Journal of Economic Theory **98**, 26–54 (2001)
doi:10.1006/jeth.2000.2746, available online at http://www.idealibrary.com on IDEAL®

A General Class of Adaptive Strategies[1]

Sergiu Hart[2]

Center for Rationality and Interactive Decision Theory; Department of Mathematics; and Department of Economics, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
hart@huji.ac.il

and

Andreu Mas-Colell

Department of Economics and Business; and CREI, Universitat Pompeu Fabra, Ramon Trias Fargas, 25–27, 08005 Barcelona, Spain
mcolell@upf.es

Received April 19, 1999; final version received August 16, 2000

We exhibit and characterize an entire class of simple adaptive strategies, in the repeated play of a game, having the Hannan-consistency property: in the long-run, the player is guaranteed an average payoff as large as the best-reply payoff to the empirical distribution of play of the other players; i.e., there is no "regret." Smooth fictitious play (Fudenberg and Levine [1995, *J. Econ. Dynam. Control* **19**, 1065–1090]) and regret-matching (Hart and Mas-Colell [2000, *Econometrica* **68**, 1127–1150]) are particular cases. The motivation and application of the current paper come from the study of procedures whose empirical distribution of play is, in the long run, (almost) a correlated equilibrium. For the analysis we first develop a generalization of Blackwell's (1956, *Pacific J. Math.* **6**, 1–8) approachability strategy for games with vector payoffs. *Journal of Economic Literature* Classification Numbers: C7, D7, C6. © 2001 Academic Press

Key Words: adaptive strategies; approachability; correlated equilibrium; fictitious play; regret; regret-matching; smooth fictitious play.

1. INTRODUCTION

Consider a game repeated through time. We are interested in strategies of play which, while simple to implement, generate desirable outcomes.

[1] The authors thank the referees and editors for their useful comments. The research was partially supported by grants of the Israel Academy of Sciences and Humanities, the Spanish Ministry of Education, the Generalitat de Catalunya, CREI, and the EU-TMR Research Network.

[2] To whom correspondence should be addressed.

0022-0531/01 $35.00

Such strategies, typically consisting of moves in "improving" directions, are usually referred to as adaptive.

In Hart and Mas-Colell [16] we presented simple adaptive strategies with the property that, if used by all players, the empirical distribution of play is, in the long run, (almost) a correlated equilibrium of the game (for other procedures leading to correlated equilibria, see[3] Foster and Vohra [8] and Fudenberg and Levine [14]. From this work we are led—for reasons we will comment upon shortly—to the study of a concept originally introduced by Hannan [15]. A strategy of a player is called *Hannan-consistent* if it guarantees that his long-run average payoff is as large as the highest payoff that can be obtained (i.e., the one-shot best-reply payoff) against the empirical distribution of play of the other players. In other words, a strategy is Hannan-consistent if, given the play of the others, there is no regret in the long run for not having played (constantly) any particular action. As a matter of terminology, the *regret* of player i for an action[4] k at period t is the difference in his average payoff up to t that results from replacing his actual past play by the constant play of action k. Hannan-consistency thus means that all regrets are nonpositive as t goes to infinity.

In this paper we concentrate on the notion of Hannan-consistency, rather than on its stronger conditional version which characterizes convergence to the set of correlated equilibria (see Hart and Mas-Colell [16]). This is just to focus on essentials. The extension to the conditional setup is straightforward; see Section 5 below.[5]

Hannan-consistent strategies have been obtained by several authors: Hannan [15], Blackwell [4] (see also Luce and Raiffa [20, pp. 482–483]), Foster and Vohra [7, 9], Fudenberg and Levine [12], Freund and Schapire [11], and Hart and Mas-Colell [16, Section 4(c)].[6] The strategy of Fudenberg and Levine [12] (as well as those of Hannan [15], Foster and Vohra [7, 9], and Freund and Schapire [11]) is a smoothed out version of fictitious play. (We note that fictitious play—which may be stated as "at each period play an action with maximal regret"—is by itself not Hannan-consistent.) In contrast, the strategy of Hart and Mas-Colell [16, Section 4(c)], called "regret-matching," prescribes, at each period, play probabilities that are proportional to the (positive) regrets. That is, if we write $D(k)$ for the regret of i for action k at time t

[3] For these and the other topics discussed in this paper, the reader is referred also to the book of Fudenberg and Levine [13] and the survey of Foster and Vohra [10] (as well as to the other papers in the special issue of *Games and Economic Behavior* **29** [1999]).

[4] Think of this as the "regret for not having played k in the past."

[5] Note that conditional regrets have been used by Foster and Vohra [8] to prove the existence of calibrated forecasts.

[6] See also Baños [2] and Megiddo [21] and, in the computer science literature, Littlestone and Warmuth [18], Auer *et al.* [1], and the book of Borodin and El-Yaniv [5].

(as defined above) and $D_+(k)$ for the *positive regret* (i.e., $D(k)$ when $D(k) > 0$ and 0 when $D(k) \leqslant 0$), then the probability of playing action k at period $t+1$ is simply $D_+(k)/\sum_{k'} D_+(k')$.

Clearly, a general examination is called for. Smooth fictitious play and regret-matching should be but particular instances of a whole class of adaptive strategies with the Hannan-consistency property. In this paper we exhibit and characterize such a class. It turns out to contain, in particular, a large variety of new simple adaptive strategies.

In Hart and Mas-Colell [16] we have introduced Blackwell's [3] approachability theory for games with vector payoffs as the appropriate basic tool for the analysis: the vector payoff is simply the vector of regrets. In this paper, therefore, we proceed in two steps. First, in Section 2, we generalize Blackwell's result: Given an approachable set (in vector payoff space), we find the class of ("directional") strategies that guarantee that the set is approached. We defer the specifics to that section. Suffice it to say that Blackwell's strategy emerges as the particular quadratic case of a continuum of strategies where continuity and, interestingly, integrability feature decisively.

Second, in Section 3, we apply the general theory to the regret framework and derive an entire class of Hannan-consistent strategies. A feature common to them all is that, in the spirit of bounded rationality, they aim at "better" rather than "best" play. We elaborate on this aspect and carry out an explicit discussion of fictitious play in Section 4. Section 5 discusses a number of extensions, including conditional regrets and correlated equilibria.

2. THE APPROACHABILITY PROBLEM

2.1. *Model and Main Theorem*

In this section we will consider games where a player's payoffs are *vectors* (rather than, as in standard games, scalar real numbers), as introduced by Blackwell [3]. This setting may appear unnatural at first. However, it has turned out to be quite useful: the coordinates may represent different commodities or contingent payoffs in different states of the world (when there is incomplete information), or, as we will see below (in Section 3), regrets in a standard game.

