strategies w.r.t. other response functions, we may obtain different notions of partially specified correlated equilibria.

Our results concerning the existence of conditional no-regret strategies extend those that study external no-regret strategies with imperfect monitoring, and those that study internal no-regret with perfect monitoring; these include Rustichini (1999), who stated that (external) regret-free strategies exists, Mannor and Shimkin (2003), who proved it when the player's choices do not affect the signals, Cesa-Bianchi, Lugosi, and Stoltz (2006), who showed a regret-free procedure under the condition that the payoff matrix of a player can be obtained through a linear transformation from the signalling structure, and Lugosi, Mannor and Stoltz (2007), who proved the general case. In another strand of the literature, Hart and Mas-Colell (1999, 2000) and Blum and Mansour (2007) dealt with internal regret in the perfect monitoring case.

The paper is arranged as follows. In Section 2 we define partially specified probabilities and partially specified correlated equilibrium. In Section 3 we present the model of sequential decision problem, and present our main results. Proofs appear in Section 4.

2 Partially-specified equilibrium

2.1 A partially-specified probability

An urn contains 90 balls, 30 are Red, 40 are Black, and 20 are White. This information is equivalent to saying that the expectation of the random variable[1] $[1(R), 0(B), 0(W)]$ is $\frac{1}{3}$, the expectation of the random variable $[0(R), 1(B), 0(W)]$ is $\frac{4}{9}$, and the expectation of the random variable $[0(R), 0(B), 1(W)]$ is $\frac{2}{9}$. Moreover, one can calculate the expectation of every random variable defined over the state space $\{R, B, W\}$.

Suppose now that it is known that 30 balls are Red and the other are either Black or White, but there is no indication as to how the Black and White balls are distributed; the distribution of colors is partially specified. The probability of only one non-trivial event is known: the probability of Red is $\frac{1}{3}$. Equivalently, the expectation of the random variable $[1(R), 0(B), 0(W)]$ is $\frac{1}{3}$, but the expectations of the two random variables $[0(R), 1(B), 0(W)]$ and $[0(R), 0(B), 1(W)]$ are unknown. Observe that the expectation of the random variable $[1(R), 1(B), 1(W)]$ is known as well — it is 1.

Suppose that as before there are 30 Red balls, 60 balls which are either Black or White, and an unknown number of Green balls. In addition it is known that the number of Green balls is the same as the number of White balls. The probability of Red in no

[1] That is, the probability of Red is 1, and the probability of both Black and White is 0. We identify a random variable with its distribution.

longer $\frac{1}{3}$. Furthermore, the probability of no non-trivial event is known. Nevertheless, some information about the distribution of colors is available. It turns out that this information is given by the expectation of two random variables.

Denote by X the random variable $[0(R), 0(B), 1(W), -1(G)]$ that takes the value 0 on Red and Black, 1 on White, and -1 on Green. Since the probabilities of White and Green are equal, the expectation of X is 0. Let Y be the random variable $[1(R), 0(B), 0(W), \frac{1}{3}(G)]$. It turns out[2] that the expectation of Y is $\frac{1}{3}$.

These examples motivate the following definition.[3]

Definition 1 *Let Ω be a finite set. A* partially-specified probability *over Ω is a pair $(P, \mathcal{Y})$, where P is a probability distribution over Ω, and $\mathcal{Y}$ is a set of random variables defined over Ω that contains the indicator of Ω.*

The interpretation of $(P, \mathcal{Y})$ is that only the expectations of the random variables in $\mathcal{Y}$ w.r.t. P are known. We call $\mathcal{Y}$ the *information structure.*

Example 1 Let Ω be a finite state space, and let P be a probability distribution over Ω. The states in Y are not directly observed. Rather, every $\omega \in \Omega$ is associated with a distribution, say $s(\omega)$, over a finite set of signals S. The states are realized sequentially. At every stage a state is randomly drawn from Ω according to P, independently of past drawings. When ω is selected the random signal $s(\omega)$ is observed. That is, the signal actually observed is randomly selected from S according to the distribution $s(\omega)$.

The long-run empirical frequency of the signals converges to $\sum_{\omega \in \Omega} P(\omega) s(\omega)$. In other words, the observer asymptotically learns the expectation of the $|S|$ random variables Y_t, $t \in S$, where for each $\omega \in \Omega$, $Y_t(\omega)$ takes the value $s(\omega)(t)$, which is the probability of the signal t according to the distribution $s(\omega)$. ■

The partially-specified probability $(P, \mathcal{Y})$ determines the set of all probability distributions that agree with P on the variables in $\mathcal{Y}$. Formally,[4] $C(P, \mathcal{Y}) = \{Q \in \Delta(\Omega);\ \mathbb{E}_Q(Y) = \mathbb{E}_P(Y) \text{ for every } Y \in \mathcal{Y}\}$. The set $C(P, \mathcal{Y})$ is convex and compact and, moreover, it is an intersection of $\Delta(\Omega)$ with a plane.

[2] Denote by n_i the number of balls of color i. Then $\frac{n_r}{n_r+n_b+n_w} = \frac{1}{3} = \frac{\frac{1}{3}n_g}{n_g}$ and therefore, $\frac{n_r+\frac{1}{3}n_g}{n_r+n_b+n_w+n_g} = \frac{1}{3}$.

[3] In the current context all the sets are finite and therefore measurability consideration is ignored. However, in general those should be taken into consideration, and the definition should be amended accordingly.

[4] For every finite set S, $\Delta(S)$ is the set of probability distributions over S.

2.2 The n-player game and partially-specified correlated strategies

Let $G = (N, \{A_i\}_{i \in N}, \{u_i\}_{i \in N})$ be an n-player game, where N is the set of the n players, A_i and u_i are player i's finite action set and utility function.

The concept of correlated equilibrium (Aumann, 1974, 1987) refers to a case where the players obtain some (correlated) information prior to playing the game. This information need not be related to the game itself (and usually is not). Neither it is related to the state of nature in the case of a game with incomplete information, nor to the payoffs. The actions taken by a player depend only on her information.

In a correlated equilibrium with a fully-specified probability, the information a player obtains prior to playing the game can be reduced to a particular pure strategy. The incentive-compatibility conditions of the equilibrium determine that this pure strategy is a best response to the conditional mixed strategy played by the others. The reason why the reduction to pure strategies is not restrictive is that when a mixed strategy is a best response, every pure strategy in its support is a best response as well. However, in the case that the probabilities are partially specified, this is no longer true. Typically, when a mixed strategy is a best response to the other players' partially-specified joint strategy, neither of the pure strategies played with a positive probability is a best response. Therefore, in partially-specified correlated equilibrium, the recommendation received by a player is to play a mixed strategy.

Formally, prior to playing the game, a point $(p_1, ..., p_n) \in \times_{i \in N} \Delta(A_i)$ is randomly selected according to a distribution Q over $\times_{i \in N} \Delta(A_i)$. Player i is informed of p_i and, in addition, receives partial information regarding $Q(\cdot|p_i)$, i.e., the expectation w.r.t. $Q(\cdot|p_i)$ of some, but not all, random variables defined over $\times_{j \neq i} \Delta(A_j)$. W.l.o.g. we assume that the information that the player receives depends only on p_i, and not on p_{-i}.[5] The set of these random variables is denoted by $\mathcal{Y}_{i,p_i}$. We denote $\mathcal{Y} = (\mathcal{Y}_{i,p_i})_{i \in N, p_i \in \Delta(A_i)}$. Having received the recommendation to play p_i, player i now possesses the partially-specified probability $(Q(\cdot|p_i), \mathcal{Y})$.

