Two-Player Repeated Games with Nonobservable Actions and Observable Payoffs
Author(s): Ehud Lehrer
Source: *Mathematics of Operations Research*, Vol. 17, No. 1 (Feb., 1992), pp. 200-224
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/3689901
Accessed: 22/08/2011 07:50

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Mathematics of Operations Research.*

http://www.jstor.org

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 17, No. 1, February 1992
Printed in U.S.A.

TWO-PLAYER REPEATED GAMES WITH NONOBSERVABLE ACTIONS AND OBSERVABLE PAYOFFS*

EHUD LEHRER

This paper studies two-person repeated games in which after each stage a player is informed about the payoff he received at the previous stage. The information can, in some cases, include more than that. Four kinds of Nash-equilibrium concepts are defined by the limit of the means. A characterization of the equilibrium-payoffs sets and several properties of these sets are given. As a specific example, the standard information case, that of the folk theorem, is provided.

1. Introduction. A repeated game with nonobservable actions consists of the repetition of a one-shot game which is repeated sequentially for an infinite number of periods. At the end of each a player gets only partial information about other players' actions which took place at that stage. At each stage a player may rely on the signal he previously got when he chooses his one-stage strategy (action) for the next repetition. The case of standard information, where at the end of each stage a player is informed of the actions taken by all other players at the previous stage, has already been studied (see [A3]). This is known as the folk theorem.

In this paper we refer to the case of a two-player game, where at the end of each stage, a player is informed of his own payoff at that stage, and sometimes of additional information. We define four notions of Nash equilibrium in the undiscounted repeated game: the lower, the upper, the Banach equilibrium, and the uniform equilibrium. The first three notions differ one from the other in the ways they evaluate diverging bounded sequences of numbers. In other words, the different notions involve three ways to measure profitability of a deviation: by the liminf, by the limsup, and by any Banach limit.

The uniform equilibrium (see [S1]), on the other hand, does not deal with evaluating the infinite sequence of partial averages. It requires that for every $\epsilon > 0$ there is a time T such that the strategy in question induces an ϵ-Nash equilibrium in the t-fold repeated game for every $t \geqslant T$. In other words, for *all* possible deviations the partial average does not exceed by more than ϵ the prescribed payoff if the game proceeds for more than T times.

We characterize the sets of the payoffs that are sustained by lower, upper, Banach, and uniform equilibria, and we find that all four sets coincide. The characterization differentiates between the nontrivial case where both players can (by playing a certain action) distinguish between two of their opponent's actions (namely, to get different signals), and the trivial case in which the signals of at least one player always reveal

*Received January 28, 1987; revised July 2, 1990.
AMS 1980 subject classification. Primary: 90D05, 90D20.
IAOR 1973 subject classification. Main: Games.
OR/MS Index 1978 subject classification. Primary: 231 Games.
Key words. Two-person repeated games, Nash equilibrium, payoff sets, folk theorem.

0364-765X/92/1701/0200/$01.25
Copyright © 1992, The Institute of Management Sciences/Operations Research Society of America

nothing about the action played by the other one. In the nontrivial case the characterization is done by the sets C_1 and C_2 consisting of joint mixed actions (p_1, p_2), where p_i is an action of player i. C_1 is the set of all the pairs (p_1, p_2), where, if there is another p'_1, which is a better response versus p_2 than p_1, then one of the following two must hold: either p'_1 changes the distribution of player 2's signals or p'_1 decreases the possibility of player 1 distinguishing between actions of player 2. Both possibilities give player 2 opportunity to detect the deviation from p_1 to p'_1. The set C_2 is defined in a similar way.

The main result of this work states that all the sets of equilibrium payoffs coincide and they are equal to the set of all individually rational payoffs that are associated with convex combinations of pairs in $C_1 \cap C_2$. According to this characterization, payoffs and information have the same importance in the sense that a player would be willing to deviate only if the deviation is nondetectable, gives the player at least the same amount of information (as the prescribed action gives), and gives a greater payoff.

The well-known folk theorem [A3] characterizes the same sets in the case of standard information as the set of all the individually rational and feasible payoffs. In this case every deviation is detectable, and therefore C_i contains all the pairs (p_1, p_2). Thus, there is no additional restriction over feasibility in saying that the payoff is associated with a combination of pairs in $C_1 \cap C_2$.

Our proof is partially based on a former result [L2] that characterizes the lower equilibrium payoffs set in a general game with nonobservable actions. It turns out to be the set of all individually rational payoffs that are associated with two (possibly different) convex combinations. The first one is of pairs in C_1 and the second is of pairs in C_2. Another case in which there is a full description of the equilibrium payoffs set in the undiscounted game is the case of semistandard information in which a player is informed about a set that contains his opponent's action rather than of the action itself (see [L1]).

The literature dealing with repeated games with nonobservable actions, especially with complete information, has rapidly expanded over the last decade. However, one of the nicest papers on this topic was written before the new wave by Kohlberg [Ko]. In this paper, to which we refer in more detail in the last section, Kohlberg dealt with zero-sum repeated games with incomplete information and general information functions for both players. Radner [R1] and Rubinstein and Yaari [RY] also studied undiscounted repeated games with one-sided moral hazard. Two players, the principal and the agent, participate in the game. The agent is fully informed and the principal is informed only partially of the agent's actions and on the outcome which is stochastically dependent on the agent's actions. Fudenberg and Maskin [FM] and Abreu, Pearce and Stachetti [APS] have studied discounted games in which all the players are informed only of a common signal which is dependent stochastically on the joint action taken by the players. This signal and the action of a player determine his payoff. Thus, payoffs are observable. Fudenberg and Levine [FL], also in a model of observable payoffs and n players, defined a set of mutually punishable and enforceable payoffs (without nondetectable profitable deviations). They proved that these payoffs are sustained by uniform equilibria. Our result shows that this set of payoffs does not exhaust all the equilibrium payoffs.

This paper is divided into six sections. The theorem is given in the third section, accompanied by a few examples. In the fourth section, several important geometric properties of the payoffs in the case of observable payoffs are proved. The proof of the main theorem, which relies on these properties, is provided in the fifth section. The sixth section is devoted to concluding remarks.

2. Definitions and notations.

Definition 2.1. A two-person repeated game G^* with nonobservable actions is defined by:

a. Finite sets Σ_1, Σ_2, called *action sets*.

b. Functions l_1, l_2; l_i: $\Sigma_1 \times \Sigma_2 \to L_i$, $i = 1,2$. l_i is called the *information function* and L_i is called the *signals set* of player i, $i = 1,2$. l_1 and l_2 satisfy:

(i) $l_1(s,t) \neq l_1(s',t')$ when $s \neq s'$ for all $t, t' \in \Sigma_2$.

(ii) $l_2(s,t) \neq l_2(s',t')$ when $t \neq t'$ for all $s, s' \in \Sigma_1$.

c. Functions h_1, h_2; h_i: $\Sigma_1 \times \Sigma_2 \to \mathbb{R}$, $i = 1,2$, called *payoff functions*.

Notation 2.2. Denote the range of h_i by A_i, $i = 1,2$.

The sets of pure strategies of a player in the repeated game, denoted by F_i, are defined as follows.

Definition 2.3. $F_i = \{(f_i^1, f_i^2, f_i^3, \dots);$ for each $n \in \mathbb{N}$, f_i^n: $L_i^{n-1} \to \Sigma_i\}$ for $i = 1,2$, where L_i^0 is a singleton.

When player i chooses the pure strategy f_i, $i = 1,2$, the game is played as follows. At the first stage, player i plays f_i^1, gets his payoff $h_i(f_1^1, f_2^1)$, and the signal $l_i(f_1^1, f_2^1)$. At the second stage, player i acts $f_i^2(l_i(f_1^1, f_2^1))$, gets his payoff $h_i(f_1^2(l_1(f_1^1, f_2^1)), f_2^2(l_2(f_1^1, f_2^1)))$ and the signal $l_i(f_1^2(l_1(f_1^1, f_2^1)), f_2^2(l_2(f_1^1, f_2^1)))$, and so forth.

A mixed strategy of player i is a probability measure μ_i on F_i.

Notation 2.4. The set of all the mixed strategies of player i is denoted by $\Delta(F_i)$, $i = 1,2$. Each pair of pure strategies $f = (f_1, f_2) \in F_1 \times F_2$ determines a string of signals $(s_1^n(f), s_2^n(f))_{n=1}^\infty \in (L_1 \times L_2)^{\mathbb{N}}$ and a string of payoffs $(x_1^n(f), x_2^n(f))_{n=1}^\infty \in (A_1 \times A_2)^{\mathbb{N}}$, where $s_i^n(f), x_i^n(f)$ are the signal and the payoff, respectively, that player i gets at stage n. If player j picks a pure strategy f_j according to $\mu_j \in \Delta(F_j)$, then s_i^n and x_i^n become random variables.

Definition 2.5. Let $\mu = (\mu_1, \mu_2) \in \Delta(F_1) \times \Delta(F_2)$ and $n \in \mathbb{N}$,

$$H_i^n(\mu_1, \mu_2) = \operatorname*{Exp}_{\mu}\left((1/n) \sum_{k=1}^{n} x_i^k(f)\right), \qquad i = 1,2.$$

$H_i^n(\mu_1, \mu_2)$ is the expectation of the average payoff of player i at the n first stages of the repeated game, when μ_1 is the strategy played by player 1, and μ_2 is that played by player 2.

Definition 2.6. (1) $H_i^*(\mu_1, \mu_2) = \lim_n H_i^n(\mu_1, \mu_2)$ if it exists, $i = 1,2$. $H^*(\mu_1, \mu_2) = (H_1^*(\mu_1, \mu_2), H_2^*(\mu_1, \mu_2))$ if both H_1^* and H_2^* are defined.

(2) Let L be a Banach limit.

$$H_i^{*L}(\mu_1, \mu_2) = L(\{H_i^n(\mu_1, \mu_2)\}_n),$$

$$H^{*L}(\mu_1, \mu_2) = \left(H_1^{*L}(\mu_1, \mu_2), H_2^{*L}(\mu_1, \mu_2)\right).$$

Definition 2.7. (1) μ is an *upper equilibrium* if $H^*(\mu)$ is well defined and if for every $\bar{\mu}_1 \in \Delta(F_1)$, $\limsup_n H_1^n(\bar{\mu}_1, \mu_2) \leqslant H_1^*(\mu_1, \mu_2)$, and for every $\bar{\mu}_2 \in \Delta(F_2)$, $\limsup_n H_2^n(\mu_1, \bar{\mu}_2) \leqslant H_2^*(\mu_1, \mu_2)$.

(2) The lower equilibrium is defined in a similar way, replacing limsup with liminf.

(3) μ is an L equilibrium if, for any $\bar{\mu}_1 \in \Delta(F_1)$, $L(\{H_1^n(\bar{\mu}_1, \mu_2)\}_n) \leqslant H_1^{*L}(\mu)$, and similarly for any $\bar{\mu}_2$, replacing 1 with 2.

(4) μ is a uniform equilibrium if $H^*(\mu)$ is well defined and if there is a sequence of numbers $\{\epsilon_n\}$, converging to zero and an increasing sequence of integers $\{k_n\}$ such that μ induces an ϵ_n-Nash equilibrium in the k_n times repeated game of G.

For an extensive study of uniform equilibria, see [S1].