Formally, we are given a game in strategic form played by a player i against an opponent $-i$ (which may be Nature and/or the other players). The *action* sets are the finite[7] sets S^i for player i and S^{-i} for $-i$. The

[7] See however Remark 2 below. Also, we always assume that S^i contains at least two elements.

payoffs are *vectors* in some Euclidean space. We denote the payoff function by[8] $A\colon S \equiv S^i \times S^{-i} \to \mathbb{R}^m$; thus $A(s^i, s^{-i}) \in \mathbb{R}^m$ is the payoff vector when i chooses s^i and $-i$ chooses s^{-i}. As usual, A is extended bilinearly to mixed actions, thus[9] $A\colon \Delta(S^i) \times \Delta(S^{-i}) \to \mathbb{R}^m$.

Let time be discrete, $t = 1, 2, ...$, and denote by $s_t = (s_t^i, s_t^{-i}) \in S^i \times S^{-i}$ the actions chosen by i and $-i$, respectively, at time t. The payoff vector in period t is $a_t := A(s_t)$, and $\bar{a}_t := (1/t) \sum_{\tau \leqslant t} a_\tau$ is the average payoff vector up to t. A *strategy*[10] *for player i* assigns to every history of play $h_{t-1} = (s_\tau)_{\tau \leqslant t-1} \in (S)^{t-1}$ a (randomized) choice of action $\sigma_t^i \equiv \sigma_t^i(h_{t-1}) \in \Delta(S^i)$ at time t, where $[\sigma_t^i(h_{t-1})](s^i)$ is, for each s^i in S^i, the probability that i plays s^i at period t following the history h_{t-1}.

Let $\mathscr{C} \subset \mathbb{R}^m$ be a convex and closed[11] set. The set $\mathscr{C}$ is *approachable* by player i (cf. Blackwell [3]; see Remark 3 below) if there is a strategy of i such that, no matter what $-i$ does,[12] $\operatorname{dist}(\bar{a}_t, \mathscr{C}) \to 0$ almost surely as $t \to \infty$. Blackwell's result can then be stated as follows.

Blackwell's Approachability Theorem. (1) *A convex and closed set $\mathscr{C}$ is approachable if and only if every half-space $\mathscr{H}$ containing $\mathscr{C}$ is approachable.*

(2) *A half-space $\mathscr{H}$ is approachable if and only if there exists a mixed action of player i such that the expected vector payoff is guaranteed to lie in $\mathscr{H}$; i.e., there is $\sigma^i \in \Delta(S^i)$ such that $A(\sigma^i, s^{-i}) \in \mathscr{H}$ for all $s^{-i} \in S^{-i}$.*

The condition for $\mathscr{C}$ to be approachable may be restated as follows (since, clearly, it suffices to consider in (1) only "minimal" half-spaces containing $\mathscr{C}$): For every $\lambda \in \mathbb{R}^m$ there exists $\sigma^i \in \Delta(S^i)$ such that

$$\lambda \cdot A(\sigma^i, s^{-i}) \leqslant w(\lambda) := \sup\{\lambda \cdot y : y \in \mathscr{C}\} \qquad \text{for all} \quad s^{-i} \in S^{-i} \tag{2.1}$$

[8] $\mathbb{R}$ is the real line and $\mathbb{R}^m$ is the m-dimensional Euclidean space. For $x = (x_k)_{k=1}^m$ and $y = (x_k)_{k=1}^m$ in $\mathbb{R}^m$, we write $x \geqslant y$ when $x_k \geqslant y_k$ for all k, and $x \gg y$ when $x_k > y_k$ for all k. The nonnegative, nonpositive, positive, and negative orthants of $\mathbb{R}^m$ are, respectively, $\mathbb{R}_+^m := \{x \in \mathbb{R}^m : x \geqslant 0\}$, $\mathbb{R}_-^m := \{x \in \mathbb{R}^m : x \leqslant 0\}$, $\mathbb{R}_{++}^m := \{x \in \mathbb{R}^m : x \gg 0\}$, and $\mathbb{R}_{--}^m := \{x \in \mathbb{R}^m : x \ll 0\}$.

[9] Given a finite set Z, we write $\Delta(Z)$ for the set of probability distributions on Z, i.e., the $(|Z| - 1)$-dimensional unit simplex $\Delta(Z) := \{p \in \mathbb{R}_+^Z : \sum_{z \in Z} p(z) = 1\}$ (the notation $|Z|$ stands for the cardinality of the set Z).

[10] We use the term "action" for a one-period choice and the term "strategy" for a multi-period choice.

[11] A set is approachable if and only if its closure is approachable; we thus assume without loss of generality that the set $\mathscr{C}$ is closed.

[12] We emphasize that the strategies of the opponents $(-i)$ are not in any way restricted; in particular, they may randomize and, furthermore, correlate their actions. All the results in this paper hold against all possible strategies of $-i$, and thus, *a fortiori*, for any specific class of strategies (like independent mixed strategies, and so on). Moreover, requiring independence over $j \neq i$ will not increase the set of strategies of i that guarantee approachability, since the worst $-i$ can do may always be taken to be pure (and thus independent).

(w is the "support function" of $\mathscr{C}$; note that only those $\lambda \neq 0$ with $w(\lambda) < \infty$ matter for (2.1)). Furthermore, the strategy constructed by Blackwell that yields approachability uses, at each step t where the current average payoff $\bar{a}_{t-1}$ is not in $\mathscr{C}$, a mixed choice σ_t^i satisfying (2.1) for that vector $\lambda \equiv \lambda(\bar{a}_{t-1})$ which goes to $\bar{a}_{t-1}$ from that point y in $\mathscr{C}$ that is closest to $\bar{a}_{t-1}$ (see Fig. 1). To get some intuition, note that the next-period expected payoff vector $b := E[a_t \mid h_{t-1}]$ lies in the half-space $\mathscr{H}$, and thus satisfies $\lambda \cdot b \leqslant w(\lambda) < \lambda \cdot \bar{a}_{t-1}$, which implies that

$$\lambda \cdot (E[\bar{a}_t \mid h_{t-1}] - \bar{a}_{t-1}) = \lambda \cdot \left(\frac{1}{t} b + \frac{t-1}{t} \bar{a}_{t-1} - \bar{a}_{t-1} \right) = \frac{1}{t} \lambda \cdot (b - \bar{a}_{t-1}) < 0.$$

Therefore the expected average payoff $E[\bar{a}_t \mid h_{t-1}]$ moves from $\bar{a}_{t-1}$ in the "general direction" of $\mathscr{C}$; in fact, it is closer than $\bar{a}_{t-1}$ to $\mathscr{C}$. Hence $E[\bar{a}_t \mid h_{t-1}]$ converges to $\mathscr{C}$, and so does the average payoff $\bar{a}_t$ (by the Law of Large Numbers).