The pair $(Q, \mathcal{Y})$, which specifies the distribution according to which profiles of mixed actions are drawn and the information known to each player, is called a *partially-specified*

[5] If this is not the case, and different mixed profiles p_{-i} give player i different information, we can consider an equivalent situation in which the information that player i receives upon receiving the recommendation p_i already incorporates the information that depends on p_{-i}.

correlated strategy. The following example clarifies the definition of this notion.

Example 2 *Consider a three player game, with the action sets $\{T, B\}$ for player 1, $\{L, R\}$ for player 2, and $\{W, E\}$ for player 3. A mixed action profile is denoted by three numbers (x_1, x_2, x_3) in $[0, 1]^3$, where x_i is the probability that player i to play his first action. Let Q be the distribution that gives weight $\frac{1}{8}$ to each of the 8 mixed action profiles in the set $\{\frac{1}{3}, \frac{1}{2}\}^3$. Define the collection $\mathcal{Y}$ that describes the information of the players as follows.*

$$\mathcal{Y}_{1,\frac{1}{3}} := \{x_2, 1\};$$

player 1, upon receiving the recommendation $\frac{1}{3}$, only knows the recommendation of player 2. In addition,

$$\mathcal{Y}_{1,\frac{1}{2}} := \{x_2 - x_3, 1\}.$$

The interpretation of this information structure is that when receiving the recommendation $\frac{1}{2}$, player 1 cannot distinguish between the cases where the recommendations of the other players differ (i.e., players 2 and 3 both received the recommendation $\frac{1}{3}$ and the case where both received the recommendation $\frac{1}{2}$). However, if the recommendations of players 2 and 3 are different, then player 1 is fully informed of these recommendations.

Define

$$\mathcal{Y}_{2,\frac{1}{3}} := \{x_1 + x_3, 1\}.$$

When receiving the recommendation $\frac{1}{3}$, player 2 cannot distinguish between cases where the recommendations of the other players differ. The sets $\mathcal{Y}_{2,\frac{1}{2}}$, $\mathcal{Y}_{3,\frac{1}{3}}$, and $\mathcal{Y}_{3,\frac{1}{2}}$ can be similarly defined.

Note that according to the definition, the information a player obtains about the others may depend on her action.

2.3 Partially-specified correlated equilibrium

A partially-specified correlated strategy $(Q, \mathcal{Y})$ is a *partially-specified correlated equilibrium* if the recommended action of a player guarantees the maxmin level against the other players' action profile that are informationally equivalent to the conditional action profile actually being used. Formally, denote by $p_{-i}|p_i$ the expected conditional mixed action profile of players $-i = N \setminus \{i\}$, w.r.t. Q, when player i is recommended to play

p_i. When player i obtains the recommendation to play p_i he obtains also a partially-specified probability $(p_{-i}|p_i,\mathcal{Y}_{i,p_i})$ about the other players' conditional action. The set $C(p_{-i}|p_i,\mathcal{Y}_{i,p_i})$ consists of all joint strategies $\widehat{p}_{-i}$ of players $-i$ that are indistinguishable from $p_{-i} \mid p_i$ given player i's information. That is, $\widehat{p}_{-i} \in C(p_{-i}|p_i,\mathcal{Y}_{i,p_i})$ if and only if under $\widehat{p}_{-i}$ the expectations of the random variables in $C(p_{-i}|p_i,\mathcal{Y}_{i,p_i})$ under $\widehat{p}_{-i}$ match the expectations under p_i. Denote,

$$U_i\Big(p_i,(p_{-i}|p_i,\mathcal{Y}_{i,p_i})\Big) = \min_{q\in C(p_{-i}|p_i,\mathcal{Y}_{i,p_i})} u_i(p_i,q).$$

$U_i\Big(p_i,(p_{-i}|p_i,\mathcal{Y}_{i,p_i})\Big)$ is the worst case payoff when player i plays p_i and players $-i$ play a strategy in $C(p_{-i}|p_i,\mathcal{Y}_{i,p_i})$.

Definition 2 *A partially-specified correlated strategy* $(Q,\mathcal{Y})$ *is an* ε-partially-specified correlated equilibrium (ε-PSCE) w.r.t. $\mathcal{Y}$ ($p_i \in \Delta(A_i)$) if with Q-probability $1-\varepsilon$, for every player $i \in N$, the mixed strategy p_i maximizes, up to an ε, player i's payoff: for every $p'_i \in \Delta(A_i)$,

$$U_i\Big(p_i,(p_{-i}|p_i,\mathcal{Y}_{i,p_i})\Big) \geq \max_{p'_i} U_i\Big(p'_i,(p_{-i}|p_i,\mathcal{Y}_{i,p_i})\Big) - \varepsilon. \quad (1)$$

If, in addition, for every player i all the sets $(\mathcal{Y}_{i,p_i})_{\{p_i\colon Q(p_i,\cdot)>0\}}$ are the same, we say that the information structure is *strategy independent*.

The information structure that we defined in Example 2 is not strategy independent.

Definition 3 *A correlated strategy $Q \in \Delta(\times_i\Delta(A_i))$ is an ε-partially-specified correlated strategy if there is an information structure $\mathcal{Y}$ such that $(Q,\mathcal{Y})$ is* ε-partially-specified correlated equilibrium (ε-PSCE) w.r.t. $\mathcal{Y}$.

In the following example, the information structure is strategy independent, namely the information a player has about the other players' strategies does not depend on the player's strategy.

Example 3 *Consider the two-player game that appears in Figure 1(A), and consider the distribution Q over $\Delta(A_1)\times\Delta(A_2)$ defined as $Q((\frac{1}{3},\frac{2}{3},0),(\frac{1}{2},\frac{1}{2},0)) = Q((\frac{1}{3},\frac{2}{3},0),(0,0,1))$ $= Q((0,0,1),(\frac{1}{2},\frac{1}{2},0)) = \frac{1}{3}$. That is, with probability $\frac{1}{3}$ each of the three joint distributions $((\frac{1}{3},\frac{2}{3},0),(\frac{1}{2},\frac{1}{2},0))$, $((\frac{1}{3},\frac{2}{3},0),(0,0,1))$ and $((0,0,1),(\frac{1}{2},\frac{1}{2},0))$ is chosen (see Figure 1(B)).*

	L	C	R
T	$18, 0$	$0, 12$	$2, 7$
M	$0, 12$	$9, 0$	$2, 7$
B	$7, 2$	$7, 2$	$0, 0$

Part A

<table>
<tr><td></td><td>$\frac{1}{2}$</td><td>$\frac{1}{2}$</td><td>1</td></tr>
<tr><td>$\frac{1}{3}$</td><td colspan="2" rowspan="2">$\frac{1}{3}$</td><td rowspan="2">$\frac{1}{3}$</td></tr>
<tr><td>$\frac{2}{3}$</td></tr>
<tr><td>1</td><td colspan="2">$\frac{1}{3}$</td><td>0</td></tr>
</table>

Part B

Figure 1: the game and the distributions.

Suppose that player 1 is informed only of the conditional probability of $\{L, C\}$ and that player 2 is informed only of the conditional probability of $\{T, M\}$. In particular, when player 1 receives the recommendation to play $(\frac{1}{3}, \frac{2}{3}, 0)$, she knows that player 2 will play a strategy in $\{L, C\}$ with probability $\frac{1}{2}$ and the pure strategy R with probability $\frac{1}{2}$. Moreover, if she receives the recommendation $(0, 0, 1)$ (i.e. the pure strategy B), player 1 knows that player 2 will play one of the strategies, L or C, with probability 1.