Notation 2.8. (1) UEP $= \{H^*(\mu) | \mu$ is an upper equilibrium$\}$.
(2) LEP $= (H^*(\mu) | \mu$ is a lower equilibrium$\}$.
(3) $\mathrm{BEP}_L = \{H^{*L}(\mu) | \mu$ is an L equilibrium$\}$.
(4) UNIF $= \{H^*(\mu) | \mu$ is a uniform equilibrium$\}$.

Notation 2.9. If Σ is a set and $s \in \Sigma$, then δ_s will denote the Dirac measure on s, and will be the measure corresponding to s in $\Delta(\Sigma)$, the set of the probability measures over Σ.

Sometimes we will refer to δ_s as s.

REMARK 2.10. The functions $h = (h_1, h_2)$ and $l = (l_1, l_2)$ can be extended to $\Delta(\Sigma_1) \times \Delta(\Sigma_2)$ in a natural way, so that h_i and l_i will be ranged to $\mathbb{R}$ and $\Delta(L_i)$, respectively, $i = 1, 2$.

From here on we will call elements in Σ_i and in $\Delta(\Sigma_i)$ actions and mixed actions, respectively.

3. The main theorem. The main theorem characterizes the various equilibrium payoffs sets in the case where the information includes the payoffs, i.e., for each player i and joint pure actions (s, t), if $h_i(s, t) \neq h_i(s', t')$, then $l_i(s, t) \neq l_i(s', t')$. In simple words: if the payoff related to (s, t) differs from the payoff related to (s', t'), then the signals related to these joint pure actions differ as well. A game of this kind will be called a game with *observable payoffs*.

The characterization is done by an equivalence relation and by a partial order defined on $\Delta(\Sigma_i)$. These relations were defined originally in [L2]. We will give the following definitions for actions of player 1. One can apply similar definitions for player 2.

DEFINITION 3.1. (1) Let $s, s' \in \Sigma_1$. s is *equivalent* to s' $(s \sim s')$ if for every $t \in \Sigma_2$, $l_2(s, t) = l_2(s', t)$.

(2) Let $p, p' \in \Delta(\Sigma_1)$. p is *equivalent* to p' $(p \sim p')$ if for every $t \in \Sigma_2$, $l_2(p', t) = l_2(p, t)$ (in the sense of Remark 2.10).

In words, $p' \sim p$ if the distributions over the signals of player 2 are the same under p as under p', for any action t.

DEFINITION 3.2. (1) Let $s, s' \in \Sigma_1$, s' is *more informative* than s if for every $t, t' \in \Sigma_2$, $l_1(s, t) \neq l_1(s, t')$ implies $l_1(s', t) \neq l_1(s', t')$. s' is *greater* than s $(s' \succ s)$ if $s \sim s'$ and if s' is more informative than s.

(2) Let $p, p' \in \Delta(\Sigma_1)$. p' is *greater* than p $(p' \succ p)$ if $p' \sim p$ and if there are nonnegative constants $\beta_{s', s}$ such that $p'_{s'} = \sum_s \beta_{s', s}$, $p_s = \sum_{s'} \beta_{s', s}$ and if $\beta_{s', s} > 0$, then s' is more informative than s.

In words, p' is greater than p, in the sense of the partial order $\succ$, if $p' \sim p$ and if by playing p' the player can distinguish between two actions of his opponent with a greater probability than he could do so by playing p. The intuitive interpretation of (2) is the following. The prescribed mixed action is to play s with probability p_s. However, according to the deviation action, p', player 1 picks first an s with probability p_s and second, with probability $\beta_{s', s}/p_s$, the action s', which he then plays.

DEFINITION 3.3.
(1) $D_1 = \{(p, q) \in \Delta(\Sigma_1) \times \Delta(\Sigma_2) | h_1(p, q) = \mathrm{Max}_{p' \sim p} h_1(p', q)\}$.
(2) $D_2 = \{(p, q) \in \Delta(\Sigma_1) \times \Delta(\Sigma_2) | h_2(p, q) = \mathrm{Max}_{q' \sim q} h_2(p, q')\}$.
(3) $D = D_1 \cap D_2$.

DEFINITION 3.4.
(1) $C_1 = \{(p, q) \in \Delta(\Sigma_1) \times \Delta(\Sigma_2) | h_1(p, q) = \mathrm{Max}_{p' \succ p} h_1(p', q)\}$.
(2) $C_2 = \{(p, q) \in \Delta(\Sigma_1) \times \Delta(\Sigma_2) | h_2(p, q) = \mathrm{Max}_{q' \succ q} h_2(p, q')\}$.
(3) $C = C_1 \cap C_2$.

I.e., C_i is the set of pairs of the mixed actions, in which player i cannot profit by any deviation without being discovered by player $3-i$, or without decreasing his potential for getting information. Intuitively, if $(p,q) \in C_1$ is played repeatedly many times, then player 1 can profit only by a detectable deviation.

DEFINITION 3.5.

(1) $d_1 = \mathrm{Min}_{q \in \Delta(\Sigma_2)} \mathrm{Max}_{p \in \Delta(\Sigma_1)} h_1(p,q)$, $d_2 = \mathrm{Min}_{p \in \Delta(\Sigma_1)} \mathrm{Max}_{q \in \Delta(\Sigma_2)} h_2(p,q)$.

(2) $IR = \{(a,b) \in \mathbb{R}^2 | a \geqslant d_1 \text{ and } b \geqslant d_2\}$;

i.e., IR is the set of all individually rational payoffs.

(3) Let σ_1 be an action of player 1 that satisfies: $d_2 = \mathrm{Max}_{q \in \Delta(\Sigma_2)} h_2(\sigma_1, q)$. σ_2 is defined in a similar way.

DEFINITION 3.6. The information of player i is *trivial* if for any $s \in \Sigma_i$ and $t, t' \in \Sigma_{3-i}$, $l_i(s,t) = l_i(s,t')$.

In the main theorem we give a characterization of the equilibrium payoffs, using terms of the one-shot game only. It turns out that in case of observable payoffs, all four notions of equilibria yield the same set of payoffs.

MAIN THEOREM 3.7. *If G^* is a two-person repeated game with observable payoffs, then*:

(1) *If the information of both players is not trivial, then*

$$\mathrm{UEP} = \mathrm{LEP} = \mathrm{BEP}_L = \mathrm{UNIF} = \mathrm{Conv}\, h(C) \cap IR$$

$$= \mathrm{Conv}\, h(C_1) \cap \mathrm{Conv}\, h(C_2) \cap IR \quad \textit{for any Banach limit } L.$$

(2) *If the information of at least one player is trivial, then*

$$\mathrm{UEP} = \mathrm{LEP} = \mathrm{BEP}_L = \mathrm{UNIF} = \mathrm{Conv}\, h(D) \cap IR$$

$$= \mathrm{Conv}\, h(D_1) \cap \mathrm{Conv}\, h(D_2) \cap IR \quad \textit{for any Banach limit } L;$$

where for each $E \subseteq \Delta(\Sigma_1) \times \Delta(\Sigma_2)$, $h(E)$ denotes the set of payoffs related to actions in E, i.e., $h(E) = \{h(p,q) | (p,q) \in E\}$.

In order to introduce the examples that follow, let us define the partitions S_1, S_2 of Σ_1 and Σ_2, respectively.

DEFINITION 3.8. If $i \in \{1,2\}$, then S_i is the partition of Σ_i into the equivalence classes of the relation $\sim$. S_1, S_2 are called the *information partitions*.

EXAMPLE 3.9. *Standard Information*. A game with standard information is a game in which each player knows the actions used by his opponent, i.e., $l_i(s,t) \neq l_i(s',t')$ iff $(s,t) \neq (s',t')$ for every $(s,t),(s',t') \in \Sigma_1 \times \Sigma_2$. In this case $S_i = \{\{1\},\{2\},\ldots,\{|\Sigma_i|\}\}$, $i = 1,2$, and $h(C_1) = h(C_2) = \mathrm{Conv}\{(h_1(s,t), h_2(s,t)) | (s,t) \in \Sigma_1 \times \Sigma_2\}$, which means that $h(C_1)$ and $h(C_2)$ are equal to the set of feasible payoffs. Thus, UEP = LEP = BEP_L = UNIF = Set of feasible payoffs, which are individually rational. This is what is known as the folk theorem [A3].

EXAMPLE 3.10. *The repeated game of*

		Player 2	
		L	R
Player 1	U	1, 1	1, 2
	D	2, 1	0, 0

where player i's information is his action and his payoff. This is not the case of standard information, since when player 2 acts L he cannot know whether player 1 acts U or D. Nevertheless, $S_1 = \{\{U\},\{D\}\}$ and $S_2 = \{\{L\},\{R\}\}$, and the result of the folk theorem holds here as well.

EXAMPLE 3.11. *The repeated game of*

		Player 2			
		W	X	Y	Z
Player 1	W	0, 0	0, 1	$\frac{3}{4}$, 0	0, 0
	X	1, 0	1, 1	0, 0	0, 0
	Y	0, 1	0, $\frac{1}{2}$	0, 1	0, 0
	Z	0, 0	0, 0	0, 0	0, 0

where player i's information is his action and his payoff. Here, $\{\{W, X\},\{Y\},\{Z\}\}$,

$$S_1 = S_2 = \{\{W\},\{X\},\{Y,Z\}\},$$

$$d_1 = d_2 = 0,$$

$$\operatorname{Conv} h(C) \cap IR = \operatorname{Conv}\{(\tfrac{3}{4},0),(1,1),(0,1),(0,0)\}$$

(see Figure 1).

Notice that $(1,0)$ is not an equilibrium payoff, due to the fact that the pair $(X, W) \in \Sigma_1 \times \Sigma_2$ is not in C_2. The action X of player 2 is greater than W and gives a greater payoff to player 2.

EXAMPLE 3.12. *The repeated game of*

		Player 2			
		W	X	Y	Z
Player 1	W	0, 0	0, 1	$\frac{3}{4}$, 0	0, 0
	X	1, 0	1, 1	0, 0	0, 0
	Y	0, 1	0, 1	0, 1	0, 0
	Z	0, 0	0, 0	0, 0	0, 0

where the information is the action and the payoff. This is almost identical to the game

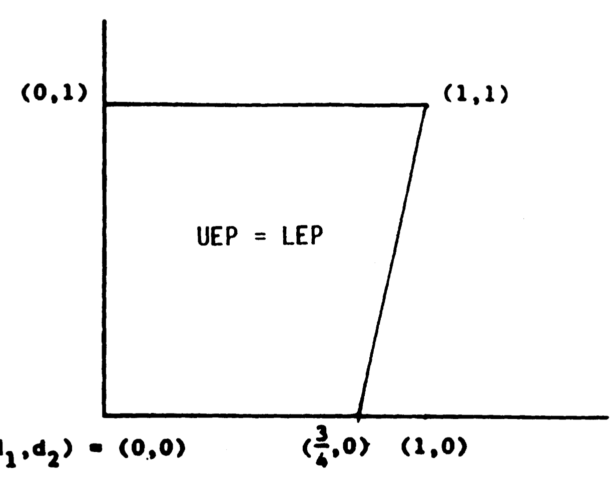

FIGURE 1

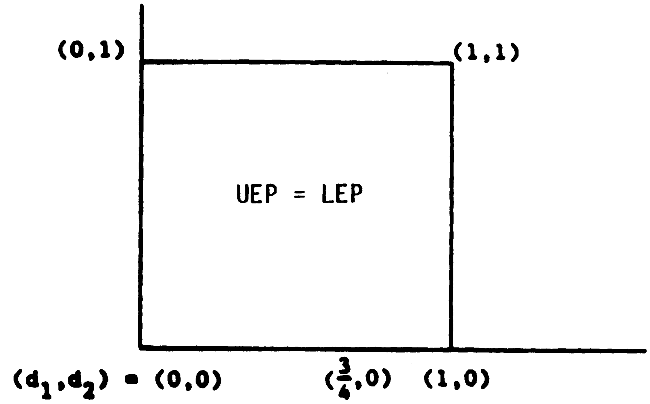

FIGURE 2

in the former example. The difference is in the payoff related to (Y, X) of $\Sigma_1 \times \Sigma_2$. By changing this payoff $(1,0)$ becomes a payoff in $h(C)$. Now the pair (X, W) belongs to C because player 2 cannot deviate to the strategy X without losing the ability to distinguish between the actions X and Y of player 1. In other words, $X \not\succ W$. Thus, the set of equilibrium payoffs is $\text{Conv}\{(0,1),(1,1),(1,0),(0,0)\}$ (see Figure 2).