Fix an approachable convex and closed set $\mathscr{C}$. We will now consider general strategies of player i which—like Blackwell's strategy above—are defined in terms of a *directional mapping*, that is, a function Λ: $\mathbb{R}^m \backslash \mathscr{C} \to \mathbb{R}^m$

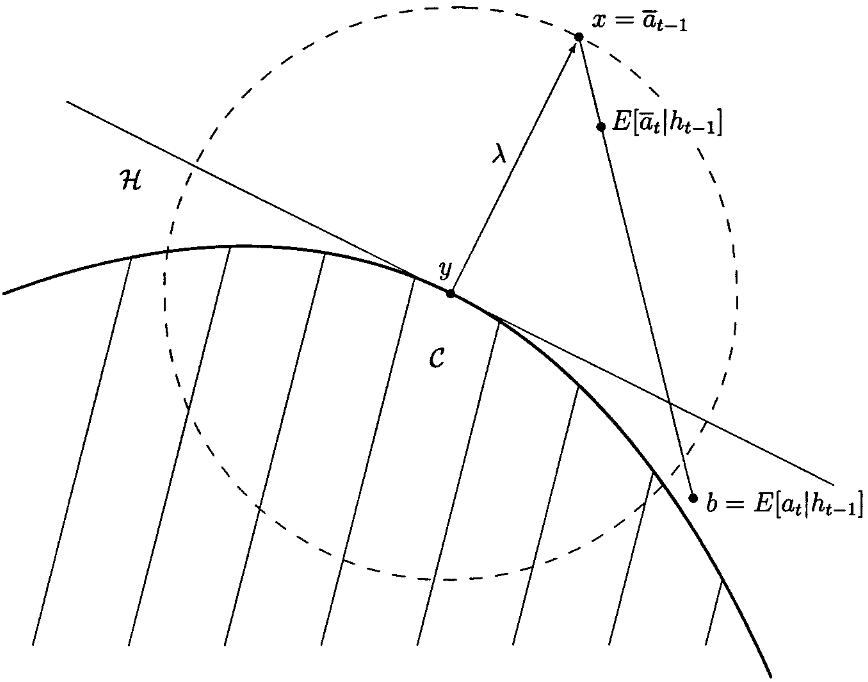

FIG. 1. Approaching the set $\mathscr{C}$ by Blackwell's strategy.

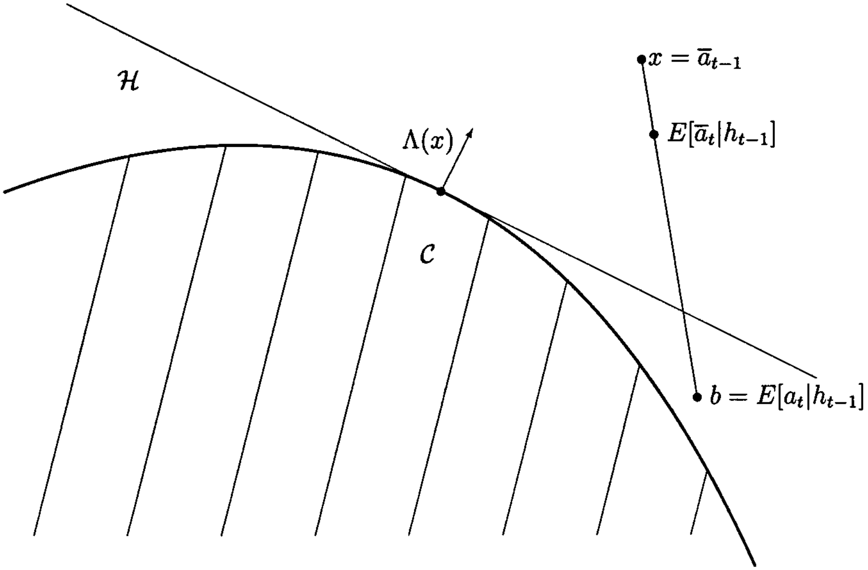

FIG. 2. A Λ-strategy.

that associates to every $x \notin \mathscr{C}$ a corresponding "direction" $\Lambda(x)$. Given such a mapping Λ, a strategy of player i is called a Λ-*strategy* if, whenever $\bar{a}_{t-1}$ does not lie in $\mathscr{C}$, it prescribes using at time t a mixed action σ_t^i that satisfies

$$\Lambda(\bar{a}_{t-1}) \cdot A(\sigma_t^i, s^{-i}) \leqslant w(\Lambda(\bar{a}_{t-1})) \qquad \text{for all} \quad s^{-i} \in S^{-i} \tag{2.2}$$

(see Fig. 2: a Λ-strategy guarantees that, when $x = \bar{a}_{t-1} \notin \mathscr{C}$, the next-period expected payoff vector $b = E[a_t \mid h_{t-1}]$ lies in the smallest half-space $\mathscr{H}$ with normal $\Lambda(x)$ that contains $\mathscr{C}$); notice that there is no requirement when $\bar{a}_{t-1} \in \mathscr{C}$. We are interested in finding conditions on the mapping Λ such that, if player i uses a Λ-strategy, then the set $\mathscr{C}$ is guaranteed to be approached, no matter what $-i$ does.

We introduce three conditions on a directional mapping Λ, relative to the given set $\mathscr{C}$.

(D1) Λ is continuous.

(D2) Λ is integrable, namely there exists a Lipschitz function[13] $P: \mathbb{R}^m \to \mathbb{R}$ such that $\nabla P(x) = \phi(x)\, \Lambda(x)$ for almost every $x \notin \mathscr{C}$, where $\phi: \mathbb{R}^m \backslash \mathscr{C} \to \mathbb{R}_{++}$ is a continuous positive function.

(D3) $\Lambda(x) \cdot x > w(\Lambda(x))$ for all $x \notin \mathscr{C}$.

[13] Note that P is defined on the whole space $\mathbb{R}^m$.

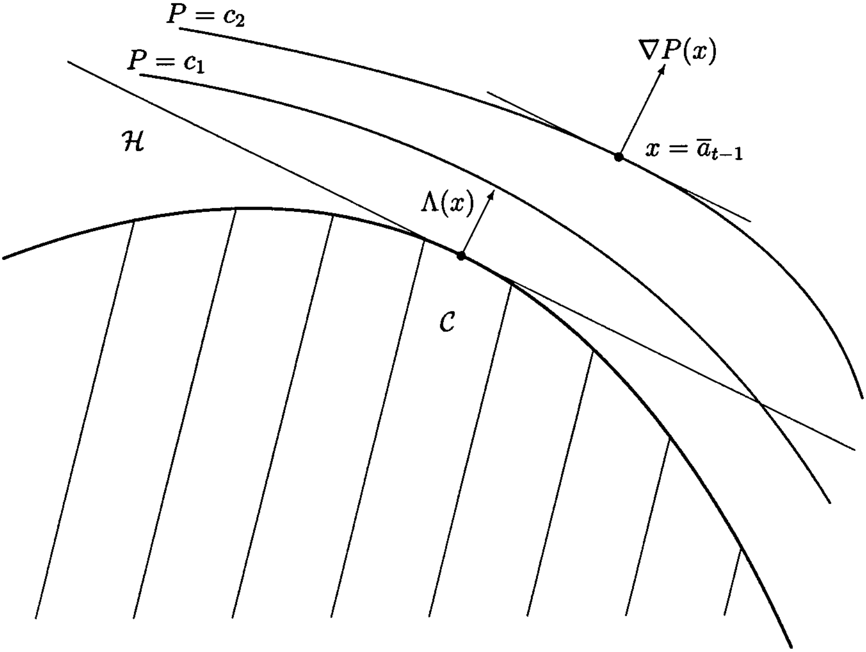

FIG. 3. The directional mapping Λ and level sets of the potential P.