The mixed strategy $(\frac{1}{3}, \frac{2}{3}, 0)$ is the unique best response to the strategy specified as $\frac{1}{2}$ on $\{L, C\}$ and $\frac{1}{2}$ on the pure strategy R. Indeed, this is the unique maxmin strategy in the upper-left game in Figure 1(A). Furthermore, playing B is a best response to the strategy that plays either L or C with probability 1. Similar arguments hold for player 2, and therefore Q is a 0-partially-specified correlated equilibrium. ■

3 Conditional regret-free strategies when probabilities are partially specified

In this section we describe our no-regret results that are used to prove the main result of the paper. These results are important on their own right.

3.1 The model of sequential decision problems with imperfect monitoring

At every period the decision maker (DM) takes an action in a finite action space A, and nature chooses a state in a finite state space Ω. The payoff function of DM is given by a function $u : A \times \Omega \to \mathbb{R}$. Without loss of generality we assume that all payoffs are bounded between -1 and 1. The monitoring structure of DM is given by a function

$s : A \times \Omega \to \Delta(S)$, where S is a finite set of signals. We assume throughout that the signal DM receives reveals his action.

Denote by $X = \Delta(A)$ and $Y = \Delta(\Omega)$ the sets of mixed strategies of DM and nature, respectively. For every $x \in X$ and $y \in Y$, $s(x,y) := \sum_{a\in A}\sum_{\omega\in\Omega} x_a y_\omega s(a,\omega)$ is the mixed signal induced by x and y. Since the signal contains the action of DM, given $s(x,y)$ one can calculate $s(a,y)$ for every $a \in \text{supp}(x)$, the average signal conditional that DM chooses repeatedly the action a.

Set $\mathcal{M} = \{s(x,y) \in \Delta(S) \colon x \in X, y \in T\}$. This set contains all possible distributions over signals that DM may observe, when both DM and nature use stationary strategies. We refer to an element $\mu \in \mathcal{M}$ as a *footprint*.

Since the signal contains the action of DM, one can uniquely recover x from $s(x,y)$. In particular, μ contains the information about the probability that a pure action a is played. We denote this probability by $\mu(a)$.

If $s(a,y) = s(a,y')$ for every action a in some subset of actions $A' \subseteq A$, then DM cannot distinguish between the two stationary mixed strategies of nature y and y' by using only actions in A'. For every $\mu \in \mathcal{M}$, denote by $Y(\mu)$ the set of all stationary strategies of nature that are consistent with the average signal μ:

$$Y(\mu) = \{y \in Y \colon s(a,y) = \mu(a) \quad \forall a \text{ s.t. } \mu(a) > 0\}.$$

When nature plays y repeatedly, and DM plays repeatedly x, all DM knows is that nature's mixed action is in $Y(s(x,y))$. As explained in Example 1, this is equivalent to obtaining a partial specification about y.

3.2 Response functions

Suppose that DM holds a partially specified probability about nature, and needs to take a mixed action from the set X. As described before, DM can calculate at every stage the set of all probability distributions over states that are consistent with his signals, or equivalently, on the set of stationary strategies of nature. Different DM's may react differently to the same set of possible stationary strategies played by nature. A decision maker is characterized by a response function defined as follows:

Definition 4 *A* response function *of player 1 is a set-valued function* $\mathrm{R} : [0,1]\times\mathcal{M} \to X$ *that satisfies* $\mathrm{R}_\varepsilon(\mu_1) = \mathrm{R}_\varepsilon(\mu_2)$ *whenever* $Y(\mu_1) = Y(\mu_2)$.

For every $\varepsilon > 0$ and every $\mu \in \mathcal{M}$, the set $\mathrm{R}_\varepsilon(\mu) = \mathrm{R}(\varepsilon, \mu)$ is the set of all DM's mixed actions that he perceives to be ε-optimal when the signal he observes is distributed according to μ. In applications we usually have $\mathrm{R}_{\varepsilon_1}(\mu) \subseteq \mathrm{R}_{\varepsilon_2}(\mu)$ whenever $\varepsilon_1 < \varepsilon_2$. We will impose a stronger condition below (see Definition 6). The requirement that $\mathrm{R}_\varepsilon(\mu_1) = \mathrm{R}_\varepsilon(\mu_2)$ whenever $Y(\mu_1) = Y(\mu_2)$ amounts to saying that the response function depends only on the footprint of nature's strategy.

Example 4 (Matching Pennies) Consider a decision problem with two states, L and R, and two pure actions T and B. The payoffs are given by:

	L	R
T	*1*	*0*
B	*0*	*1*

Assume that the DM has no information about nature's behavior. From the DM's point of view, the set of nature's possible strategies consists of the set Y of all the distributions over $\{L, R\}$. Assume furthermore that DM wishes to guarantee the best payoff possible when nature acts in an adversarial manner. The mixed action $[\frac{1}{2}(T), \frac{1}{2}(B)]$ is then the unique best response to Y, and[6] $\mathrm{R}^*_\varepsilon = [\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon]$, *for every* $\varepsilon \in [0, \frac{1}{2}]$, *and* $\mathrm{R}^*_\varepsilon = \mathrm{R}^*_{\frac{1}{2}}$ *for every* $\varepsilon \in (\frac{1}{2}, 1]$.

It is worth emphasizing that when DM wishes to maximize the worst case payoff his best response is often mixed. ■

We will prove a regret-free result that applies to response functions that satisfy two properties.

Definition 5 *The response function* R *is* robust *if for every* $\mu \in \mathcal{M}$ *and for every* $\varepsilon' \in (0, 1)$, $\mathrm{R}_{\varepsilon'}(\mu)$ *contains infinitely many distinct elements in* X *that have a full support.*

Suppose that $\mathrm{R}_0(\mu)$ represents the set of all mixed actions of player 1 that are optimal in some sense w.r.t. μ. A sufficient condition for robustness, is that any ε-perturbation of a strategy in $\mathrm{R}_0(\mu)$ is $K\varepsilon$-optimal, for some fixed $K > 0$, and thus in $\mathrm{R}_{K\varepsilon}(\mu)$.

Definition 6 *The response function* R *is* uniformly continuous *if for every* $\varepsilon > 0$ *there is* $\delta > 0$ *such that* $\|\mu_1 - \mu_2\|_\infty \leq \delta$ *implies* $\mathrm{R}_\eta(\mu_2) \subseteq \mathrm{R}_{\eta+\varepsilon}(\mu_1)$, *for every* $\eta \in [0, 1 - \varepsilon]$.

[6] Since DM receives no information, the parameter μ in the definition of R^* is superfluous.

In words, R is uniformly continuous if every mixed action that is η-optimal when the signals are μ_2 is also $(\eta + \varepsilon)$-optimal when the signals are μ_1, provided μ_1 and μ_2 are sufficiently close. Since the domain of a response function, $[0, 1] \times \mathcal{M}$, is a compact set, uniform continuity of R is equivalent to continuity.