Notice that by increasing *player 2's* payoff associated with (Y, X) from $\frac{1}{2}$ to 1 the Pareto optimal frontier is pushed *rightward*.

4. Properties of the sets D_i, C_i. The geometric properties of the sets D_i and C_i will play a significant role in the proof of the main theorem. We start with the simple observation:

LEMMA 4.1. *D_i and C_i are closed sets, $i = 1, 2$.*

PROOF. Clear. //

The following lemma is used repeatedly in what is to come.

LEMMA 4.2. *Let $(p, q) \in \Delta(\Sigma_1) \times \Delta(\Sigma_2)$.*

(i) *If $p' \succ p$ and $q' \succ q$ satisfy:*

$$h_1(p', q) = \operatorname*{Max}_{\bar{p} \succ p} h_1(\bar{p}, q), \tag{4.1}$$

$$h_2(p, q') = \operatorname*{Max}_{\bar{q} \succ q} h_2(p, \bar{q}), \tag{4.2}$$

then $(p', q') \in C$.

(ii) *If $p' \sim p$ and $q' \sim q$ satisfy* (4.1) *and* (4.2), *replacing $\succ$ with $\sim$, then $(p', q') \in D$.*

(iii) *If $p' \sim p$ (resp., $p' \succ p$) and $q \succ q$ (resp., $q' \sim q$) satisfy* (4.1) *(resp.,* (4.2)*) replacing $\succ$ with $\sim$, then $(p', q') \in D_1 \cap C_2$ (resp., $C_1 \cap D_2$).*

PROOF. We will show (i). Assume to the contrary that $(p', q') \notin C$. Without loss of generality, $(p', q') \notin C_1$, therefore there is $\bar{p} \succ p'$ s.t.

$$h_1(\bar{p}, q') > h_1(p', q'). \tag{4.3}$$

Since $q' \succ q$ and since the information includes the payoffs

$$h_1(\cdot, q) = h_1(\cdot, q'). \tag{4.4}$$

The relation $\succ$ is transitive and, therefore, $\bar{p} \succ p$. (4.3) and (4.4) imply that

$$h_1(\bar{p}, q) = h_1(\bar{p}, q') > h_1(p', q') = h_1(p', q),$$

a contradiction to the choice of p'. //

PROPOSITION 4.3. *There is a point* $(u_1, u_2) \in h(D)$ *such that* $u_i \leqslant d_i$, $i = 1, 2$.

PROOF. Recalling Notation 3.6, let $p^* \in \Delta(\Sigma_1)$ and $q^* \in \Delta(\Sigma_2)$ have the following properties: $p^* \sim \sigma_1$, $q^* \sim \sigma_2$, and
(i) $u_1 = h_1(p^*, \sigma_2) = \mathrm{Max}_{p \sim \sigma_1} h_1(p, \sigma_2) \leqslant \mathrm{Max}_{p \in \Delta(\Sigma_1)} h_1(p, \sigma_2) = d_1$, and
(ii) $u_2 = h_2(\sigma_1, q^*) = \mathrm{Max}_{q \sim \sigma_2} h_2(\sigma_1, q) \leqslant \mathrm{Max}_{q \in \Delta(\Sigma_2)} h_2(\sigma_1, q) = d_2$.
By Lemma 4.2, $(p^*, q^*) \in D$. Let $(u_1, u_2) = h(p^*, q^*)$. By the definition $u_i \leqslant d_i$, $i = 1, 2$. //

DEFINITION 4.4. Let $Q \subseteq \mathbb{R}^2$ be a closed set. A point $(a, b) \in Q$ is called an upper boundary point (UBP) in Q of type I if $a = \mathrm{Max}\{c | (c, b) \in Q\}$ and UBP in Q of type II if $b = \mathrm{Max}\{c | (a, c) \in Q\}$.

REMARK 4.5. Any boundary point in $\mathrm{Conv}\, h(C)$ which is strictly individually rational is an UPB in $\mathrm{Conv}\, h(C)$.

DEFINITION 4.6. Let B_i be a closed subset of $\Delta(\Sigma_i)$, $i = 1, 2$. The pair (B_1, B_2) has the C–I property[1] if any upper boundary point in $\mathrm{Conv}\, h(B_1) \cap \mathrm{Conv}\, h(B_2)$ is contained in $\mathrm{Conv}\, h(B_1 \cap B_2)$.

LEMMA 4.7. *The pairs* (D_1, D_2) *and* (C_1, C_2), (D_1, C_2), (C_1, D_2) *have the C–I property.*

PROOF. We will show it for (C_1, C_2).
Assume that (a, b) is an UBP in $\mathrm{Conv}\, h(C_1) \cap \mathrm{Conv}\, h(C_2)$. So that

$$(a, b) = \sum_{t=1}^{3} \alpha_t h(p^t, q^t) = \sum_{t=1}^{3} \beta_t h(\bar{p}^t, \bar{q}^t) \quad \text{where}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \beta_1 + \beta_2 + \beta_3 = 1,$$

$$\alpha_t \geqslant 0, \quad \beta_t \geqslant 0, \quad (p^t, q^t) \in C_1, \quad (\bar{p}^t, \bar{q}^t) \in C_2, \qquad t = 1, 2, 3.$$

Without loss of generality, (a, b) is an UBP of type II. We will show that $(p^t, q^t) \in C_2$. Assume to the contrary that $(p^t, q^t) \notin C_2$. It means that there exists $\hat{q}^t \succ q^t$ such that

$$h_2(p^t, \hat{q}^t) = \underset{q \sim q^t}{\mathrm{Max}}\, h_2(p^t, q) > h_2(p^t, q^t). \tag{4.5}$$

Since the information includes the payoffs

$$h_1(p^t, \hat{q}^t) = h_1(p^t, q^t). \tag{4.6}$$

[1] For Convex-Intersection.

By Lemma 4.2, $(p^t, \hat{q}^t) \in C$. Using (4.5) and (4.6) we obtain

$$\sum_{t=1}^{3} \alpha_t h_1(p^t, \hat{q}^t) = \sum_{t=1}^{3} \alpha_t h_1(p^t, q^t) \quad \text{and}$$

$$\sum_{t=1}^{3} \alpha_t h_2(p^t, \hat{q}^t) > \sum_{t=1}^{3} \alpha_t h_2(p^t, q^t),$$

which contradicts the assumption that (a, b) is of type II. The same proof, replacing $\succ$ with $\sim$ at the appropriate places, holds for all other pairs. //

The next proposition is responsible for the fact that LEP $\subseteq$ UEP in the observable payoffs case. (The inverse inclusion always exists.)

PROPOSITION 4.8.

$$\operatorname{Conv} h(C) \cap IR = \operatorname{Conv} h(C_1) \cap \operatorname{Conv} h(C_2) \cap IR,$$

and a similar equation for D_1, D_2.

PROOF. We will show it for C_i, and a similar proof holds for D_i. It is enough to show that $\operatorname{Conv} h(C) \cap IR \supseteq \operatorname{Conv} h(C_1) \cap \operatorname{Conv} h(C_2) \cap IR$. Let $(a', b') \in \operatorname{Conv} h(C_1) \cap \operatorname{Conv} h(C_2) \cap IR$, and let (a, b') and (a', b) be two upper boundary points of $\operatorname{Conv} h(C_1) \cap \operatorname{Conv} h(C_2)$ of types I and II, respectively. By Lemma 4.7, $(a', b), (a, b') \in \operatorname{Conv} h(C)$.

By Proposition 4.3 there is $(u_1, u_2) \leqslant (d_1, d_2)$, and $(u_1, u_2) \in h(D)$. By the definition $h(D) \subseteq h(C)$ and, thus,

$$(a', b') \in \operatorname{Conv}\{(a', b), (a, b'), (u_1, u_2)\} \subseteq \operatorname{Conv} h(C). \quad //$$

Notation 4.9.

$$\text{SPO} = \{(a, b) \in \mathbb{R}^2 | \text{If } (a', b') \in \operatorname{Conv} h(\Sigma_1 \times \Sigma_2)$$
$$\text{and } a' \geqslant a,\ b' \geqslant b, \text{ then } a' = a \text{ and } b' = b\},$$

i.e., SPO is the set of the strong Pareto optimal payoffs.

PROPOSITION 4.10. SPO $\subseteq \operatorname{Conv} h(D)$.

PROOF. Let $(a, b) \in$ SPO be an extreme point of $\operatorname{Conv} h(\Sigma)$. Since (a, b) is an extreme point there are p and q satisfying $h(p, q) = (a, b)$. Let p', q' satisfy (4.1) and (4.2) written with $\sim$, instead of with $\succ$. By Lemma (4.2)(ii), $(p', q') \in D$. Since $(a, b) \in$ SPO, $h_1(p', q') = a$ and $h_2(p', q') = b$. Therefore, those extreme points that are also strong Pareto optimum payoffs are included in $h(D)$. Thus, SPO $\subseteq \operatorname{Conv} h(D)$. //

REMARK 4.11. As has already been shown in Example 3.11, the individually rational and weak Pareto optimal payoffs are not always contained in $\operatorname{Conv} h(C) \cap IR$.

Notation 4.12. Let $B \subseteq \mathbb{R}^2$ be a convex and a closed set. Denote by $EX_{\mathrm{I}}(B)$ (resp., $EX_{\mathrm{II}}(B)$) the set of all the extreme points which are UBP of type I (resp., type II) in B.

LEMMA 4.13. (i) $EX_{\mathrm{I}}(\operatorname{Conv} h(C)) \subseteq h(D_1 \cap C_2)$ *and*
(ii) $EX_{\mathrm{II}}(\operatorname{Conv} h(C)) \subseteq h(D_2 \cap C_1)$.

PROOF. Let $(a, b) \in EX_I(\mathrm{Conv}\, h(C))$. Since (a, b) is an extreme point there is $(p, q) \in C$ so that $(a, b) = h(p, q)$. If $(p, q) \notin D_1$ there is $p' \sim p$ which satisfies $h_1(p', q) > h_1(p, q)$, $h_2(p', q) = h_2(p, q)$ and $(p', q) \in D_1$. By Lemma 4.7, $(p', q) \in D_1 \cap C_2$. This contradicts the assumption that (a, b) is of type I. The proof of (ii) is similar. //

The purpose of the following proposition is to facilitate the definition of an equilibrium which sustains payoffs in $\mathrm{conv}\, h(C) \setminus \mathrm{conv}\, h(D)$. It turns out that such a payoff which is also an extreme point of type I can be written as $h(p, q)$, where (i) q is a pure action and it is the best response up to a discovery, and (ii) by playing q only player 2 can already detect any profitable deviation of player 1. Thus, player 2 can play a pure action by which he can detect any profitable deviation of his opponent. Moreover, player 1 expects to get at any time only one signal, because player 2 should play a *pure* action.