See Fig. 3. The geometric meaning of (D3) is that the point x is strictly separated from the set $\mathscr{C}$ by $\Lambda(x)$. Note that (D3) implies that all λ with $w(\lambda) = \infty$, as well as $\lambda = 0$, are not allowable directions. Also, observe that the combination of (D1) and (D2) implies that P is continuously differentiable on $\mathbb{R}^m \backslash \mathscr{C}$ (see Clarke [6, Corollary to Proposition 2.2.4 and Theorem 2.5.1]). We will refer to the function P as the *potential* of Λ.

The main result of this section is

THEOREM 2.1. *Suppose that player i uses a Λ-strategy, where Λ is a directional mapping satisfying* (D1), (D2), *and* (D3) *for the approachable convex and closed set $\mathscr{C}$. Then the average payoff vector is guaranteed to approach the set $\mathscr{C}$; that is,* $\text{dist}(\bar{a}_t, \mathscr{C}) \to 0$ *almost surely as $t \to \infty$, for any strategy of $-i$.*

Before proving the theorem (in the next subsection), we state a number of comments.

Remarks. 1. The conditions (D1)–(D3) are independent of the game A (they depend on $\mathscr{C}$ only). That is, given a directional mapping Λ satisfying

(D1)–(D3), a Λ-strategy is guaranteed to approach $\mathscr{C}$ for *any* game A for which $\mathscr{C}$ is approachable (of course, the specific choice of action depends on A, according to (2.2)). It is in this sense that we refer to the Λ-strategies as "universal."

2. The action sets S^i and S^{-i} need not be finite; as we will see in the proof, it suffices for the range of A to be bounded.

3. As in Blackwell's result, our proof below yields *uniform* approachability: For every ε there is $t_0 \equiv t_0(\varepsilon)$ such that $E[\text{dist}(\bar{a}_t, \mathscr{C})] < \varepsilon$ for all $t > t_0$ and all strategies of $-i$ (i.e., t_0 is independent of the strategy of $-i$).

4. The conditions on P are invariant under strictly increasing monotone transformations (with positive derivative); that is, only the level sets of P matter.

5. If the potential P is a convex function and $\mathscr{C} = \{y : P(y) \leqslant c\}$ for some constant c, then (D3) is automatically satisfied: $P(x) > P(y)$ implies $\nabla P(x) \cdot x > \nabla P(x) \cdot y$.

6. Given a norm $\|\cdot\|$ on $\mathbb{R}^m$, consider the resulting "distance from $\mathscr{C}$" function $P(x) := \min_{y \in \mathscr{C}} \|x - y\|$. If P is a smooth function (which is always the case when either the norm is smooth—i.e., the corresponding unit ball has smooth boundary—or when the boundary of $\mathscr{C}$ is smooth), then the mapping $\Lambda = \nabla P$ satisfies (D1)–(D3) (the latter by the previous Remark 5). In particular, the Euclidean l_2-norm yields precisely the Blackwell strategy, since then $\nabla P(x)$ is proportional to $x - y(x)$, where $y(x) \in \mathscr{C}$ is the point in $\mathscr{C}$ closest to x. The l_p-norm is smooth for $1 < p < \infty$; therefore it yields strategies that guarantee approachability for *any* approachable set $\mathscr{C}$. However, if the boundary of $\mathscr{C}$ is not smooth—for instance, when $\mathscr{C}$ is an orthant, an important case in applications—then (D1) is *not* satisfied in the extreme cases $p = 1$ and $p = \infty$ (see Fig. 4; more on these two cases below).

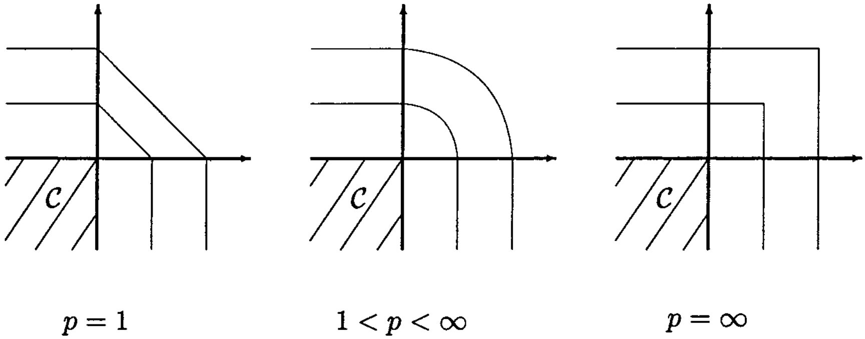

FIG. 4. The l_p-potential for an orthant $\mathscr{C}$.

7. When $\mathscr{C} = \mathbb{R}^m_-$ and P is given by (D2), condition (D3) becomes $\nabla P(x) \cdot x > 0$ for every $x \notin \mathscr{C}$, which means that P is increasing along any ray from the origin that goes outside the negative orthant.

2.2. *Proof of Theorem* 2.1

We begin by proving two auxiliary results. The first applies to functions Q that satisfy conditions similar to but stronger than (D1)–(D3); the second allows us to reduce the general case to such a Q. The set $\mathscr{C}$, the mappings Λ and P, and the strategy of i (which is a Λ-strategy) are fixed throughout. Also, let K be a convex and compact set containing in its interior the range of A (recall that S is finite).

Lemma 2.2. *Let Q: $\mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function that satisfies*

(i) $Q(x) \geqslant 0$ *for all* x;

(ii) $Q(x) = 0$ *for all* $x \in \mathscr{C}$;

(iii) $\nabla Q(x) \cdot x - w(\nabla Q(x)) \geqslant Q(x)$ *for all* $x \in K \backslash \mathscr{C}$; *and*

(iv) $\nabla Q(x)$ *is nonnegatively proportional to* $\Lambda(x)$ (*i.e.*, $\nabla Q(x) = \phi(x)$ $\Lambda(x)$ *where* $\phi(x) \geqslant 0$) *for all* $x \notin \mathscr{C}$.

Then $\lim_{t \to \infty} Q(\bar{a}_t) = 0$ *a.s. for any strategy of* $-i$.