Definition 7 *The response function* R *is* polynomially uniform continuous *if there is a polynomial φ which is positive in a neighborhood of* 0*, such that for every $\varepsilon > 0$, $\|\mu_1 - \mu_2\|_\infty \leq \varphi(\varepsilon)$ implies $\mathrm{R}_\eta(\mu_2) \subseteq \mathrm{R}_{\eta+\varepsilon}(\mu_1)$, for every $\eta \in [0, 1 - \varepsilon]$.*

Observe that if R is polynomial uniform continuous then we can assume w.l.o.g. that $\varphi(\varepsilon) = \varepsilon^r$ for some natural number r.

Example 5 (Worst case) As in Example 3, suppose DM wishes to maximize the worst possible payoff given his information. For every $\mu \in \Delta(S)$ denote

$$v(\mu) := \max_{x \in X} \min_{y \in Y(\mu)} u(x, y). \tag{2}$$

This is the highest payoff that DM can guarantee when his information is described by μ. Let $\mathrm{R}^*_\varepsilon(\mu)$ be the set of all mixed actions x such that $\min_{y \in Y(\mu)} u(x, y) \geq v(\mu) - \varepsilon$. This response function is robust. By Rosenberg et al. (2004, Lemma 4) the set-valued function $\mu \to Y(\mu)$ is continuous, and therefore also uniformly continuous. *Since the value operator is non-expansive, this response function is polynomially uniform continuous.* ■

Example 6 (Maximum entropy) For every $\mu \in \mathcal{M}$ let $Y^{\mathrm{ent}}(\mu) \in Y(\mu)$ be nature's mixed action that is consistent with μ and maximizes the entropy, that is

$$Y^{\mathrm{ent}}(\mu) \ := \ \mathrm{argmax}_{y \in Y(\mu)} \Big(- \sum_{b \in B} y_b \ln(y_b) \Big).$$

Let $\mathrm{R}_\varepsilon(\mu)$ be the set of all ε-best response strategies to $Y^{\mathrm{ent}}(\mu)$. Then, R is robust, and as in Example 5 it is uniformly continuous. ■

3.3 Conditional regret-free strategies

As exhibited by Example 5, an optimal strategy of DM often involves choosing a mixed action, we allow DM to choose a mixed action at every stage, and to condition his play on the mixed action he chose in past stages.

Definition 8 *A* (behavior) strategy *of DM is a function $\sigma : H \to \Delta(X)$, where $H = \bigcup_{t=1}^{\infty}(X \times S)^{t-1}$ is the space of all finite histories.*

Since DM only observes the signals, for every strictly mixed action x he plays, he can calculate at every stage t the empirical frequency of the past signals he received up-to that stage, whenever he was mixing according to x. Let $\nu_x^t \in \Delta(S)$ be the empirical distribution of signals up-to stage t over all past stages where DM mixed according to x. The distribution ν_x^t need not be in $\mathcal{M}$. However, by the strong law of large numbers, with probability 1 ν_x^t gets closer and closer to $\mathcal{M}$ as the number of times x is played increases. Denote by $\mu_x^t \in \mathcal{M}$ the closest point in $\mathcal{M}$ to ν_x^t. If $x \in \mathrm{R}_\varepsilon(\mu_x^t)$, then x is ε-optimal against μ_x^t, implying that DM does not have a severe regret for playing x instead of another mixed action.

Since the space of mixed actions is infinite, DM may play each mixed action only finitely many times, in which case the regret for playing a certain mixed action cannot diminish. We therefore partition the set of mixed actions into finitely many small sets, and lump all mixed actions in the same set. For every subset $X' \subseteq X$, let $\nu_{X'}^t \in \Delta(S)$ be the empirical frequency of signals over the stages where DM mixed according to a mixed action in X'. Let $\mu_{X'}^t$ be the closest footprint in $\mathcal{M}$ to $\nu_{X'}^t$, and let $I_{X'}^t$ be the set of stages up to stage t in which DM played a mixed action in X'.

Definition 9 *A strategy σ is* ε-conditional regret-free w.r.t. the response function R *if there is $\delta > 0$ such that for every partition of X into sets with diameter smaller than δ, and for every nature's strategy τ,*

$$\lim_{T\to\infty} \mathbb{P}_{\sigma,\tau}\left(\frac{1}{t}\#\left\{k \le t\colon X(k) \subseteq \mathrm{R}_\varepsilon(\mu_{X(k)}^t)\right\} \ge 1-\varepsilon \quad \forall t \ge T\right) = 1, \tag{3}$$

where $X(k)$ is the atom in the partition of X that contains x_k, the mixed action played at stage k. A strategy is conditional regret-free w.r.t. the response function R *if it is* ε-conditional regret-free w.r.t. the response function R *for every $\varepsilon > 0$.*

Note that the footprint $\mu_{X(k)}^t$ is (close to) the average signal in those stages where DM played a mixed action in $X(k)$, and $\mathrm{R}_\varepsilon(\mu_{X(k)}^t)$ is the set of all mixed actions that are ε-optimal against it. Thus, a strategy is ε-conditional regret-free w.r.t. the response function R if, as t goes to infinity, the probability that in most of the stages up to stage t the mixed actions that ????

are played are ε-optimal increases to 1.

Example 7 *Consider a decision problem with two states, L and R, and two pure actions T and B. The payoffs (at the center) and the signals (at the upper-right corner) are given by:*

	L	R
T	$1^{\,a}$	$3^{\,a}$
B	$2^{\,b}$	$0^{\,c}$

Observe that the signal reveals the action of DM. The stationary strategy T is 0-conditional regret-free w.r.t. the response function R *of maximum entropy that was defined in Example 6. Indeed, when DM plays repeatedly T he receives no information about nature's behavior, the empirical distribution over signals is $[1(a)]$. Then $Y([1(a)]) = Y$, and the stationary strategy y^* that maximizes the entropy is $[\frac{1}{2}(L), \frac{1}{2}(R)]$. The best response of player 1 against y^* is indeed T, so that the stationary strategy T is 0-conditional regret-free w.r.t.* R. *Even though playing B with low distribution may improve DM's performance, as he will obtain additional information on nature's behavior, this pure strategy is 0-conditional regret-free. In our construction of ε-conditional regret-free strategies DM will always play each action with a probability which is bounded away from 0, so that the signals will be as informative as possible.*

The stationary strategy T is not 0-conditional regret-free w.r.t. the response function R^* *of worst case that was defined in Example 5. Indeed, when $Y([1(a)]) = Y$ the mixed action of DM that maximizes the worst case scenario is $[\frac{1}{2}(T), \frac{1}{2}(B)]$.*

We now explain why it is necessary to require that the response function is uniformly continuous to obtain regret-free strategies. Denote by $y^t_{X(k)}$ the average mixed strategy of nature over the stages where DM played a mixed action in $X(k)$. When the monitoring is imperfect, $y^t_{X(k)}$ is not known to DM. From the strong law of large numbers, if DM plays infinitely often mixed actions in $X(k)$, then the difference between the theoretical footprint $s(y^t_{X(k)})$ and the empirical footprint $\nu^t_{X(k)}$ goes to 0 as t increases, and therefore the difference between $s(y^t_{X(k)})$ and $\mu^t_{X(k)}$ goes to 0 as well. If R is uniformly continuous, and if $X(k) \subseteq \mathrm{R}_\varepsilon(\mu^t_{X(k)})$, then $X(k) \subseteq \mathrm{R}_{\varepsilon+\eta}(\nu^t_{X(k)})$, so that $X(k) \subseteq \mathrm{R}_{\varepsilon+\eta}(s(y^t_{X(k)}))$ for every $\eta > 0$, provided $t = t(\eta)$ is large enough. Therefore, the mixed actions played by DM are $\varepsilon + \eta$-optimal against the true play of nature, and not only against the observed signals. Without the requirement that the response function is uniformly continuous one cannot relate the optimal response against the observed footprint to the optimal response against the actual behavior of nature, or even to the optimal response against the footprint which is closest to the observed signals.