PROPOSITION 4.14. (i) *If* $(a, b) \in EX_I(\mathrm{Conv}\, h(C)) \setminus \mathrm{Conv}\, h(D)$, *then there is* $(p, q) \in (\Delta(\Sigma_1) \times \Sigma_2) \cap C_2$ *such that*

$$h_1(p, q) = \mathrm{Max}\{h_1(p', q) | l_2(p', q) = l_2(p, q)\}.$$

(ii) *Similar to* (i) *if* (a, b) *is of type* II, *interchanging* 1 *and* 2.

PROOF. We will prove (i).

By the previous lemma there is $(p, q) \in D_1 \cap C_2$ such that $h(p, q) = (a, b)$. If, to the contrary, $q \notin \Sigma_2$ then q is a convex combination of pure actions, say, $q = \sum \alpha_i q_i$. Furthermore, for some i, $\alpha_i > 0$ and $(p, q_i) \notin D_1 \cap C_2$ (because (a, b) is an extreme point). Since $(p, q_i) \in C_2$, $(p, q_i) \notin D_1$ and that there is some $p_i \sim p$ such that $h_1(p_i, q_i) > h_1(p, q_i)$ and $(p_i, q_i) \in D_1$. Since $(p, q_i) \in C_2$ and $p \sim p_i$, we obtain by Lemma 4.2(iii) $(p_i, q_i) \in D_1 \cap C_2$. We have obtained that

$$h_1(p_i, q_i) > h_1(p, q_i) \quad \text{and} \quad h_2(p_i, q_i) = h_2(p, q_i).$$

One can find such p_i for every i and get

$$\sum \alpha_i h_1(p_i, q_i) > a, \quad \sum \alpha_i h_2(p_i, q_i) = b.$$

This means that (a, b) is not an UBP of type I in $h(\mathrm{Conv}\, D_1 \cap C_2) = \mathrm{Conv}\, h(D_1 \cap C_2)$, which is a contradiction. Thus, $q \in \Sigma_2$.

Auxiliary Claim. We can assume, without loss of generality, that there is no other $q' \in \Sigma_2$ such that $q' \succ q$ and $q \not\succ q'$.

PROOF OF THE CLAIM. If there is such a pure action let q' be a maximum with respect to $\succ$. We obtain $h_2(p, q') \leqslant h_2(p, q)$, because $h_2(p, q') > h_2(p, q)$ means that $(p, q) \notin C_2$. We have two cases. In the first one, $h_2(p, q') = h_2(p, q)$, we can take q' instead of taking q and get the desired claim. The second case, $h_2(p, q') < h_2(p, q)$, will lead to a contradiction. We assumed that $h(p, q) \notin D$. Thus, by Proposition 4.10, $h(p, q) \notin \mathrm{SPO}$. However, if the second case holds and q' is a maximum with respect to $\succ$, we get that $(p, q') \in D_1 \cap C_2$. Thus $h(p, q) \notin EX_I(\mathrm{Conv}\, h(C))$, a contradiction.

Return to the proposition's proof. It remains to show that (p, q) satisfies the second requirement mentioned in (i). If, to the contrary, there is $p' \in \Delta(\Sigma_1)$ such that $l_2(p', q) = l_2(p, q)$ (in particular, $h_2(p', q) = h_2(p, q)$), and $h_1(p', q) > h_1(p, q)$, then $(p', q) \notin C$ (because (a, b) is of type I). Take such p' that ensures that $(p', q) \in D_1$. By the previous observation, we get $(p', q) \notin C_2$. This means that there is $q' \succ q$ such that $h_2(p', q') > h_2(p', q)$. However, by the auxiliary claim, $q \succ q'$ and by the assumption that the information includes the payoffs, we get:

$$h_2(p, q') = h_2(p', q') > h_2(p', q) = h_2(p, q). \tag{4.7}$$

The first equality of (4.7) holds because $q \succ q'$ and thus $l_2(p', q) = l_2(p, q)$ implies $l_2(p', q') = l_2(p, q')$ which implies in turn $h_2(p', q') = h_2(p, q')$. (4.7) is a contradiction to the fact that $(p, q) \in C_2$. Hence, $h_1(p, q) = \text{Max}\{h_1(p', q) | l_2(p', q) = l_2(p, q)\}$, as desired. //

REMARK 4.15. Notice that (p, q) of Proposition 4.14(i) satisfies the following properties:

(a) q is a pure strategy and it is a best response versus p among all the strategies that are greater ($\succ$) than q.

(b) p is the best response versus q among all the strategies that preserve the same distribution on L_2, while player 2 plays q. In other words, p is not only the best response among all p' such that $p' \succ p$, it is the best response among a bigger set, namely the set of the actions that preserve the distribution over player 2's signals while he is playing q (and not necessarily while playing other pure actions).

The second property is crucial for the construction of the strategies in the third step of the next section. Player 2 will be required to adhere to q. It will be his best response up to a discovery, on one hand, and it will be sufficient, otherwise, for detecting any profitable deviation of player 1.

5. The proof of the main theorem.

DEFINITION 5.1. A *behavior strategy* of player i, $i = 1, 2$, in the repeated game is a sequence $f_i = (f_i^1, f_i^2, \ldots)$ of functions $f_i^n \colon L_i^{n-1} \to \Delta(\Sigma_i)$, $n = 1, 2, \ldots$.

A pair (f_1, f_2) of behavior strategies induces measures on $F_1 \times F_2$, and thus on $(A_1 \times A_2)^{\mathbb{N}}$.

REMARK 5.2. A repeated game with nonobservable actions is a game with a perfect recall. Therefore, according to the Kuhn theorem ([K], [A1]), we can deal with behavior strategies in the sequel, whenever it is convenient.

For any pair of behavior strategies (f_1, f_2), we can define $H^n(f_1, f_2)$, $n = 1, 2, \ldots$, and $H^*(f_1, f_2)$ as in §2. In addition, we can define for any $n \in \mathbb{N}$ a probability measure, $\text{prob}_{(f_1, f_2)}(\cdot)$, on L_i^n for $i \in (1, 2)$ and on $L_1^n \times L_2^n$, in a natural way.

PROOF. We will divide the proof into seven steps, as follows:

1. LEP $\subseteq$ Conv $h(C) \cap IR$.
2. Conv $h(D) \cap IR \subseteq$ UEP.
3. $[\partial \operatorname{Conv} h(C) \setminus \operatorname{Conv} h(D)] \cap IR \subseteq$ UEP.

By combining Steps 2 and 3 we arrive at:

4. Conv $h(C) \cap IR \subseteq$ UEP.
5. BEP_L = LEP = UEP.
6. UNIF = UEP. (The first six steps provide the proof of the nontrivial case.)
7. The seventh step proves the trivial information case.

Step 1. LEP $\subseteq$ Conv $h(C) \cap IR$. By [L2], the LEP of any two-player repeated game with nonobservable actions is equal to Conv $h(C_1) \cap$ Conv $h(C_2) \cap IR$. By Proposition 4.8, it is equal to Conv $h(C) \cap IR$.

Step 2. Conv $h(D) \cap IR \subseteq$ UEP. It is readily seen that SPO $\cap IR \subseteq$ UEP, but to see that Conv $h(D) \cap IR$ is contained in UEP is much more difficult.

SKETCH OF THE PROOF. We divide the set of periods, $\mathbb{N}$, into an infinite number of blocks: $M_1, M_2, \ldots,$ and then gather them into super blocks: $B_1 = M_1 \cup M_2 \cup \cdots \cup M_{k_1}$, $B_2 = M_{k_1+1} \cup \cdots \cup M_{k_2}, \ldots$. The lengths of M_i and k_i increase with

i. On each super block the strategies are defined independently of the signals received in the previous super blocks. In other words, each player loses his memory at the beginning of each super block. On each block the player checks his opponent to see whether he has deviated from the agreed strategy or not. The checking is done by some statistical tests on the appearance of the signals, based on Chebyshev inequality. If he finds a definition he "punishes" the opponent at those following blocks that are contained in the same super block. A player has the ability to "punish" his opponent because the payoffs we deal with are all individually rational. However, using those statistical tests, a player can come to the conclusion that the other player has deviated even when he actually has not. To prevent this type of mistake from influencing all the future payoffs, there are "moratorium" times at the beginning of each super block, where the players lose their memories.

Let $(a, b) \in \operatorname{Conv} h(D) \cap IR$. There are $(p^j, q^j) \in D$ and $\alpha^j \geqslant 0$, $j = 1, 2, 3$ such that $\sum_{j=1}^{3} \alpha^j = 1$ and $(a, b) = \sum_{j=1}^{3} \alpha^j h(p^j, q^j)$.

Define the sequences $\{n_i\}_{i=1}^{\infty}$, $\{\epsilon_i\}_{i=1}^{\infty}$, and $\{k_i\}_{i=1}^{\infty}$ as follows:

$$n_i = i^{100}, \quad \epsilon_i = i^{-10}, \quad k_i = 2^i, \qquad i = 1, 2, \dots .$$

Divide $\mathbb{N}$, the set of stages, into infinite consecutive blocks: $M_1, M_2, \dots,$ where M_i contains n_i stages. Gather blocks into consecutive super blocks: $B_1, B_2, \dots,$ where B_l contains 2^l blocks. Divide each block M_k into three parts, M_k^1, M_k^2 and M_k^3, so that for some sequence $\{\delta_k\}$, which tends to zero, and for some constant γ

$$\left| \# M_k^j / n_k - \alpha^j \right| < \gamma \cdot \delta_k \quad \text{for } j = 1, 2, 3.$$

Notation 5.3. Let $p = (p(1), \dots, p(m)) \in \Delta^m$ and $\epsilon > 0$. Then denote by p_ϵ the point in $\{\bar{p} \in \Delta^m | \bar{p}(i) \geqslant \epsilon\}$ which achieves the minimum distance (with respect to the maximum norm) from p.

In other words, p_ϵ is a modification of p in which every coordinate is at least ϵ. For the sake of convenience denote $p_k^j = p_{\epsilon_k}^j$, $q_k^j = q_{\epsilon_k}^j$.

In order to describe the equilibrium pair of strategies (f_1, f_2), we will define what we will call the "master plan"[2] of player 1 in block M_k. The master plan of player 1 in block M_k is to play p_k^j at the stages of M_k^j.

During the game, player 1 does some statistical tests. The relevant data for these tests are the following. Let $j \in \{1, 2, 3\}$, $e \in L_1$ and let M_k be some former block. Denote

$$O_k^j(e) = \#\left\{ n \in M_k^j | \text{The signal given to player 1 at stage } n \text{ was } e \right\},$$

i.e., $O_k^j(e)$ is the number of stages in M_k^j in which signal e was observed by player 1.

If player 1 plays p_k^j and player 2 plays q_k^j, then the probability that player 1 will get the signal e will be denoted by $\operatorname{prob}_{(p_k^j, q_k^j)}(e)$.

f_1 is defined as follows: player 1 plays his master plan, unless he has come to the conclusion that player 2 had deviated in some previous block of the same super block. If he finds a deviation in the same super block, player 1 acts so as to punish player 2, i.e., by playing σ_1 (see Definition 3.5(3)). If player 1 finds $e \in L_1$, and $j \in \{1, 2, 3\}$ so that

$$(5.1) \qquad \left| \left(O_k^j(e) / \# M_k^j \right) - \operatorname{prob}_{(p_k^j, q_k^j)}(e) \right| > \epsilon_k^2,$$

[2]This terminology is due to Hart [H].

he comes to the conclusion that player 2 has deviated at block M_k. The punishment takes place from M_{k+1} on, until the end of the super block.