Proof. We have $\bar{a}_t - \bar{a}_{t-1} = (1/t)(a_t - \bar{a}_{t-1})$; thus, writing x for $\bar{a}_{t-1}$,

$$Q(\bar{a}_t) = Q(x) + \nabla Q(x) \cdot \frac{1}{t}(a_t - x) + o\left(\frac{1}{t}\right), \tag{2.3}$$

since Q is (continuously) differentiable. Moreover, the remainder $o(1/t)$ is uniform, since all relevant points lie in the compact set K. If $x \notin \mathscr{C}$ then player i plays at time t so that

$$\nabla Q(x) \cdot E[a_t \mid h_{t-1}] \leqslant w(\nabla Q(x)) \tag{2.4}$$

(by (2.2) and (iv)); if $x \in \mathscr{C}$ then $\nabla Q(x) = 0$ (by (i) and (ii)), and (2.4) holds too. Taking conditional expectation in (2.3) and then substituting (2.4) yields

$$\begin{aligned} E[Q(\bar{a}_t) \mid h_{t-1}] &\leqslant Q(x) + \frac{1}{t} w(\nabla Q(x)) - \nabla Q(x) \cdot x) + o\left(\frac{1}{t}\right) \\ &\leqslant Q(x) - \frac{1}{t} Q(x) + o\left(\frac{1}{t}\right), \end{aligned}$$

where we have used (iii) when $x \notin \mathscr{C}$ and (i), (ii) when $x \in \mathscr{C}$. Thus

$$E[Q(\bar{a}_t) \mid h_{t-1}] \leqslant \frac{t-1}{t} Q(\bar{a}_{t-1}) + o\left(\frac{1}{t}\right).$$

This may be rewritten as[14]

$$E[\zeta_t \mid h_{t-1}] \leqslant o(1), \tag{2.5}$$

where $\zeta_t := tQ(\bar{a}_t) - (t-1)\, Q(\bar{a}_{t-1})$. Hence $\limsup_{t\to\infty} (1/t) \sum_{\tau \leqslant t} E[\zeta_\tau \mid h_{\tau-1}] \leqslant 0$. The Strong Law of Large Numbers for Dependent Random Variables (see Loève [19, Theorem 32.1.E]) implies that $(1/t) \sum_{\tau \leqslant t} (\zeta_\tau - E[\zeta_\tau \mid h_{\tau-1}]) \to 0$ a.s. as $t \to \infty$ (note that the ζ_t's are uniformly bounded, as can be immediately seen from (2.3), $\zeta_t = Q(\bar{a}_{t-1}) + \nabla Q(\bar{a}_{t-1}) \cdot (a_t - \bar{a}_{t-1}) + o(1)$, and from the fact that everything happens in the compact set K). Therefore $\limsup_{t\to\infty} (1/t) \sum_{\tau \leqslant t} \zeta_\tau \leqslant 0$. But $0 \leqslant Q(\bar{a}_t) = (1/t) \sum_{\tau \leqslant t} \zeta_\tau$, so $\lim_{t\to\infty} Q(\bar{a}_t) = 0$. ▮

LEMMA 2.3. *The function P satisfies*:

(c1) *If the boundary of $\mathscr{C}$ is connected, then there exists a constant c such that*

$$\begin{aligned} P(x) &= c, && \textit{if} \quad x \in \operatorname{bd} \mathscr{C}; \\ P(x) &> c, && \textit{if} \quad x \notin \mathscr{C}. \end{aligned}$$

(c2) *If the boundary of $\mathscr{C}$ is not connected, then there exists a $\lambda \in \mathbb{R}^m \backslash \{0\}$ such that*[15] $\mathscr{C} = \{x \in \mathbb{R}^m : -w(-\lambda) \leqslant \lambda \cdot x \leqslant w(\lambda)\}$ (*where $w(\lambda) < \infty$ and $w(-\lambda) < \infty$*), *and there are constants c_1 and c_2 such that*

$$\begin{aligned} P(x) &= c_1, && \textit{if} \quad x \in \operatorname{bd} \mathscr{C} && \textit{and} && \lambda \cdot x = w(\lambda); \\ P(x) &= c_2, && \textit{if} \quad x \in \operatorname{bd} \mathscr{C} && \textit{and} && (-\lambda) \cdot x = w(-\lambda); \\ P(x) &> c_1, && \textit{if} \quad x \notin \mathscr{C} && \textit{and} && \lambda \cdot x > w(\lambda); \\ P(x) &> c_2, && \textit{if} \quad x \notin \mathscr{C} && \textit{and} && (-\lambda) \cdot x > w(-\lambda). \end{aligned}$$

Proof. Let $x_0, x_1 \in \operatorname{bd} \mathscr{C}$ and denote by λ_j, for $j = 0, 1$, an outward unit normal to $\mathscr{C}$ at x_j; thus $\|\lambda_j\| = 1$ and $\lambda_j \cdot x_j = w(\lambda_j)$.

[14] Recall that the remainder term $o(1/t)$ was uniform; that is, for every $\varepsilon > 0$ there is $t_0(\varepsilon)$ such that $o(1) < \varepsilon$ is guaranteed for all $t > t_0(\varepsilon)$.

[15] This is a general fact about convex sets: The only case where the boundary of a convex closed set $\mathscr{C} \subset \mathbb{R}^m$ is not path-connected is when $\mathscr{C}$ is the set of points lying between two parallel hyperplanes. We prove this in Steps 1–3 below, independently of the function P.

Step 1. If $\lambda_1 \neq -\lambda_0$, we claim that there is a path on $\mathrm{bd}\,\mathscr{C}$ connecting x_0 and x_1, and moreover that $P(x_0)=P(x_1)$. Indeed, there exists a vector[16] $z \in \mathbb{R}^m$ such that $\lambda_0 \cdot z > 0$ and $\lambda_1 \cdot z > 0$. The straight line segment connecting x_0 and x_1 lies in $\mathscr{C}$; we move it in the direction z until it reaches the boundary of $\mathscr{C}$. That is, for each $\eta \in [0,1]$, let $y(\eta) := \eta x_1 + (1-\eta)\, x_0 + \alpha(\eta)\, z$, where $\alpha(\eta) := \max\{\beta : \eta x_1 + (1-\eta)\, x_0 + \beta z \in \mathscr{C}\}$; this maximum exists by the choice of z. Note that $y(\cdot)$ is a path on $\mathrm{bd}\,\mathscr{C}$ connecting x_0 and x_1.