Our first result is that an ε-conditional regret-free strategy always exists.

Theorem 1 *For every robust and uniformly continuous response function* R*, and for every $\varepsilon > 0$, there exists an ε-conditional regret-free strategy w.r.t.* R*.*

By playing in blocks of increasing size, and executing an ε_k-conditional regret-free strategy in block k, where $(\varepsilon_k)_{k \in \mathbf{N}}$ is a sequence that decreases to 0, one obtains the existence of a conditional regret-free strategy.

Theorem 2 *For every robust and polynomially uniform continuous response function* R *there exists an conditional regret-free strategy w.r.t.* R*.*

We now explain why it is necessary to require that the response function is polynomially uniform continuous, and it is not enough to require uniform continuity. The strategy that we construct plays in blocks; the length of block k is T_k and within that block the player plays an ε_k-conditional regret-free strategy. The length T_k guarantees that the probability of the event defined in Eq. (3) is at least $1 - \varepsilon_k$. It implies that in block k, the probability that DM will have a regret is ε_k.

In order for this strategy to be regret-free, it needs to have two properties. The first is that $\sum_{k=1}^{\infty} \varepsilon_k$ should be finite, which by the Borel-Cantelli Lemma, implies that DM will have a regret only in finitely many blocks. In particular, the regret of DM at the end of the blocks goes to 0 as the game evolves. The second property is that the sequence $(T_k)_{k \in \mathbb{N}}$ increases sufficiently slowly, so that the regret at the **end** of the blocks being small guarantees that the regret in **all** stages is small. As ε_k decreases, T_k needs to increase, but in order to satisfy the second property, it needs to grow relatively slowly, which can be done when the response function is polynomially uniform continuous.

It is important to note that an ε-conditional regret-free strategy w.r.t. R (or a conditional regret-free strategy w.r.t. R) does not guarantee that the unobserved long-run average payoff is high. The next definition of no-regret takes care of this issue.

A partition $X_1, X_2, \ldots, X_L$ of the set of mixed actions X is *support preserving* if $\mathrm{supp}(x) = \mathrm{supp}(x')$ for every X_l and every $x, x' \in X_l$. Recall that as long as DM plays a given set of actions with positive probability, the set of distributions among which he can distinguish remains the same: if $x, x' \in X_l$ then $Y(s(x, y)) = Y(s(x', l))$.

Definition 10 *A strategy σ of DM is ε-*conditional regret-free *if there is $T \in \mathbf{N}$ and $\delta > 0$ such that for every support preserving partition of X into sets with diameter smaller than δ, and for every nature's strategy τ, one has*

$$\lim_{T\to\infty} \mathbb{P}_{\sigma,\tau}\left(\frac{1}{t}\#\left\{k \le t \colon \frac{1}{|I^t_{X(k)}|}\sum_{j\in I^t_{X(k)}} u(a_j,b_j) \ge v_k - \varepsilon\right\} > 1-\varepsilon \quad \forall t \ge T\right) = 1,$$

where v_k is the maxmin level when nature is restricted to play actions in $Y(x_k, s(y^j_{X(k)}))$ and DM is unrestricted. The strategy is conditional regret-free *if it is ε-conditional regret-free for every $\varepsilon > 0$.*

Since player 1 does not observe his payoffs, and does not even observe the realized distribution of the actions of player 2, he cannot calculate his long-run average payoff. Nevertheless, playing an ε-conditional regret-free strategy guarantees him (up to an ε) a payoff of at least $v(s(y^t_{X(k)}))$, whenever he played a mixed action in $X(k)$. This safety level is the maxmin payoff when nature plays an action that is informationally equivalent to what it really played.

Since the response function R* introduced in Example 5 is polynomially uniform continuous, applying Theorem 2 to this response function implies that a conditional regret-free strategy exists. Formally,

Theorem 3 *DM has a conditional regret-free strategy.*

3.4 Back to n-player games

Suppose that the n-player game $G = (N, \{A_i\}_{i\in N}, \{u_i\}_{i\in N})$ is repeatedly played. At every stage each player $i \in N$ chooses an action $a^t_i \in A_i$, and receives a signal $s^t = s(a^t)$, where $a = (a^t_i)_{i\in N}$ is the joint action played at stage t, and $s_i : A \to \Delta(S)$ is a signalling function. If each player regards the other $n-1$ players as "nature", we reduce the game into n sequential decision problems.

Theorem 4 *In an n-player game with signalling functions, if every player employs an ε-conditional regret-free strategy, then with probability* 1 *the empirical distribution of joint actions converges to the set of* $2n\varepsilon$-PSCE .

The goal of this paper is the following straightforward implication of Theorem 3 and Theorem 4:

Theorem 5 *In every game with signalling functions, there is a decentralized process which induces with probability* 1 *an empirical distribution of joint actions that converges to the set of* ε-PSCE .

Remark 1 *The empirical distribution of joint actions in the procedure we describe below not only converge to the set of* ε-PSCE . *It has the additional feature that the information the players have about others are not correlated. In other words, only the strategies the players play are correlated, while the information each player has does not depend on the strategy he is recommended to play.*

4 Proofs

4.1 Random vector-payoffs games – a background

Blackwell's approachability theory (Blackwell, 1956) is a useful tool in the study of regret-free strategies. Luce and Raifa (1958) cite the Blackwell's[7] proof that uses his own approachability theory of Hannan's (1957) no-regret theorem. To make the presentation complete, we briefly review the definitions and results that we need below.

A two-player game with vector payoffs is given by an $n \times m$ matrix, whose entries are distributions over $\mathbb{R}^d$, such that for every $i = 1, \ldots, n$ and every $j = 1, \ldots, m$, the distribution $W_{i,j}$ that corresponds to the entry (i, j) has mean $w_{i,j} \in \mathbb{R}^d$. At every stage t the two players, independently and simultaneously, choose actions $i^t \in \{1, \ldots, n\}$ and $j^t \in \{1, \ldots, m\}$, and player 1 obtains an $\mathbb{R}^d$-dimensional payoff w^t that is chosen according to the distribution W_{i^t,j^t}. Player 1 is not informed of the action j^t that player 2 chose, but only of the realization w^t of W_{i^t,j^t}.

Denote by $\overline{w}^t := \frac{1}{k} \sum_{k=1}^{t} w^k$ the average vector payoff up to stage t. A set $C \subseteq \mathbb{R}^d$ is *approachable* by player 1 if there is a strategy σ of player 1 that guarantees that $\overline{w}^t$ gets closer and closer to C: for every $\varepsilon > 0$ there is $T \in \mathbb{N}$ such that the following holds

[7] Luce and Raifa (1958) refer to Blackwell's invited address to the Institute of Mathematical Statistics, Seattle, August 1956, entitled "Controlled random walks".

inequality for every strategy τ of player 2.[8]

$$\lim_{T\to\infty} \mathbb{P}_{\sigma,\tau}\left(d(\overline{w}^t, C) \leq \varepsilon, \quad \forall t \geq T\right) = 1.$$

We say that the strategy σ *approaches* C.