The master plan of player 2 in M_k and his strategy f_2 are defined in a similar way.

Notation 5.4. (1) $\alpha_k = (1/n_k)\Sigma_{t\in M_k} x_1^t$, $\alpha_k^j = (1/\#M_k^j)\Sigma_{t\in M_j^k} x_1^t$, (2) $\beta_k = (1/n_k)\Sigma_{t\in M_k} x_2^t$, $\beta_k^j = (1/\#M_k^j)\Sigma_{t\in M_j^k} x_2^t$, i.e., α_k and β_k are the average payoffs of players 1 and 2, respectively, at block M_k.

LEMMA 5.5. $H^*(f_1, f_2) = (a, b)$.

PROOF. If in block M_k the players play their master plan, then the expectation of an average payoff in M_k is close to (a, b), i.e., there is a constant c_1 such that

$$\begin{aligned} &\left|E_{(f_1,f_2)}(\alpha_k) - a\right| \leqslant c_1 \cdot \mathrm{Max}\{\epsilon_k, \delta_k\} \quad \text{and} \\ &\left|E_{(f_1,f_2)}(\beta_k) - b\right| \leqslant c_1 \cdot \mathrm{Max}\{\epsilon_k, \delta_k\}. \end{aligned} \tag{5.2}$$

$E_{(f_1,f_2)}(\alpha_k, \beta_k)$ is not precisely (a, b), for the following two reasons:

1. The strategies played are not (p^j, q^j), but their modifications, (p_k^j, q_k^j);
2. (p_k^j, q_k^j) is played $\#M_k^j$ times and not exactly $n_k \cdot \alpha^j$ times, $j = 1, 2, 3$ (recall $|(\#M_k^j/n_k) - \alpha^j| < \gamma \cdot \delta_k$).

If, at block M_{k+1}, at least one player does not play his master plan, then one of two things must have occurred:

(i) Both players played their master plan at M_k, but it so happened that one of them came to the conclusion that his opponent had deviated. If player 1 came to this conclusion, it means that he found $t \in \{1, 2, 3\}$ and $e \in L_1$ so that (5.1) holds. Similarly for player 2 if he came to the conclusion that his opponent had deviated. What is the probability of these events? We will find an upper bound of the probability that (5.1) holds for certain $j \in (1, 2, 3)$, $k \in \mathbb{N}$ and $e \in L_1$. By Chebyshev inequality we get

$$\mathrm{prob}_{(f_1,f_2)}\left\{\left|(O_k^j(e)/\#B_k^j) - \mathrm{prob}_{(p_k^j,q_k^j)}(e)\right| > \epsilon_k^2\right\} < c_2/\#B_k^j(\epsilon_k)^4 < c_3/k^{50} \tag{5.3}$$

for some constants c_2, c_3, whenever k is sufficiently large. The probability that (5.1) has occurred for any j and e in block M_k is less than $(c_3/k^{50}) \cdot (\#\{1, 2, 3\} \cdot |L_1|) = c_4/k^{50}$. We now obtain that the probability of case (i) occurring is bounded by $2 \cdot c_4/k^{50}$ (the 2 is for two players).

(ii) The second case is when one of the players did not play his master plan at the former block, M_k. However, if (f_1, f_2) is played, it means that case (i) has occurred in some previous block, $M_{k'}$, which belongs to the same super block, say, B_l. The probability of this case is bound by

$$\sum_{k=k_l}^{k_{l+1}} 2c_4/k^{50} \leqslant c_5/k_l^{49}. \tag{5.4}$$

Combine (5.2) and (5.4) to get, for any k, which satisfies $M_k \subseteq B_l$,

$$\left|\mathrm{Exp}_{(f_1,f_2)}(\alpha_k, \beta_k) - (a, b)\right| < c_1 \cdot \mathrm{Max}\{\epsilon_k, \delta_k\} + c \cdot c_5/k_l^{49}, \tag{5.5}$$

where $c = 2\,\mathrm{Max}\{|x| \mid x \in A_1 \cup A_2\}$. Define $m_k = \mathrm{Max}\, M_k$.

By (5.5), $H^{m_k}(f_1, f_2) \to_{k \to \infty} (a, b)$. Now, if $n \in M_k$, then

$$\|H^n(f_1, f_2) - H^{m_{k-1}}(f_1, f_2)\|_\infty$$

$$= \left\| E_{(f_1, f_2)}\left((1/n) \sum_{t=1}^{n} (x_1^t, x_2^t) - (1/m_{k-1}) \sum_{t=1}^{m_{k-1}} (x_1^t, x_2^t) \right) \right\|_\infty$$

$$= \left\| E_{(f_1, f_2)}\left((1/n) \sum_{t=m_{k-1}+1}^{n} (x_1^t, x_2^t) + (1/n - 1/m_{k-1}) \sum_{t=1}^{m_{k-1}} (x_1^t, x_2^t) \right) \right\|_\infty$$

$$\leqslant (1/n)(n - m_{k-1})c + ((n - m_{k-1})/n) \|H^{m_{k-1}}(f_1, f_2)\|_\infty.$$

However, $n - m_{k-1} \leqslant n_k$, and

$$n_k/m_{k-1} = k^{100} \Big/ \sum_{j=1}^{k-1} j^{100} \to 0,$$

thus $\|H^n(f_1, f_2) - H^{m_{k-1}}(f_1, f_2)\|_\infty$ tends to zero when k tends to infinity. Hence, we have $H^*(f_1, f_2) = \lim_n H^n(f_1, f_2) = (a, b)$, as desired. //

In order to prove that (f_1, f_2) is an upper equilibrium, we need Lemma 5.6 of [L1]:

LEMMA 5.6. *Let $R_1, \ldots, R_n$ be a sequence of identically distributed Bernoulli random variables with parameter p, and let $Y_1, \ldots, Y_n$ be a sequence of Bernoulli random variables, such that for each $1 \leqslant l \leqslant n$, R_l is independent of $R_1, \ldots, R_{l-1}$, $Y_1, \ldots, Y_l$. Then we have*

$$\operatorname{prob}\{|(R_1 Y_1 + \cdots + R_n Y_n)/n - p(Y_1 + \cdots + Y_n)/n| > \epsilon\} < 1/n\epsilon^2.$$

LEMMA 5.7. *(f_1, f_2) is an upper equilibrium.*

PROOF. As previously noted, c is twice the maximal absolute value of the payoffs appearing in the game. Let g_2 be a mixed strategy of player 2, and let μ be the measure on $\Sigma_1^* \times \Sigma_2^*$ induced by (f_1, g_2). We will show that

$$\limsup_t (x_2^1 + \cdots + x_2^t)/t \leqslant b, \quad \mu\text{-a.s.},$$

which implies that $\limsup_t H_2^t(f_1, g_2) \leqslant b$. In a similar way one can show that for every strategy g_1 of player 1 there is $\limsup_t H_1^t(g_1, f_2) \leqslant a$. Fix $s_1 \in \Sigma_1$, $s_2 \in \Sigma_2$, $k \in \mathbb{N}$ and $j \in \{1, 2, 3\}$, and let $t \in M_l^j$. Set $R_t(s_1) = 1$ if player 1 plays s_1 at t and 0 otherwise; $Y_t(s_1, e) = 1$ if player 2 plays some action of $\{s_2 \in \Sigma_2 | l_1(s_1, s_2) = e\}$ at stage t, and 0 otherwise. Because of Definition 2.1.b(i),

$$O_k^j(e) = \sum_{t \in M_k^j} Y_t(s_1, e) R_t(s_1).$$

We have to calculate the probability of several events. We are interested in the event where player 2 gains by a deviation without being detected. By the first computation we will find the probability that player 2's payoff in M_k does not exceed by much his prescribed payoff, b, given that no deviation has been detected by player 1 in M_k (i.e., (5.1) does not hold for M_k) and that player 1 plays in M_k according to the master plan (i.e., no deviation has been found by him in that particular super block so far). By a similar method we will later compute the probability of the same event without assuming that player 1 plays according to the master plan. By Lemma 5.6,

$$\text{prob}\left\{\left|O_k^j(e)/\#M_k^j - \left(\text{prob}(R_t(s_1) = 1)/\#M_k^j\right) \sum_{t \in M_k^j} Y_t(s_1, e)\right| < (\epsilon_k)^2\right\}$$

$$> 1 - 1/\left(\#M_k^j \epsilon_k^4\right) > 1 - c_1/k^{50},$$

for a certain constant c_1. Given that (5.1) does not hold for M_k, with probability of at least $1 - (c_1/k^{50})$ the following holds:

$$\left|\text{prob}_{(p_k^j, q_k^j)}(e) - \left(\text{prob}(R_t(s_1) = 1)/\#M_k^j\right) \sum_{t \in M_k^j} Y_t(s_1, e)\right| < 2(\epsilon_k)^2.$$

Given that (5.1) does not hold for all previous blocks in the same super block, we have

$$\text{prob}(R_t(s_1) = 1) = p_k^j(s_1) \geqslant \epsilon_k, \quad \text{and}$$

$$\text{prob}_{(p_k^j, q_k^j)}(e) = p_k^j(s_1)\left[\sum_{s_2 :\, l_1(s_1, s_2) = e} q_k^j(s_2)\right].$$

Hence, with probability of at least $1 - (c_1/k^{50})$,

$$\left|\left(1/\#M_k^j\right) \sum_{t \in M_k^j} Y_t(s_1, e) - \sum_{l_1(s_1, s_2) = e} q_k^j(s_2)\right| < 2\epsilon_k. \tag{5.6}$$

The last inequality means that with a high probability (at least $1 - (c_1/k^{50})$) if player 1 did not detect any deviation, then the empirical distribution of player 2's actions in $\{s_2 \in \Sigma_2 | l_1(s_1, s_2) = e\}$ is close up to $2\epsilon_k$ to their prescribed probability, $\Sigma_{l_1(s_1, s_2) = e}\, q_k^j(s_2)$. Fix k and j and denote $Z_t(s_2) = 1$ when player 2 plays s_2 at stage t, and 0 otherwise. Let $W(s_2)$ denote the empirical distribution of s_2 in M_k^j, i.e.,

$$W(s_2) = \left(1/\#M_k^j\right) \sum_{t \in M_k^j} Z_t(s_2).$$

$$\beta_k^j = \left(1/\#M_k^j\right) \sum_{t \in M_k^j} \sum_{s_2 \in \Sigma_2} \sum_{s_1 \in \Sigma_1} R_t(s_1) Z_t(s_2) h_2(s_1, s_2)$$

$$= \sum_{s_2 \in \Sigma_2} \sum_{s_1 \in \Sigma_1} \left[\left(h_2(s_1, s_2)/\#M_k^j\right) \sum_{t \in M_k^j} R_t(s_1) Z_t(s_2)\right] = (*).$$

Again, apply Lemma 5.6 for $R_t(s_1)$, $Z_t(s_2)$, $t \in M_k^j$, and (given the fact that (5.1) does not hold for all previous blocks of the same super block, i.e., in particular, that

$\operatorname{prob}(R_t(s_1) = 1) = p_k^j(s_1)$, $s_1 \in \Sigma_1$) obtain with probability of at least $1 - (c_1/k^{50})$

$$\left|(1/\#M_k^j) \sum_{t \in M_k^j} R_t(s_1) Z_t(s_2) - (1/\#M_k^j) p_k^j(s_1) \sum_{t \in M_k^j} Z_t(s_2)\right| < \epsilon_k.$$