It is easy to verify that $\alpha(0)=\alpha(1)=0$ and that $\alpha\colon [0,1] \to \mathbb{R}_+$ is a concave function, and thus differentiable a.e. For each $k=1,2,...,$ define $y_k(\eta) := y(\eta) + (1/k)\, z$; then $y_k(\cdot)$ is a path in $\mathbb{R}^m \backslash \mathscr{C}$, the region where P is continuously differentiable. Let $\bar{\eta} \in (0,1)$ be a point of differentiability of $\alpha(\cdot)$, thus also of $y(\cdot)$, $y_k(\cdot)$, and $P(y_k(\cdot))$; we have $dP(y_k(\bar{\eta}))/d\eta = \nabla P(y_k(\bar{\eta})) \cdot y_k'(\bar{\eta}) = \nabla P(y_k(\bar{\eta})) \cdot y'(\bar{\eta})$. By (D3), $\nabla P(y_k(\bar{\eta})) \cdot y_k(\bar{\eta}) > w(\nabla P(y_k(\bar{\eta}))) \geqslant \nabla P(y_k(\bar{\eta})) \cdot y(\eta)$ for any $\eta \in [0,1]$ (the second inequality since $y(\eta) \in \mathscr{C}$). Thus, for any accumulation point q of the bounded[17] sequence $(\nabla P(y_k(\bar{\eta})))_{k=1}^{\infty}$ we get $q \cdot y(\bar{\eta}) \geqslant q \cdot y(\eta)$ for all $\eta \in [0,1]$. Therefore $q \cdot y(\eta)$ is maximized at $\eta = \bar{\eta}$, which implies that $q \cdot y'(\bar{\eta}) = 0$. This holds for *any* accumulation point q, hence $\lim_{k \to \infty} dP(y_k(\bar{\eta}))/d\eta = 0$ for almost every $\bar{\eta}$. Therefore

$$P(x_1) - P(x_0) = P(y(1)) - P(y(0)) = \lim_k \left[P(y_k(1)) - P(y_k(0))\right]$$

$$= \lim_k \int_0^1 \frac{dP(y_k(\eta))}{d\eta}\, d\eta = \int_0^1 \lim_k \frac{dP(y_k(\eta))}{d\eta}\, d\eta = 0$$

(again, P is Lipschitz, so $dP(y_k(\eta))/d\eta$ are uniformly bounded).

Step 2. If $\lambda_1 = -\lambda_0$ and there is another boundary point x_2 with outward unit normal λ_2 different from both $-\lambda_0$ and $-\lambda_1$, then we get paths on $\mathrm{bd}\,\mathscr{C}$ connecting x_0 to x_2 and x_1 to x_2, and also $P(x_0)=P(x_2)$ and $P(x_1)=P(x_2)$—thus we get the same conclusion as in Step 1.

Step 3. If $\lambda_1 = -\lambda_0$ and no x_2 and λ_2 as in Step 2 exist, it follows that the unit normal to every point on the boundary of $\mathscr{C}$ is either λ_0 or $-\lambda_0$; thus $\mathscr{C}$ is the set bounded between the two parallel hyperplanes $\lambda_0 \cdot x = w(\lambda_0)$ and $-\lambda_0 \cdot x = w(-\lambda_0)$. In particular, the boundary of $\mathscr{C}$ is not connected, and we are in Case (c2). Note that in this case when x_0 and x_1 lie on the same hyperplane then $P(x_0)=P(x_1)$ by Step 1 (since $\lambda_1 = \lambda_0 \neq -\lambda_0$).

[16] Take for instance $z = \lambda_0 + \lambda_1$.

[17] Recall that P is Lipschitz.

Step 4. If it is case (c1)—thus not (c2)—then the situation of Step 3 is not possible; thus for any two boundary points x_0 and x_1 we get $P(x_0) = P(x_1)$ by either Step 1 or Step 2.

Step 5. Given $x \notin \mathscr{C}$, let $x_0 \in \operatorname{bd} \mathscr{C}$ be the point in $\mathscr{C}$ that is closest to x. Then the line segment from x to x_0 lies outside $\mathscr{C}$, i.e., $y(\eta) := \eta x + (1-\eta)\, x_0 \notin \mathscr{C}$ for all $\eta \in (0, 1]$. By (D3) and $x_0 \in \mathscr{C}$, it follows that $\nabla P(y(\eta)) \cdot y(\eta) > w(\nabla P(y(\eta))) \geqslant \nabla P(y(\eta)) \cdot x_0$, or, after dividing by $\eta > 0$, that $\nabla P(y(\eta)) \cdot (x - x_0) > 0$, for all $\eta \in (0, 1]$. Hence $P(x) - P(x_0) = \int_0^1 \nabla P(y(\eta)) \cdot y'(\eta)\, d\eta = \int_0^1 \nabla P(y(\eta)) \cdot (x - x_0)\, d\eta > 0$, showing that $P(x) > c$ in Case (c1) and $P(x) > c_1$ or $P(x) > c_2$ in Case (c2). ▮

We can now prove the main result of this section.

Proof of Theorem 2.1. First, use Lemma 2.3 to replace P by P_1 as follows: When the boundary of $\mathscr{C}$ is connected (Case (c1)), define $P_1(x) := (P(x) - c)^2$ for $x \notin \mathscr{C}$ and $P_1(x) := 0$ for $x \in \mathscr{C}$; when the boundary of $\mathscr{C}$ is not connected (Case (c2)), define $P_1(x) := (P(x) - c_1)^2$ for $x \notin \mathscr{C}$ with $\lambda \cdot x > w(\lambda)$, $P_1(x) := (P(x) - c_2)^2$ for $x \notin \mathscr{C}$ with $(-\lambda) \cdot x > w(-\lambda)$, and $P_1(x) := 0$ for $x \in \mathscr{C}$. It is easy to verify: P_1 is continuously differentiable; $\nabla P_1(x)$ is positively proportional to $\nabla P(x)$ and thus to $\Lambda(x)$ for $x \notin \mathscr{C}$; $P_1(x) \geqslant 0$ for all x; and $P_1(x) = 0$ if and only if $x \in \mathscr{C}$.

Given $\varepsilon > 0$, let $k \geqslant 2$ be a large enough integer such that

$$\frac{\nabla P_1(x) \cdot x - w(\nabla P_1(x))}{P_1(x)} \geqslant \frac{1}{k} \tag{2.6}$$

for all x in the compact set $K \cap \{x : P_1(x) \geqslant \varepsilon\}$ (the minimum of the above ratio is attained and it is positive by (D3)). Put[18] $Q(x) := ([P_1(x) - \varepsilon]_+)^k$. Then Q is continuously differentiable (since $k \geqslant 2$) and it satisfies all the conditions of Lemma 2.2. To check (iii): When $Q(x) = 0$ we have $\nabla Q(x) = 0$, and when $Q(x) > 0$ we have

$$\begin{aligned} \nabla Q(x) \cdot x - w(\nabla Q(x)) &= k(P_1(x) - \varepsilon)^{k-1}\, [\nabla P_1(x) \cdot x - w(\nabla P_1(x))] \\ &\geqslant (P_1(x) - \varepsilon)^{k-1}\, P_1(x) \\ &\geqslant Q(x) \end{aligned}$$

(the first inequality follows from (2.6)).