For every mixed action y of player 2, that is, a probability distribution over $\{1,\ldots,m\}$, define

$$R_2(y) := \left\{\sum_{i=1}^{m} x_i y_j w_{i,j} \colon \sum_{i=1}^{m} x_i = 1, \quad x_i \geq 0 \quad \forall i\right\}.$$

When player 2 plays the mixed action y, it is guaranteed that the average of the means will be in $R_2(y)$, whatever player 1 plays.

Blackwell (1956) provided a geometric condition on the set C that guarantees that it is approachable. The strategy σ that Blackwell constructed to approach such a set C does it in a uniform rate, which is independent of the strategy employed by player 2:[9]

$$\mathbb{P}_{\sigma,\tau}\left(d(\overline{w}^t, C) \leq \frac{11}{t^{1/3}}, \quad \forall t \geq T\right) \geq 1 - \frac{2}{\sqrt{T}}, \quad \forall T, \forall \tau. \tag{4}$$

Finally, Blackwell, (1956, Theorem 3) proved that a closed and convex set C is approachable if and only if $R_2(y) \cap C \neq \emptyset$ for every y; that is, if for every mixed action of player 2, there is a mixed action x of player 1 such that the average mean w.r.t. (x, y) lies in C.

4.2 The proof of Theorem 1.

The outline of the proof is as follows. After some definitions (step 1) we define an auxiliary game with random vector-payoffs (step 2). Unlike in most of the literature in this area, in our game, as in Blackwell's (1956) seminal work, payoffs are random, and the player only observes the realized payoff, and not the action of the opponent. We study some properties of the average payoff in this game (step 3), define a target set C and prove that it is approachable (step 4). Finally, we prove that every strategy that approaches C is ε-conditionally regret-free (step 5). The target set that we define is close in spirit to the one defined by Blackwell in his alternative proof of Hannan's (1957) result.

Step 1: Preparations.

[8] The metric we use throughout is that induced by the sup-norm.

[9] See Maschler, Solan and Zamir (2007) for this derivation.

Fix $\varepsilon \in (0, \frac{1}{2})$, and let $\delta \in (0, \varepsilon)$ be given by the definition of uniform continuity: if $\|\mu - s(y)\| < 4\delta$ then $\mathrm{R}_\eta(\mu) \subseteq \mathrm{R}_{\eta+\varepsilon}(s(y))$ for every $\eta \in [0, 1-\varepsilon]$.

Let $\{y_1, y_2, \ldots, y_L\}$ be a finite δ-grid of Y: for every $y \in Y$ there is $l \in \{1, 2, \ldots, L\}$ such that $\|y - y_l\| \leq \delta$. For every l define the closed ball around y_l with radius δ:

$$Y_l = \{y \in Y \colon \|y - y_l\| \leq \delta\}.$$

Then each Y_l is convex and closed, and $\bigcup_{l=1}^L Y_l = Y$. Moreover, $\|s(x,y) - s(x,y_l)\| \leq \delta$ for every l, every $y \in Y_l$, and every $x \in X$. Let $x^* \in X$ be a mixed action with full support. For every $l \in \{1, \ldots, L\}$ choose $x_l \in \mathrm{R}_\varepsilon(s(x^*, y_l^*))$, such that all $(x_l)_{l=1}^L$ are distinct and have full support (i.e. $\mathrm{supp}(x_l) = A$). Since R is robust this is possible. Since both x^* and x_l have full support, $x_l \in \mathrm{R}_\varepsilon(s(x_l, y_l^*))$. Define $X_* := \{x_l : 1 \leq l \leq L\}$. Observe that the number L of mixed actions in X_* depends on δ, and therefore on ε.

Step 2: Defining an auxiliary game with random vector-payoffs.
Define an auxiliary two-player game with random vector-payoffs as follows. The action set of player 2 is $B = \Omega$, and the action set of player 1 is X_*. The payoff is $L|S|$-dimensional, and it serves as a means to keep track of the average signal. Observe that the dimension depends on δ, and therefore on ε. For every x_l and every $b \in B$, the random payoff $W(x_l, b) = (W_{l',s'}(x_l, b))_{l'=1,\ldots,L}^{s' \in S} \in \mathbb{R}^{L|S|}$ is defined as follows:

- The vector $W(x_l, b)$ is a unit vector: all coordinates are equal to 0, except of one that is equal to 1.
- For every $l' \neq l$ and every $s' \in S$, $W_{l',s'}(x_l, b) = 0$.
- For every $s' \in S$, the probability that $W_{l,s'}(x_l, b) = 1$ is equal to $s(x_l, b)(s')$.

We therefore obtain that the expectation of $W(x_l, b)$, $\mathbb{E}(W(x_l, b))$, is 0 for all coordinates (l', s') such that $l' \neq l$, and is equal to $s(x, b)(s')$ in all coordinates (l', s') when $l' = l$.

Step 3: Properties of the average payoff in the game with payoff vectors.

Since the stage payoff is a unit vector, the average payoff vector up to stage t, $\overline{w}^t$, is an element of the unit simplex of $\mathbb{R}^{L|S|}$. Denote by y^t the mixed action played by player 2 at stage t of the auxiliary game, and by I_l^t the set of all stages up to stage t in which player 1 played x_l. Denote by $\overline{y}_l^t$ the average play up to stage t in all stages in I_l^t:

$$\overline{y}_l^t = \frac{1}{|I_l^t|} \sum_{t' \in I_l^t} y^{t'}.$$

By the strong law of large numbers, the distance $d(\overline{w}_l^t, s(x_l, \overline{y}_l^t))$ goes to 0, provided that $|I_l^t|$ goes to infinity, and by the Azuma inequality, the convergence is uniform over the strategies of player 2. Formally, for every $t \in \mathbb{N}$, for every strategy τ of player 2, for every l,

$$\mathbb{P}_{\sigma,\tau}\left(d(\overline{w}_l^t, s(x_l, \overline{y}_l^t)) \leq \delta\right) \geq 1 - 2\delta \exp\left(-\frac{\delta^2 |I_l^t|}{2}\right). \quad (5)$$

Let L_t be the set of all l's such that $\frac{|I_l^t|}{t} \geq \frac{\delta}{|L|}$. Thus,

$$\sum_{l \in L_t} \frac{|I_l^t|}{t} = 1 - \sum_{l \notin l_t} \frac{|I_l^t|}{t} \geq 1 - |L| \frac{\delta}{|L|} = 1 - \delta.$$

Furthermore, by Eq. (5), for every $l \in L_t$,

$$\mathbb{P}_{\sigma,\tau}\left(d(\overline{w}_l^t, s(x_l, \overline{y}_l^t)) \leq \delta\right) \geq 1 - 2\exp\left(-\frac{\delta^3 t}{2|L|}\right). \quad (6)$$

This implies that

$$\mathbb{P}_{\sigma,\tau}\left(d(\overline{w}_l^t, s(x_l, \overline{y}_l^t)) \leq \delta,\ \forall l \in L_t\right) \geq 1 - 2|L| \exp\left(-\frac{\delta^3 t}{2|L|}\right). \quad (7)$$

Since $\sum_{t=T}^{\infty} \exp\left(-\frac{\delta^3 t}{2|L|}\right) \leq \int_{T-1}^{\infty} \exp\left(-\frac{\delta^3 t}{2|L|}\right) \mathrm{d}t = \frac{2L}{\delta^3} \exp\left(-\frac{\delta^3 (T-1)}{2|L|}\right)$, for every $T \in \mathbb{N}$,

$$\mathbb{P}_{\sigma,\tau}\left(\sum_{l \in L_t} \frac{|I_l^t|}{t} \geq 1 - \delta \text{ and } d(\overline{w}_l^t, s(x_l, \overline{y}_l^t)) \leq \delta,\ \forall t \geq T, \forall l \in L_t\right) \geq 1 - \frac{4|L|^2}{\delta^3} \exp\left(-\frac{\delta^3 (T-1)}{2|L|}\right). \quad (8)$$