Thus,

$$(*) \leqslant \sum_{s_2 \in \Sigma_2} \sum_{s_1 \in \Sigma_1} \left[\left(h_2(s_1, s_2) p_k^j(s_1)/\#M_k^j\right)\left(\sum_{t \in M_k^j} Z_t(s_2) + c\epsilon_k\right)\right]$$

$$= \sum_{s_2 \in \Sigma_2} \left(h_2(p_k^j, s_2)/\#M_k^j\right)\left(\sum_{t \in M_k^j} Z_t(s_2) + c\epsilon_k |\Sigma_1|\right)$$

$$= \sum_{s_2 \in \Sigma_2} \left[\left(h_2(p_k^j, s_2) W(s_2)\right) + c\epsilon_k |\Sigma_1|\right]$$

$$= h_2(p_k^j, W) + |\Sigma_2|\,|\Sigma_1| c\epsilon_k = (**),$$

where $W = (w(s_2))_{s_2 \in \Sigma_2} \in \Delta(\Sigma_2)$. For every $(s_1, e) \in \Sigma_1 \times L_1$ we get by the definitions,

$$\text{(5.7)} \qquad \sum_{l_1(s_1, s_2) = e} W(s_2) = \sum_{l_1(s_1, s_2) = e} \left((1/\#M_k^j) \sum_{t \in M_k^j} Z_t(s_2)\right)$$

$$= (1/\#M_k^j) \sum_{t \in M_k^j} Y_t(s_1, e).$$

If (5.6) is satisfied for all s_1 and e then by (5.7), W is far, by at most $3\epsilon_k$, from being equivalent to q^j. Since $(p^j, q^j) \in D_2$, we get with a probability of at least $1 - (2|\Sigma_1|\,|\Sigma_2| c_1/k^{50})$

$$(**) \leqslant h_2(p^j, q^j) + 4|\Sigma_1| \cdot |\Sigma_2| \cdot c \cdot \epsilon_k.$$

To recapitulate: given that (5.1) does not hold for all blocks $M_{k'}$ $(k' \leqslant k)$ of the same super block B_l, with probability of at least $1 - (2|\Sigma_1|\,|\Sigma_2| c_1/k^{50})$

$$\text{(5.8a)} \qquad \beta_k^j \leqslant h_2(p^j, q^j) + 4|\Sigma_1|\,|\Sigma_2| c\epsilon_k,$$

for all k big enough and all $j \in \{1, 2, 3\}$. Hence,

$$\text{(5.8b)} \qquad \beta_k \leqslant b + 3 \cdot 2|\Sigma_1|\,|\Sigma_2| c\epsilon_k + c\gamma\delta_k$$

(the 3 is for $j \in \{1, 2, 3\}$ and the third term appears because $|\#M_k^j/n_k - \alpha_j| < \gamma\delta_k$).

(5.8b) can be written as follows. Let $\eta > 0$. The probability that

$$\text{(5.9)} \qquad \beta_k > b + \eta/4$$

and that player 1, playing according to the master plan, does not come to the conclusion that player 2 had deviated at block M_k (i.e., that (5.1) does not hold) is less than c_2/k^{50} when k is big enough and c_2 is a constant.

Now we proceed by evaluating the probability that player 2 gains by more than $6|\Sigma_1|\,|\Sigma_2| c\epsilon_k$ above his prescribed payoff b, while player 1 punishes rather than adheres to the master plan.

By an argument similar to the one above, given that $R_t(s_1) = \sigma_1(s_1)$ (which means that for some $M_{k'}$, $k' < k$ in the same B_l (5.1) does hold), with probability of at least $1 - (|\Sigma_1|\,|\Sigma_2|c_1)/k^{50}$ we have

$$\beta_k \leqslant d_2 + 6|\Sigma_1|\,|\Sigma_2|c\epsilon_k \leqslant b + 6|\Sigma_1|\,|\Sigma_2|c\epsilon_k. \tag{5.10}$$

The lengths of the blocks and of the super blocks have been defined in such a way that the length of a super block B_l relative to its past tends to infinity with l, and that the length of $M_k \subseteq B_l$ relative to the length of B_l tends to zero with l. Therefore, the event $\{\limsup_T (1/T)\Sigma_{t=1}^T x_2^t > b + \eta\}$ is included in the event

$$\left\{(1/\#B_l) \sum_{t \in B_l} x_2^t > b + \eta/2 \text{ for infinitely many } l\text{'s}\right\}.$$

Since $\#M_k/\#B_l \to_{l\to 0} 0$ for $m_k \subseteq B_l$, the event $\{(1/\#B_l)\Sigma_{t\in B_l} x_2^t > b + \eta/2\}$ is included in the union

$$\bigcup \{V_{k,k'} | k < k', \quad \text{and} \quad M_k \cup M_{k'} \subseteq B_l\},$$

where $V_{k,k'}$ is the event $\{\beta_k > b + \eta/4 \text{ and } \beta_{k'} > b + \eta/4\}$. In other words, $V_{k,k'}$ is the event in which, both in block M_k and in block $M_{k'}$, player 2's average payoff is greater than $b + \eta/4$.

The event $V_{k,k'}$ is included in the union of two events. The first one is that $\beta_k > b + \eta/4$ and (5.1) does not hold for M_k (whose probability is by (5.9) less than c_2/k^{50}) and the second one is that $\beta_k > b + \eta/4$ and (5.1) does hold (namely player 1 punishes player 2 at $M_{k'}$) and even though $\beta_{k'} > b + \eta/4$ (whose probability is, by (5.10), less than $|\Sigma_1|\,|\Sigma_2|c_1/k'^{50}$). Thus, for a fixed k,

$$\operatorname{prob}\left(\bigcup_{k<k'} V_{k,k'}\right) \leqslant (c_2/k^{50}) + \sum_{k<k'} |\Sigma_1|\,|\Sigma_2|c_1/k'^{50} \leqslant c_3/k^{49},$$

for some constant c_3. Fix an l. We then have that

$$\operatorname{prob}\left(\bigcup_{M_k \subseteq B_l} \bigcup_{k<k'} V_{k,k'}\right) \leqslant \sum_{k:\, M_k \subseteq B_l} (c_3/k^{49}) \leqslant c_4/k_l^{48}$$

for some constant c_4.

The sum of $c_4(k_l)^{48}$ over l is finite. Hence, by the Borel-Cantelli lemma (see [LO, p. 240]) the event $\limsup_T \Sigma_{t=1}^T x_2^t > b + \eta$ has probability zero. This concludes the proof. //

REMARK 5.8. a. The strategy constructed here is also a uniform equilibrium. The intermediate results (5.9) and (5.10) show it.

b. One can show that the method employed here goes through not only in the case of observable actions but also in any information structure. Notice that the definition of D as well as the proof do not rely on the particular information structure assumed here.

Step 3. $[\partial \operatorname{Conv} h(C) \setminus \operatorname{Conv} h(D)] \cup IR \subseteq$ UEP. The difference between payoffs in $h(D)$ and payoffs in $h(C) \setminus \operatorname{Conv} h(D)$ is that if $(p, q) \in D$, the player can take care of a deviation of the opponent during the regular game by playing (p^ϵ, q^ϵ) for some ϵ, and counting the number of times each signal has appeared. However, if $(p, q) \in C \setminus D$, it can happen that player 1, for example, will have p' such that $p' \sim p$ but $p' \not\succ p$, which increases his payoff. $p' \sim p$ means that p' does not change the probability of any signal in L_2, while player 2 continues playing q. Therefore, player 2 cannot take care of player 1's deviations during the regular game.

How can player 2 ensure that player 1 will play only strategies $\bar{p}$ that are greater than p? Player 2 can check player 1 whether the latter knows about the actions that took place at the former stages. Whenever $\bar{p} \succ p$, if by playing p player 1 can tell the difference between two actions, then he can also tell the difference by playing $\bar{p}$. Thus if player 1 does not know something he could have known by playing p, player 2 comes to the conclusion that an action which is not greater than p was played. In general, the players, in order to transmit information to one another, have to play an action that does not sustain the equilibrium payoff. For this reason this information is transmitted in a set of stages that have no influence on the payoff, namely a set with zero density. However, if a deviation is detected long after it has occurred, the deviator profits in the meantime. If the punishment would hold only for a while, and then the player would return to the master plan, the upper limit of the average payoffs of the deviator might exceed his prescribed payoff. This means that the strategy would have not been an upper equilibrium. Thus, the punishment has to take place from the detection moment on, forever. With this kind of punishment, however, the strategies must be qualified in such a way that if a player punishes his opponent, then there is a probability 1 that a deviation has actually taken place. We are able to define such a strategy by using the properties of the extreme points of $\operatorname{Conv} h(C)$, which are not in $\operatorname{Conv} h(D)$. Indeed, by Proposition 4.14, each extreme point is sustained by a pair of strategies (p_1, p_2), where one, say p_i, is a pure action and the other, say p_{3-i}, is a best response versus p_i among all the actions that preserve the distribution over L_i. This property of p_{3-i} enables player i to prevent his opponent from deviating only by playing p_i; no other action is needed for detecting profitable deviations. So player i plays with probability one the pure action p_i, which enables player $3-i$ to punish only in those cases where deviation has actually occurred, i.e., the probability for him to be mistaken is zero.

Let $(a,b) \in [(\partial \operatorname{Conv} h(C) \setminus \operatorname{Conv} h(D)] \cap IR$. By Remark 4.5, (a,b) is a UBP, without loss of generality it is of type I, and therefore it lies in a segment connecting two extreme points of the same type. By Proposition 4.14(i) there are $(p^1, q^1), (p^2, q^2) \in (\Delta(\Sigma_1) \times \Sigma_2) \cap C_2$ and $0 \leqslant \alpha \leqslant 1$ such that

$$h_1(p^i, q^i) = \operatorname{Max}\{h_1(\bar{p}, q^i) | l_2(\bar{p}, q^i) = l_2(p^i, q^i)\}, \qquad i = 1, 2, \quad \text{and}$$

$$(a,b) = \alpha h(p^1, q^1) + (1-\alpha) h(p^2, q^2).$$

We will describe the behavior strategies f_1 and f_2 so that $H^*(f_1, f_2) = (a,b)$. The joint strategy will not be a uniform equilibrium. A modified strategy, to be defined in Step 6, will be a uniform strategy. Take an infinite set $W \subseteq \mathbb{N}$ with zero density, i.e., $\lim_n |W \cap \{1, \dots, n\}|/n = 0$. As in the third step, divide $\mathbb{N} \setminus W$ into consecutive blocks $M_1, M_2, \dots$. Take an infinite sequence of real numbers $\{\epsilon_k\}_1^\infty$ such that $\epsilon_k = k^{-10}$. Divide each M_k into M_k^1 and M_k^2 in such a way that $|(\#M_k^1/\#M_k) - \alpha| \leqslant c_1 \delta_k$, where $c_1 \in \mathbb{R}$ and $\delta_k \to_{k \to \infty} 0$.

The strategies are defined as follows: for each $k \in \mathbb{N}$, in stages of M_k, player 2 plays the pure action q^i and player 1 plays $p_k^i = p_{\epsilon_k}^i$, except in the cases described below. After each block player 1 checks the signal he got at that block. Player 2 must always play a pure action; thus, by checking his signals, player 1 can discover deviations to strategies q satisfies that $q \nsim q_j$ with probability 1. In case he finds a defection, he punishes player 2 forever. However, player 2 has the option to deviate to an action $\bar{q}$ which is equivalent to q_j, but $\bar{q} \nsucc q_j$.