By Lemma 2.2, it follows that the Λ-strategy guarantees a.s. $\lim_{t \to \infty} Q(\bar{a}_t) = 0$, or $\limsup_{t \to \infty} P_1(\bar{a}_t) \leqslant \varepsilon$. Since $\varepsilon > 0$ is arbitrary, this yields a.s. $\lim_{t \to \infty} P_1(\bar{a}_t) = 0$ or $\bar{a}_t \to \mathscr{C}$. ▮

[18] We write $[z]_+$ for the positive part of z, i.e., $[z]_+ := \max\{z, 0\}$.

Remark. P may be viewed (up to a constant, as in the definition of P_1 above) as a generalized distance to the set $\mathscr{C}$ (compare with Remark 6 in Subsection 2.1).

2.3. *Counterexamples*

In this subsection we provide counterexamples showing the indispensability of the conditions (D1)–(D3) for the validity of Theorem 2.1. The first two examples refer to (D1), the third to (D2), and the last one to (D3).

EXAMPLE 2.4. *The role of (D1):* Consider the following two-dimensional vector payoff matrix A

	C1	C2
R1	$(0, -1)$	$(0, 1)$
R2	$(1, 0)$	$(-1, 0)$

Let i be the Row player and $-i$ the Column player. The set $\mathscr{C} := \mathbb{R}^2_-$ is approachable by the Row player since $w(\lambda) < \infty$ whenever $\lambda \geqslant 0$, and then the mixed action $\sigma^{\text{Row}}(\lambda) := (\lambda_1/(\lambda_1+\lambda_2), \lambda_2/(\lambda_1+\lambda_2))$ of the Row player yields $\lambda \cdot A(\sigma^{\text{Row}}(\lambda), \gamma) = 0 = w(\lambda)$ for any action γ of the Column player.

We define a directional mapping Λ_∞ on $\mathbb{R}^2 \backslash \mathbb{R}^2_-$,

$$\Lambda_\infty(x) := \begin{cases} (1, 0), & \text{if } \; x_1 > x_2; \\ (0, 1), & \text{if } \; x_1 \leqslant x_2. \end{cases}$$

Clearly Λ_∞ is *not continuous*, i.e., it does not satisfy (D1); it does however satisfy (D3) and (D2) (with $P(x) = \max\{x_1, x_2\}$, the l_∞-potential; see Remark 6 in Subsection 2.1). Consider a Λ_∞-strategy for the Row player that, when $x := \bar{a}_{t-1} \notin \mathscr{C}$, plays $\sigma^{\text{Row}}(\Lambda_\infty(x))$ at time t; that is, he plays R1 when $x_1 > x_2$, and R2 when $x_1 \leqslant x_2$. Assume that the Column player plays[19] C2 when $x_1 > x_2$ and C1 when $x_1 \leqslant x_2$. Then, starting with, say, $a_1 = (0, 1) \notin \mathscr{C}$, the vector payoff a_t will always be either $(0, 1)$ or $(1, 0)$, thus on the line $x_1 + x_2 = 1$, so the average $\bar{a}_t$ does not converge to $\mathscr{C} = \mathbb{R}^2_-$.

[19] In order to show that the strategy of the Row player does not guarantee approachability to $\mathscr{C}$, we exhibit one strategy of the Column player for which $\bar{a}_t$ does not converge to $\mathscr{C}$.

EXAMPLE 2.5. *The role of (D1), again:* The same as in Example 2.4, but now the directional mapping is Λ_1, defined on $\mathbb{R}^2 \backslash \mathbb{R}^2_-$ by

$$\Lambda_1(x) := \begin{cases} (1, 1), & \text{if} \quad x_1 > 0 \quad \text{and} \quad x_2 > 0; \\ (1, 0), & \text{if} \quad x_1 > 0 \quad \text{and} \quad x_2 \leqslant 0; \\ (0, 1), & \text{if} \quad x_1 \leqslant 0 \quad \text{and} \quad x_2 > 0. \end{cases}$$

Again, the mapping Λ_1 is *not continuous*—it does not satisfy (D1)—but it satisfies (D3) and (D2) (with $P(x) := [x_1]_+ + [x_2]_+$, the l_1-potential). Consider a Λ_1-strategy for the Row player where at time t he plays $\sigma^{\mathrm{Row}}(\Lambda_1(x))$ when $x := \bar{a}_{t-1} \notin \mathscr{C}$, and assume that the Column player plays C1 when $x_1 \leqslant 0$ and $x_2 > 0$, and plays C2 otherwise. Thus, if $x \notin \mathscr{C}$ then a_t is

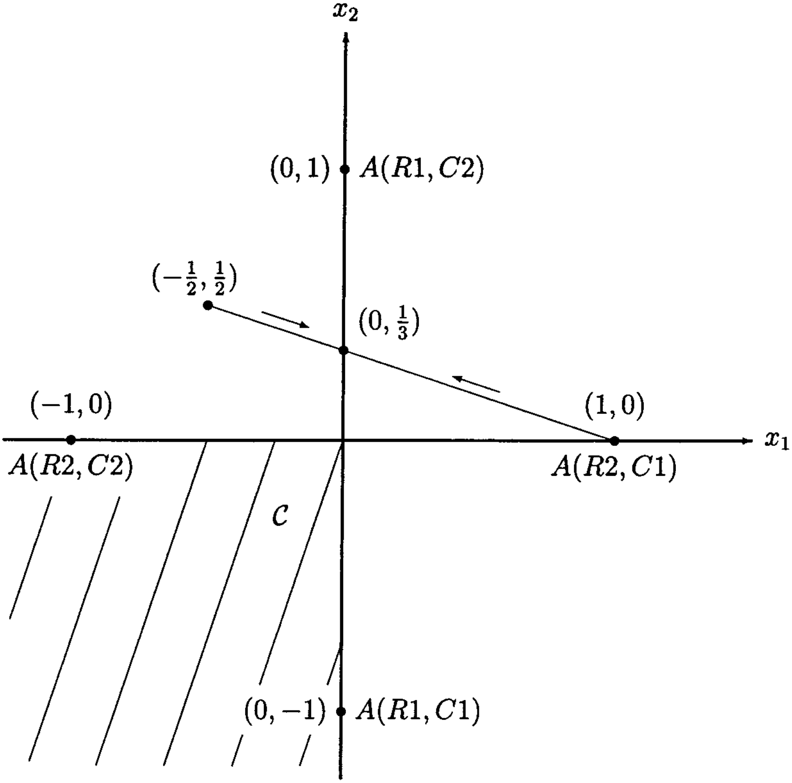

FIG. 5. The deterministic dynamic in Example 2.4.

- $(0, 1)$ or $(-1, 0)$ with equal probabilities, when $x_1 > 0$ and $x_2 > 0$;
- $(0, 1)$, when $x_1 > 0$ and $x_2 \leqslant 0$;
- $(1, 0)$, when $x_1 \leqslant 0$ and $x_2 > 0$.