Step 4: Defining a set C, and showing that it is approachable by player 1.
Since $\bigcup_{l=1}^{L} Y_l = Y$, there is at least one l such that $s(y) \in Y_l$. For every such index l the average signal $s(x_l, y)$ is a probability distribution over S. Denote by $\overrightarrow{0}$ the origin of $\mathbb{R}^{|S|}$, and define the following vector $\psi_l(s(x_l, y)) \in \mathbb{R}^{L|S|}$:

$$\psi_l(s(x_l, y)) := (\overrightarrow{0}, ..., \overrightarrow{0}, s(x_l, y), \overrightarrow{0}, ..., \overrightarrow{0}),$$

where $s(x_l, y)$ sits on the coordinates $(l, \cdot)$. Denote $D_l = \{s(x_l, y) \colon y \in Y_l\}$ and $C_l = \{\psi_l(s(x_l, y)) \colon s(x_l, y) \in D_l\}$. Finally let

$$C := \operatorname{conv}\left(\bigcup_l C_l\right).$$

Since for every $x \in X$ the function $y \mapsto s(x,y)$ is linear, and since each Y_l is convex and closed, each C_l is a convex and closed set, so that C is closed. Moreover, a member of C is a convex combination of elements in the sets C_l. That is, every $c \in C$ has a unique representation, $c = \sum_{l=1}^{L} \alpha_l \psi_l(s(x_l, z_l))$, where $s(z_l) \in M_l \subseteq B(s(y_l^*), \delta)$, $\alpha_l \geq 0$, and $\sum_{l=1}^{L} \alpha_l = 1$.

We now verify that C is an approachable set. By Blackwell (1956, Theorem 3), it is sufficient to prove that for every $y \in Y$, the intersection $R_2(y) \cap C$ is not empty. However, for every $y \in Y$, there is l such that $y \in Y_l$. Thus, $\psi_l(s(x_l, y)) \in R_2(y) \cap C_l \subseteq R_2(y) \cap C$.

Denote by σ a strategy that approaches C in the auxiliary game with vector payoffs. The strategy σ can also be interpreted as a strategy in the original repeated game, as the only data that it uses is the sequence of past mixed actions in X_* that player 1 played, and the sequence of past signals that he received.

Step 5: σ is 2ε-conditional regret-free strategy w.r.t. R.

Since σ approaches C, from Eq. (4) we obtain that for every strategy τ of player 2,

$$\mathbb{P}_{\sigma,\tau}\left(d(\overline{w}^t, C) \leq \frac{\delta^2}{|L|}, \quad \forall t \geq T \right) \geq 1 - \frac{2}{\sqrt{T}}, \quad \forall T \geq \frac{(11L)^{1/3}}{\delta^{2/3}}.$$

Recall that $\overline{w}^t$ is an $L|S|$-dimensional vector such that the sum of its $(l,\cdot)$ coordinates is $\frac{|I_l^t|}{t}$. Thus, $d(\overline{w}^t, C) = \max_l\{\frac{|I_l^t|}{t} d(\overline{w}_l^t, C_l)\}$. If $d(\overline{w}^t, C) \leq \frac{\delta^2}{|L|}$, then $d(\overline{w}_l^t, C_l) \leq \delta$ for every $l \in L_t$ (see step 2 above). Together with Eq. (8) we obtain for every strategy τ of player 2, and every $T \geq \frac{(11L)^{1/3}}{\delta^{2/3}}$,

$$\mathbb{P}_{\sigma,\tau}\left(\forall t \geq T, \sum_{l \in L_t} \frac{|I_l^t|}{t} \geq 1 - \delta \text{ and } \quad \forall l \in L_t,\ d(s(x_l, \overline{y}_l^t), C_l) \leq 2\delta \right) \geq$$
$$1 - \frac{2}{\sqrt{T}} - \frac{4|L|^2}{\delta^3} \exp\left(-\frac{\delta^3(T-1)}{2|L|} \right).$$

If $d(s(x_l, \overline{y}_l^t), C_l) \leq 2\delta$, then $d(s(x_l, \overline{y}_l^t), x_l, s(y_l^*)) \leq 4\delta$. Since R is uniformly continuous, and since $x_l \in \mathrm{R}_\varepsilon(s(x_l, \overline{y}_l^*))$, we obtain $x_l \in \mathrm{R}_{2\varepsilon}(s(x_l, y_l^t))$ (recall the choice of δ at the beginning of step 1. Therefore,

$$\mathbb{P}_{\sigma,\tau}\left(\forall t \geq T, \sum_{l \in L_t} \frac{|I_l^t|}{t} \geq 1 - \delta \text{ and } \forall l \in L_t, x_l \in \mathrm{R}_{2\varepsilon}(s(\overline{y}_l^t)) \right) \geq 1 - \frac{2}{\sqrt{T}} - \frac{4|L|^2}{\delta^3} \exp\left(-\frac{\delta^3(T-1)}{2|L|} \right). \tag{9}$$

Since $\delta < \varepsilon$, σ is indeed 2ε-conditional regret-free. ■

4.3 The proof of Theorem 2.

As the argument is standard we provide only a sketch of the proof. We construct a strategy played in blocks. Set $\varepsilon_k := \frac{1}{k}$, and let σ_k be the ε_k-conditional regret-free strategy that was constructed in the proof of Theorem 1. Since R is polynomially uniform continuous, there is a polynomial $\varphi(\varepsilon) = \varepsilon^r$ such that Definition 7 is satisfied. Set $\delta_k := \varepsilon^r = k^{-r}$. δ_k takes the role of δ used in the proof of Theorem 1.

As mentioned in Footnote **??**, we can assume that the number L_k of mixed actions used in σ_k satisfies $|L_k| \leq (\frac{2+\delta_k}{\delta_k})^{|S|} \leq 3^{|S|}k^{r|S|}$. The length of block k is set to be $T_k = k^{3(r|S|+4r)}$. From Eq. (9) we obtain that the probability that for at least one $l \in L_{T_k}$, $x_l \notin \mathrm{R}_{2k^{-1}}(s(\overline{y}_l^{T_k}))$ at the end of block k is at most $\frac{12 \cdot 3^{|S|}k^{r|S|}k^{2r}}{k^{r|S|+4r}} + 4 \cdot 2^{2|S|}k^{2r|S|}k^{2r}\exp(-\frac{k^{3(r|S|+4r)}}{2 \cdot 3^{|S|}k^{r|S|} \cdot k^{3r}}) \leq b_1 k^{-2} + \exp(-b_2 k)$, where b_1 and b_2 are positive constants.

Summing up these terms over the blocks $k \geq k_0$, we obtain that for every $k_0 \in \mathbf{N}$, with probability of at most $b_1 k^{-1} + \frac{\exp(-b_2 k)}{b_2}$, there exists $k \geq k_0$ and $l \in L_{T_k}$ such that at the end of block k, $x_l \notin \mathrm{R}_{2k^{-1}}(s(\overline{y}_l^{T_k}))$. This shows that the probability to have a regret at the end of at least one block shrinks to zero.