The way to prevent player 2 from deviating to such $\bar{q}$ is to "ask" him, at the stages of the zero density set, W, about signals he received at previous stages. By asking a finite number of "Yes-No" questions, one can know what signals player 2 received at any former stage. Player 2 can answer these questions by acting a certain $s_1 \in \Sigma_2$ for "Yes" and a certain $s_2 \in \Sigma_2$ for "No." s_1 and s_2 have to be chosen in such a way that player 1 will be able to distinguish between them, i.e., there is $v \in \Sigma_1$ such that

$$l_1(v, s_1) \neq l_1(v, s_2). \tag{5.11}$$

This can be done, since player 1 has nontrivial information. There are infinite stages $w \in W$, so that for any stage $n \notin W$ we will correlate r states $w_1(n), \dots, w_r(n)$ in W, where $n \leqslant w_i(n)$, $1 \leqslant i \leqslant r$. In these stages player 2 will have to answer "Yes-No" questions about the signal he got at stage n. It is enough to ask[3] $r = |L_2| - 1$ "Yes-No" questions to know what signal player 2 got at any stage. Call these questions $\phi_1, \phi_2, \dots, \phi_r$. When player 1 finds a wrong answer he comes to the conclusion that $\bar{q}$ such that $q \not\succ q^j$ was played and he then punishes player 2 forever.

Player 2 also checks his opponent at the end of M_k. If the relative frequency of appearance of each signal does not exceed the expected number by more than ϵ_k, he punishes player 1 and the punishment is carried out forever. Therefore, player 1 has to be careful not to repeat any action too many times. At any stage $r \in M_k$, player 1 has to check for each action $u \in \Sigma_1$ how many times he acted u. In case he finds $\varnothing \neq T \subseteq \Sigma_1$ so that every $u \in T$ is carried out at M_k^j more than $\#M_k^j(p_k^j(u) + 2\epsilon_k)$ times, then at the rest of the M_k^j it is not p that will be played but rather $p_{k_T}^j$ where $p_{k_T}^j$ is defined as follows:

DEFINITION 5.9. Let $p = (p(1), \dots, p(l)) \in \Delta^l$ for some $l \in \mathbb{N}$, and $T \subseteq \{1, \dots, l\}$ if $1 > \sum_{s \notin T} p(s)$. Then $p_T = (p_T(1), \dots, p_T(l))$ is defined by

$$p_T(u) = \begin{cases} 0, & u \in T, \\ p(u) / \sum\limits_{s \notin T} p(s), & u \notin T. \end{cases}$$

In other words, if player 2 has found no deviation at any previous block, he plays either q_1 or q_2, depending on whether the stage belongs to M_k^1 or to M_k^2. He answers at stage $w_\alpha(n)$ to the question ϕ_α, $\alpha = 1, \dots, r - 1$, referring to stage n. If the answer is "Yes," he plays s_1, and if the answer is "No," he plays s_2 (see (5.11)).

Player 2 comes to the conclusion that there has been a deviation at some previous block M_k if he finds $j \in \{1, 2\}$ and a signal $s \in L_1$ whose relative frequency at M_k^j $(O_k^j(a)/\#M_k^j)$ is far from the expected number. In a precise way, if

$$\left| O_k^j(a)/\#M_k^j - \mathrm{prob}_{(p_k^j, q^j)}(a) \right| > 2\epsilon_k, \tag{5.12}$$

then player 2 plays σ_2 from M_{k+1} on, forever.

Player 1 plays according to the following rules:

1. He starts by playing p_k^j at M_k^j and continues for as long as he does not find that some action $u \in \Sigma_1$ has been played in M_k^j more than $\#M_k^j(p_k^j(u) + \epsilon_k)$ times.

2. At the moment he finds such an action u, he starts playing $p_{k_T}^j$ where T is the set of all these u.

3. At the stages of W, player 1 plays v (see (5.11)).

4. He starts punishing his opponents by playing σ_1 if either:

[3] r could be chosen to be $[\log_2 |L_2|] + 1$, but then at the stages of W the questions should have been conditioned by the previous answers. We take $r = |L_2| - 1$ for the sake of simplicity.

(i) he finds that player 2 did not play q^j at some previous stage (for example if, at stage $t \in M_k$, he acted $u \in \Sigma_1$ and his signal in that stage differs from $l_1(u, q^j)$), or

(ii) he finds a wrong answer at some stage $w_\alpha(n) \in W$ (for example, if at stage $t \in M_k$, player 1 had played u and $l_2(s', q) = \alpha$, but at stage $w_\alpha(n)$ player 1 got the signal $l_1(v, s_2)$), which means the answer to the question "Did you get the signal α at stage n?" was "No."

The proof that $H^*(f_1, f_2) = (a, b)$ and that (f_1, f_2) is an upper equilibrium point is based on arguments similar to those employed in the proof of the second step. Notice that the probability that player 1 will not play p_k^j at stage $t \in M_k^j$, but rather $p_{k_T}^j$ for some $T \neq \varnothing$, in infinitely many blocks is by using the Chebyshev inequality and the Borel-Cantelli lemma, equal to zero. Therefore, the probability that player 2 will profit by a deviation on those stages where player 1 does not play p_k^j but $p_{k_T}^j$ for some $T \neq \varnothing$ infinitely many times is zero. This means that the limsup of the expected average payoffs does not exceed the prescribed payoff.

REMARK 5.10. a. The strategy of the previous step could be defined in another way, namely player 1 will play always p_k^j and player 2 will punish player 1 in a case where the relative frequency of a certain signal is far from the expected frequency by at least $2\epsilon_k$. As in Step 2, this punishment has to take place in a finite number of stages. The proof that such a punishment will occur infinitely many times, while player 1 plays according to the prescribed strategy has probability zero, involves the arguments that appeared in Step 2.

b. Notice that the strategy defined here is not uniform. The reason is that information about what has been played at stage t is transmitted long after time t. Thus, for any time T, there are many stages t, $t \leqslant T$, about which a player will have to inform long after T. Therefore, in the T-fold repeated game, the strategy does not induce an almost Nash equilibrium. In Step 6, we will modify the strategy, still using the idea of asking and answering "Yes"–"No" questions, to define a uniform equilibrium.

Step 4. $\operatorname{Conv} h(C) \cap IR \subseteq$ UEP. Any point in $\operatorname{Conv} h(C) \cap IR$ is the convex combination of $(a, b) \in \partial(\operatorname{Conv} h(C)) \setminus \operatorname{Conv} h(D) \cap IR$, and $(c, d) \in \operatorname{Conv} h(D) \cap IR$.

According to the former steps there are upper equilibrium strategies (f_1, f_2) and (g_1, g_2) such that $H^*(f_1, f_2) = (a, b)$, and $H^*(g_1, g_2) = (c, d)$.

Take $0 \leqslant \alpha \leqslant 1$ and divide $\mathbb{N}$ into two parts, V and U, with density α and $(1-\alpha)$, respectively, $V = \{v_1 < v_2 < \dots\}$, $U = \{u_1 < u_2 < \dots\}$. Define strategies k_1, k_2 as follows: for every $i \in \{1, 2\}$, $n \in \mathbb{N}$ and $a \in L_i^{n-1}$

$$k_i^n(a) = \begin{cases} f_i^j(a|V) & \text{if } n = v_j \in V, \\ g_i^j(a|U) & \text{if } n = u_j \in U, \end{cases}$$

where $a|V = (a_{v_j})_{v_j \leqslant n-1}$ and $a|U = (a_{u_j})_{u_j \leqslant n-1}$, $k_i = (k_i^1, k_i^2, \dots)$. It is clear that $H^*(k_1, k_2) = \alpha(a, b) + (1-\alpha)(c, d)$.

The proof that (k_1, k_2) is an upper equilibrium is achieved by employing the proof of Step 2.

Step 5. BEP_L = LEP = UEP. Obviously UEP $\subseteq \mathrm{BEP}_L$ for any L. That side of the proof in [L2] that proves LEP $\subseteq \operatorname{Conv} h(C_1) \cap \operatorname{Conv} h(C_2) \cap IR$ can be translated easily into terms of Banach equilibrium. Since, by Steps 1–4, LEP $\subseteq$ UEP, and by the definitions UEP $\subseteq$ LEP, we get the desired equality.

Step 6. UNIF = UEP. It is clear that UNIF $\subseteq$ UEP. To show that UEP $\subseteq$ UNIF one should construct for every point in UEP a strategy that sustains it. As was noted in Remark 5.8(b), the strategy defined in Step 2 is a uniform equilibrium, and thus $\operatorname{Conv} h(D) \cap IR \subseteq$ UNIF. If the strategy of Step 3 could have been modified so as to become a uniform equilibrium, we could have used the method of Step 4 to conclude that $\operatorname{Conv} h(C) \cap IR \subseteq$ UNIF.

We will use a method described by Sorin [S2] in order to modify the strategy of Step 3.

Recall that $\#M_k = k^{100}$. Any block M_k will be divided into k^{99} subblocks, each of which with length k. Between any two subblocks we will insert two segments of stages. In the first one, a state from the previous subblock will be chosen randomly by player 2. In the second segment, player 1 will have to inform about the signal that he received at the chosen stage.

In the first segment, which is of length $[\log k] + 1$, player 1 plays with probability $\frac{1}{2}$ each of the actions s_1 and s_2. These random moves generate a random string of length $[\log k] + 1$. The random strings encode stages in the subblock and thereby assign any stage a probability of at least $1/2k$. Player 2 can observe the random string by playing v, while player 1 randomizes. Immediately after observing the random string, player 2 should report the signals he got at the stage encoded by that string. This report takes place in the second segment of stages that follows every subblock. Thus, this segment should be of length $|\Sigma_2| - 1$.

To recapitulate, we have the block M_k which is divided into subblocks of length k each. Two segments of stages proceed any subblock. The purpose of these segments is to check possible deviations of player 2 in the previous subblock. At the first segment, player 1 announces the stage at which player 2 will have to inform in the following segment, the second one.

From the moment player 1 discovers a deviation, he punishes his opponent forever. The description of player 2's strategy is very much the same as the one given in Step 3. Denote the strategy defined here by f.

As opposed to the strategy defined in Step 3, in f the report player 2 must send to player 1 is not delayed until the far future. Here it is done immediately after any subblock, at the expense that on signals of any stage t, player 2 is asked only with a positive probability and not with probability one.

Let $n_k = \sum_{k' \leqslant k} \#M_{k'}$. To see that, indeed, f is a uniform strategy, observe the following.

1. $\#M_k / n_{k-1} \to 0$.

2. The payoff collected during the two segments preceding a subblock do not have much influence because their length is small compared to the length of the previous subblock.

3. To see that f is a uniform equilibrium it is enough, by 1., to show that f induces a δ_k-Nash equilibrium in the n_k-fold repeated game, G_{n_k}, for some δ_k converging to zero.

4. For any $\eta > 0$ there are K and $\epsilon > 0$ s.t. if $K < k$ and player 2 gets in M_k η more than his prescribed payoff, then in a fraction of at least ϵ subblocks there are at least ϵk stages in which player 2 plays in a probability of at least ϵk stages in which player 2 plays in a probability of at least ϵ worse action (i.e., some action $\bar{q}$, where $\bar{q} \not\succ q$) than the prescribed action q.

5. The probability that one of the stages, mentioned in 4., will be picked by player 1 for the report of player 2 is at least $\epsilon/2$.

6. The probability that player 2 will not be able to distinguish between the actions that he could have been able to by playing q is ϵ_k (because every action of player 1 is played at least with probability ϵ_k) times ϵ (at least the probability of deviation).

7. The probability not to discover deviation after a subblock of the type mentioned in 4. is at most $1 - \epsilon^2 \epsilon_k / 2$.

8. The probability not to discover the deviation in M_k is at most $(1 - \epsilon^2\epsilon_k/2)^{\#M_k/k}$. Recalling $\epsilon_k = k^{-10}$, we obtain that the probability in question is at least of the order of $\exp(-\epsilon^2 k^{89}/2)$, which tends to zero as k goes to infinity.

9. For k relatively big (to K), we will show that f induces a δ_k-Nash equilibrium in G_{n_k}, where δ_k will be specified later. Let M_k be the first block after K in which player 2's payoff exceeds his prescribed payoff by more than η. Player 2's total payoff (not the average) in G_{n_k} is then bounded from above by

(5.13)

$$cn_K + (b + \eta)(n_{k'-1} - n_K) + c\#M_k$$

$$+(1 - \exp(-\epsilon^2 k^{89}/2))d_2(n_k - n_{k'}) + \exp(-\epsilon^2 k^{89}/2)c(n_k - n_{k'}).$$

The first term stands for all stages preceding M_K. The second term stands for the blocks between M_K and $M_{k'}$, the third term for $M_{k'}$ itself. The fourth term is the punishment payoff (d_2) multiplied by the probability of punishment in the stages after M_k. The last term stands for the probability that player 2 will deviate in M_k and the deviation will not be discovered by player 1 multiplied by the maximal payoff, c.

10. Notice that after dividing (5.13) with n_k, in order to get the average payoff, we obtain $b + \delta_k$ where $\delta_k \to 0$.

This finishes the proof that f is a uniform equilibrium. //

Step 7. The Trivial Case. By Step 2, $\operatorname{Conv} h(D_1) \cap \operatorname{Conv} h(D_2) \cap IR \subseteq$ UEP. By [L2], whenever at least one player has trivial information in any repeated game with nonobservable actions, LEP $\subseteq \operatorname{Conv} h(D_1) \cap \operatorname{Conv} h(D_2) \cap IR$. Since UEP $\subseteq$ LEP, the proof is finished.

6. Concluding remarks.

6.1. (i) We defined an equilibrium as a joint mixed strategy that satisfies incentive compatibility. Kuhn's theorem enabled us to consider behavior strategy without restricting generality. We could use Kuhn's theorem because we assumed that each player is informed about his own actions (see Definition 2.1.b). However, it turns out that this assumption is not needed.

The main theorem also holds in a case where the information includes the payoffs and not necessarily the actions. For the construction of the behavior strategy involved in the proof, we can deal with an auxiliary game in which the players are provided with the information of their own actions in addition to the information given by the original information function. These behavior strategies induce mixed strategy in the auxiliary game. However, by the Dalkey theorem [D], any mixed strategy of player i can be played without knowing player i's own actions. In other words, the induced mixed strategy in the auxiliary game induces in turn an equivalent mixed strategy in the original game. Equivalent strategies in the sense that with any strategy of the opponent both induce the same distribution over histories.

To recapitulate, the assumption of Definition 2.1.b is not essential for proving Theorem 3.7 and it is valid even when the players are not informed of their own actions.

(ii) The proof of Step 2 does not depend on the particular information structure, and therefore it proves that in general two-player repeated games with nonobservable action, $h(D) \cap IR \subseteq$ UEP.

(iii) There are few similarities[4] between the construction defined at Step 2 and the strategy constructed by Kohlberg in [Ko]. He defined an ϵ-strategy of the uninformed player in a zero-sum repeated game with incomplete information, lack of information on one side and with a general information structure. In that strategy player 2 (traditionally, the uninformed player) divides the set of stages into blocks. At all the stages belonging to a certain block any player plays the same one-shot game mixed action. This mixed action is determined by the relative frequency of the signals at the former block. So, player 2 is updating his actions, relying on the relative frequency of the signals he formerly got. The length of blocks is defined in such a way that player 2 (the minimizer) will achieve at most the value $+\epsilon$.

6.2. *A trial to extend the characterization of the general case.* The question of characterizing UEP in the general case is still open, while the characterization of LEP in the two-player general case appears in [L2].

The first guess is that UEP $=$ Conv $h(C) \cap IR$ is valid also in the general case. We will provide an example in which UEP $\setminus$ (Conv $h(C) \cap IR$) is nonvoid.

EXAMPLE 6.1 (Based on an example of Aumann [A2]). The repeated game of

	W	X	Y	Z
W	0, 0	0, 0	0, 0	0, 0
X	0, 0	0, 0	0, 0	0, 0
Y	0, 0	0, 0	6, 6	2, 7
Z	0, 0	0, 0	7, 2	0, 0

Payoffs

	W	X	Y	Z
W	a, a	a, b	a, c	a, d
X	b, a	b', b'	b, c	b, d
Y	c, a	c, b	c, c	c, d
Z	d, a	d, b	d, c	d, d

Signals

We will show that the payoff $(5, 5)$ which is sustained by a one-shot game correlated equilibrium (see [A2]) is contained in UEP. In this example, Conv $h(C) \cap IR$ is the convex hull of all the Nash equilibrium payoffs, and then $(5, 5)$ is not contained in it. The strategy is based on the following procedure for both players. Play W with probability $\frac{1}{2}$ and X with probability $\frac{1}{2}$. If the signal is a, play Y; if the signal is b, play Z; and if the signal is b', play again W and X with probability $\frac{1}{2}$ each until the first time where the signal is not b'. Given that b' was not the signal, the joint distribution of the signals is given by

	a	b
a	$\frac{1}{3}$	$\frac{1}{3}$
b	$\frac{1}{3}$	0

Therefore the correlation between players is done by the information matrix and the strategies without the need of a mediator. These instructions do not define an equilibrium strategy that sustains the payoff $(5, 5)$, because a player can cheat in the first part (the construction of the correlation matrix) and play with probability 1 the action X, for example, and by that to profit at the second part. However, the other player can control this kind of cheating by repeating the procedure many times and detecting possible deviations by considering the frequency of the appearance of the signal b'. Then to play the second part many times, and so on, repeatedly, infinitely many times. By a strategy of this sort $(5, 5)$ is becoming an upper equilibrium payoff.

A more extensive study of correlation by histories can be found in [L3].

[4]This similarity was pointed out by S. Sorin.

6.3. *Another definition of* $\sim$. We could define the equivalence relation $\sim$ on mixed actions in another way. $p \sim p'$ $(p, p' \in \Delta(\Sigma_i))$ if p and p' give the same probability to every equivalence class of Σ_i. By this definition we could define in a similar way the partial order $\succ$, and the sets C_i', D_i'. Since the new definition of $\sim$ is less restrictive, $C_i \subseteq C_i'$ and $D_i \subseteq D_i'$.

It can be proven that $\operatorname{Conv} h(C_1' \cap C_2') \cap IR = \operatorname{Conv} h(C) \cap IR$, and the same equality with D_i'. Thus, the main theorem could have been stated also in terms of the new definition, which gives an earlier way for calculating UEP.

6.4. *More propositions that can be proven.* (i) The set $\operatorname{Conv} h(C)$ is a polygon.

(ii) Let $|S_i| = c_i$, i.e., Σ_i is divided into c_i equivalence classes. For any pair of vectors in the unit simplex Δ^{c_i}, $i = 1, 2$, (α, β), there is a pair $(p, q) \in D$ such that p (resp., q) gives the jth class of S_1 (resp., S_2) the total probability α_j (resp., β_j). This can be proven by a standard fixed point argument.

Acknowledgements. I am grateful to R. J. Aumann for his supervision, and to F. Forges, J. F. Mertens, A. Neyman and S. Sorin for their helpful comments.

References

[APS] Abreu, D, Pearce, D. and Stachetti, E. (1986). Optimal Cartel Equilibria with Imperfect Monitoring. *J. Economic Theory* (forthcoming).

[A1] Aumann, R. J. (1964). Mixed Behavior Strategies in Infinite Extensive Games. In *Advances in Game Theory. Ann. Math. Studies*, **52**, M. Dresher et al. (Eds.), Princeton University Press, Princeton, NJ, 627–650.

[A2] ______. (1974). Subjectivity and Correlation in Randomized Strategies. *J. Math. Economics* **1** 67–95.

[A3] ______. (1981). Survey of Repeated Games. In *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Bibliographisches Institute, Mannheim/Wein/Zurich, 11–42.

[D] Dalkey, N. (1953). Equivalence of Information Patterns and Essentiality Determinant Games. In Kuhn, H. W. and Tucker, A. W. (Eds.), *Contribution to the Theory of Games. II. Ann. Math. Studies*, **28**, Princeton University Press, Princeton, NJ, 217–243.

[FL] Fudenberg, D. and Levine, D. (1989). An Approximate Folk Theorem with Imperfect Private Information. Mimeo.

[FM] ______. and Maskin, E. (1986). Discounted Repeated Games with Unobservable Actions. I. One-Sided Moral Hazard. Mimeo.

[H] Hart, S. (1985). Nonzero-Sum Two-Person Repeated Games with Incomplete Information. *Math. Oper. Res.* **10** 117–153.

[Ko] Kohlberg, E. (1975). Optimal Strategies in Repeated Games with Incomplete Information. *Internat. J. Game Theory* **4** 1–24.

[K] Kuhn, H. W. (1953). Extensive Games and the Problem of Information. In *Contributions to the Theory of Games II, Ann. Math. Studies*, **28**, H. W. Kuhn and A. W. Tucker (Eds.), Princeton University Press, Princeton, NJ, 193–216.

[L1] Lehrer, E. (1990). Nash Equilibria of n-Player Repeated Games with Semi-Standard Information. *Internat. J. Game Theory* **19** 191–217.

[L2] ______. (1989). Lower Equilibrium Payoffs in Two-Player Repeated Games with Non-Observable Actions. *Internat. J. Game Theory* **18** 57–89.

[L3] ______. (1991). Internal Correlation in Repeated Games. *Internat. J. Game Theory* **19** 431–456.

[LO] Loeve, M. (1987). *Probability Theory I.* (4th Ed.), Springer-Verlag, Berlin and New York.

[R1] Radner, R. (1981). Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship. *Econometrica* **49** 1127–1148.

[R2] ______. (1986). Repeated Partnership Games with Imperfect Monitoring and No Discounting. *Rev. Economic Studies* **53**, **1** 43–58.

[Ru] Rubinstein, A. (1979). Repeated Insurance Contracts and Moral Hazard. *J. Economic Theory* **21** 7–9.

[RY] ______. and Yaari, M. (1983). Repeated Insurance Contracts and Moral Hazard. *J. Economic Theory* **30** 74–97.

[S1] Sorin, S. (1988). Repeated Games with Complete Information. CORE Discussion Paper No. 8822.

[S2] ______. (1988). Supergames. In *Game Theory and Applications*. T. Ichiichi et al. (Eds.), Academic Press, New York, 46–63.

DEPARTMENT OF MANAGERIAL ECONOMICS AND DECISION SCIENCES, J. L. KELLOGG GRADUATE SCHOOL OF MANAGEMENT AND DEPARTMENT OF MATHEMATICS, NORTHWESTERN UNIVERSITY, 2001 SHERIDAN ROAD, EVANSTON, ILLINOIS 60208