To conclude the proof we should consider stages at the middle of the blocks. However, since the weight of a block relative to its history becomes negligible as k increases (because $\frac{T_k}{T_1+T_2+\cdots+T_{(k-1)}}$ goes to 0 as k grows to infinity), the regret at the middle of a block is predominantly affected by the regret at the end of the previous block. This implies that the regret goes to zero as time goes by, which shows that the strategy constructed is ε-conditional regret-free for every $\varepsilon > 0$. ■

4.4 The proof of Theorem 3.

We apply Theorem 1 to the response function R^* defined in Example 5. Thus, $\mathrm{R}^*_\varepsilon(\mu)$ is the set of ε-optimal mixed actions of player 1 in the game $G(\mu)$ in which the sets of mixed actions of the two players are X and $Y(\mu)$ respectively, and the payoff function is u. Denote by $v(\mu)$ the value of $G(\mu)$. Then

$$u(x,y) \geq v(\mu) - \varepsilon, \quad \forall x \in \mathrm{R}^*_\varepsilon(\mu), y \in Y(\mu). \tag{10}$$

As explained in Example 5, R^* is polynomially uniformly continuous. Therefore, for every $\eta > 0$ there is $\delta > 0$ such that whenever $\|\mu_1 - \mu_2\| \leq \delta$ we have

$$u(x,y) \geq v(\mu_1) - \varepsilon - \eta, \quad \forall x \in \mathrm{R}^*_\varepsilon(\mu_1), y \in Y(\mu_2). \tag{11}$$

It now follows from Definitions 9 and 10 that every strategy that is ε-conditional regret-free w.r.t. R^* is ε-conditional regret-free. The result follows from Theorem 2, when applied to R. ■

4.5 The proof of Theorem 4.

Suppose that each player employs an ε-conditional regret-free. Denote by σ_i player i's strategy. For every stage t denote by Q_t the empirical frequency of the (mixed) joint actions played up to stage t. We will show that with a high probability Q_t is $2n\varepsilon$-PSCE, provided t is sufficiently large.

Fix a player i, and refer to all other players, denoted $-i$, as nature. Specifically, X is the set of player i's actions and Y is that of the joint strategies (mixed) of $-i$. For simplicity denote $\sigma = \sigma_i$, $s = s_i$ and τ the (joint) strategy of players $-i$.

Let $X_1, ..., X_K$ be a partition of X into sets, the diameter of each is sufficiently small so that $x, x' \in X_k$ guarantees $\|s(x,y) - s(x',y)\|_1 < \varepsilon$ for every $y \in Y$. Recall that $I^t_{X_k}$ denotes the set of stages until t that a strategy in X_k has been used. Since σ is ε-conditional regret-free, there is a set $K'_t \subseteq K$ such that the probability that $\sum_{k \in K'_t} \frac{|I^t_{X_k}|}{t} \geq 1 - \varepsilon$ and $\forall k \in K'_t, \forall t \geq T$, $X_k \subseteq \mathrm{R}_\varepsilon(\mu^t_{X_k})$ grows to 1 with T.

Let $p^k_t = \frac{|I^t_k|}{t}$ be the frequency of strategies from X_k up to stage t. Denote by K''_t the set of k's such that $p^k_t \geq \frac{\varepsilon}{K}$. Sets X_k with $k \notin K'_t$ have been used at most ε of the stages until stage t. Finally denote by y^k_t the average strategy used by $-i$ on I^t_k. The strong law of large numbers implies that for every $k \in K''_t$, the asymptotic difference between $s(x, y^k_t)$ and the empirical frequency of signals over $I^t_{X_k}$ is not larger than ε.

Denote $K_t = K'_t \cap K''_t$. We obtain that for every $k \in K_t$, $X_k \subseteq \mathrm{R}_{2\varepsilon}(\mu^t_{X_k}), \forall t \geq T$ with a probability that increases to 1 as T increases to infinity. Moreover, $\sum_{k \in K_t} \frac{|I^t_{X_k}|}{t} \geq 1 - 2\varepsilon$. As in Example 1, the distribution $s(x, y^k_t)$ is equivalent to getting a partially specified probability on the distribution over signals induced by (x, y^k_t).

We obtained that after excluding 2ε of player i's strategies, his strategies are 2ε best response to the worst strategy consistent with the signals. In order to obtain this property for all players one should exclude $2n\varepsilon$ of the distribution, which implies that the probability that Q_t is $2n\varepsilon$-PSCE grows to 1 as t grows to infinity. ■

References

[1] Aumann, R. J. (1974) "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, **1**, 67-96.

[2] Aumann, R. J. (1987) "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, **55**, 1-18.

[3] Blackwell, D. (1956) "An Analog of the Minmax Theorem for Vector Payoffs," *Pacific Journal of Mathematics*, **6**, 1-8.

[4] Blum, A. and Y. Mansour (2007) "From External to Internal Regret," *Journal of Machine Learning Research*, **8**, 1307-1324.

[5] Cesa-Bianchi, N. G. Lugosi, and G. Stoltz (2006) "Regret Minimization under Partial Monitoring," *Mathematics of Operations Research*, **31**, 562-580.

[6] Foster, D. and R. V. Vohra (1997) "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, **21**, 40-55.

[7] Foster, D. and R. V. Vohra (1999) "Regret in the On-line Decision Problem," *Games and Economic Behavior*, **29**, 7-35.

[8] Fudenberg, D. and D. K. Levine (1999) "Conditional Universal Consistency," *Games and Economic Behavior*, **29**, 104-130.

[9] Hannan, J. (1957) "Approximation to Bayes Risk in Repeated Play," *Contributions to the theory of games*, **3**, 97-139.

[10] Hart, S. and A. Mas-Colell (2000) "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica*, **68**, 1127-1150.

[11] Hart, S. and A. Mas-Colell (2001) "A General Class of Adaptive Strategies," *Journal of Economic Theory*, **98**, 26-54.

[12] Lehrer, E. (2005) "Partially Specified Probabilities: Decisions and Games," mimeo.

[13] Luce, D. R. and H. Raiffa (1958) *Games and Decisions*, John Wiley, N.Y.

[14] Lugosi, G., S. Mannor and G. Stoltz (2007) "Strategies for Prediction under Imperfect Monitoring," mimeo.

[15] Mannor, S. and N. Shimkin (2003) "On-line learning with imperfect monitoring," in *Proceedings of the 16th Annual Conference on Learning Theory,* 552-567. Springer.

[16] Maschler M., E. Solan and S. Zamir (2007) *Game Theory*, A textbook in preparation.

[17] Nash, J. (1951) "Non-cooperative games," *The Annals of Mathematics,* 2nd Ser., **54**, 286-295.

[18] Rosenberg, D., E. Solan and N. Vieille (2003) "Stochastic Games with Imperfect Monitoring," *International Journal of Game Theory*, **32**, 133-150.

[19] Rustichini, A. (1999) "Minimizing Regret: the General Case," *Games and Economic Behavior*, **29**, 224-243.

[20] Weissman, T. and N. Merhav. (2001) "Universal Prediction of Binary Individual Sequences in the Presence of Noise," *IEEE Trans. Inform. Theory*, **47**, 2151-2173.

[21] Weissman, T., N. Merhav and A. Somekh-Baruch. (2001) "Twofold Universal Prediction Schemes for Achieving the Finite State Predictability of a Noisy Individual Binary Sequence," *IEEE Trans. Inform. Theory,* **47**, 1849-1866.