In all cases the second coordinate of a_t is nonnegative; therefore, if we start with, say, $a_1 = (0, 1) \notin \mathscr{C}$, then, inductively, the second coordinate of $\bar{a}_{t-1}$ will be strictly positive, so that $\bar{a}_{t-1} \notin \mathscr{C}$ for all t. But then $E[a_t \mid h_{t-1}] \in \mathscr{D} := \text{conv}\{(-1/2, 1/2), (0, 1), (1, 0)\}$, and $\mathscr{D}$ is disjoint from $\mathscr{C}$ and at a positive distance from it. Therefore $(1/t) \sum_{\tau \leqslant t} E[a_\tau \mid h_{\tau-1}] \in \mathscr{D}$ and so, by the Strong Law of Large Numbers, $\lim \bar{a}_t = \lim (1/t) \sum_{\tau \leqslant t} a_\tau \in \mathscr{D}$ too (a.s.), so $\bar{a}_t$ does not approach[20] $\mathscr{C}$.

To get some intuition, consider the deterministic system where a_t is replaced by $E[a_t \mid h_{t-1}]$. Then the point $(0, 1/3)$ is a stationary point for this dynamic. Specifically (see Fig. 5), if $\bar{a}_{t-1}$ is on the line segment joining $(-1/2, 1/2)$ with $(1, 0)$, then $E[\bar{a}_t \mid h_{t-1}]$ will be there too, moving toward $(-1/2, 1/2)$ when $\bar{a}_{t-1}$ is in the positive orthant and toward $(1, 0)$ when it is in the second orthant.

EXAMPLE 2.6. *The role of (D2):* Consider the following two-dimensional vector payoff matrix A

	C1	C2	C3	C4
R1	$(0, 1)$	$(0, 0)$	$(0, -1)$	$(0, 0)$
R2	$(-1, 0)$	$(0, 0)$	$(1, 0)$	$(0, 0)$
R3	$(0, 0)$	$(0, -1)$	$(0, 0)$	$(0, 1)$
R4	$(0, 0)$	$(-1, 0)$	$(0, 0)$	$(1, 0)$

.

Again, the Row player is i and the Column player is $-i$. Let $\mathscr{C} := \{(0, 0)\}$. For every $\lambda \in \mathbb{R}^2 \backslash \{(0, 0)\}$, put $\mu_1 := |\lambda_1|/(|\lambda_1| + |\lambda_2|)$ and $\mu_2 := |\lambda_2|/(|\lambda_1| + |\lambda_2|)$ and define a mixed action $\sigma^{\text{Row}}(\lambda)$ for the Row player and a pure action $c(\lambda)$ for the Column player, as follows:

- If $\lambda_1 \geqslant 0$ and $\lambda_2 \geqslant 0$ then $\sigma^{\text{Row}}(\lambda) := (\mu_1, \mu_2, 0, 0)$ and $c(\lambda) := \text{C1}$.
- If $\lambda_1 < 0$ and $\lambda_2 \geqslant 0$ then $\sigma^{\text{Row}}(\lambda) := (0, 0, \mu_1, \mu_2)$ and $c(\lambda) := \text{C2}$.

[20] One way to see this formally is by a separation argument: Let $f(x) := x_1 + 3x_2$; then $E[f(a_t) \mid h_{t-1}] \geqslant 1$, so $\liminf f(\bar{a}_t) = \liminf (1/t) \sum_{\tau \leqslant t} f(a_t) \geqslant 1$, whereas $f(x) \leqslant 0$ for all $x \in \mathscr{C}$.

- If $\lambda_1 < 0$ and $\lambda_2 < 0$ then $\sigma^{\mathrm{Row}}(\lambda) := (\mu_1, \mu_2, 0, 0)$ and $c(\lambda) := \mathrm{C3}$.
- If $\lambda_1 \geqslant 0$ and $\lambda_2 < 0$ then $\sigma^{\mathrm{Row}}(\lambda) := (0, 0, \mu_1, \mu_2)$ and $c(\lambda) := \mathrm{C4}$.

It is easy to verify that in all four cases:

(a1) $\lambda \cdot A(\sigma^{\mathrm{Row}}(\lambda), \gamma) \leqslant 0 = w(\lambda)$ for any action γ of the column player; and

(a2) $A(\sigma^{\mathrm{Row}}(\lambda), c(\lambda)) = (|\lambda_1| + |\lambda_2|)^{-1}\, \hat{\lambda}$, where $\hat{\lambda} := (-\lambda_2, \lambda_1)$.

Condition (a1) implies, by (2.1), that $\mathscr{C}$ is approachable by the Row player; and Condition (a2) means that $A(\sigma^{\mathrm{Row}}(\lambda), c(\lambda))$ is 90° counterclockwise from λ.

Consider now the directional mapping Λ given by $\Lambda(x) := (x_1 + \alpha x_2, x_2 - \alpha x_1)$, where $\alpha > 0$ is a fixed constant.[21] Then (D1) and (D3) hold (for the latter, we have $x \cdot \Lambda(x) = (x_1)^2 + (x_2)^2 > 0 = w(\Lambda(x))$ for all $x \notin \mathscr{C}$), but *integrability* (D2) *is not satisfied*. To see this, assume that $\nabla P(x) = \phi(x)\, \Lambda(x)$ for all $x \notin \mathscr{C}$, where $\phi(x) > 0$ is a continuous function. Consider the path $y(\eta) := (\sin \eta, \cos \eta)$ for $\eta \in [0, 2\pi]$. We have $0 = P(y(2\pi)) - P(y(0)) = \int_0^{2\pi} (dP(y(\eta))/d\eta)\, d\eta = \int_0^{2\pi} \nabla P(y(\eta)) \cdot y'(\eta)\, d\eta$. But the integrand equals $(\phi(y(\eta)))\, (\sin \eta + \alpha \cos \eta, \cos \eta - \alpha \sin \eta) \cdot (\cos \eta, -\sin \eta) = \alpha \phi(y(\eta))$ which is everywhere positive, a contradiction.

We claim that if the Row player uses the Λ-strategy where at time t he plays[22] $\sigma^{\mathrm{Row}}(\Lambda(\bar{a}_{t-1}))$, and if the Column player chooses $c(\Lambda(\bar{a}_{t-1}))$, then the distance to the set $\mathscr{C} = \{(0, 0)\}$ does not approach 0. Indeed, let $b_t := E[a_t \mid h_{t-1}]$; then the strategies played imply that the vector $b_t = A(\sigma^{\mathrm{Row}}(\Lambda(\bar{a}_{t-1})), c(\Lambda(\bar{a}_{t-1})))$ is perpendicular to $\Lambda(\bar{a}_{t-1})$ and makes an acute angle with the vector $\bar{a}_{t-1}$. Specifically,[23] $b_t \cdot \bar{a}_{t-1} = \beta\, \|b_t\|\, \|\bar{a}_{t-1}\|$, where $\beta := \alpha/\sqrt{1+\alpha^2} \leqslant 1$. Therefore

$$\begin{aligned}(E[t\, \|\bar{a}_t\| \mid h_{t-1}])^2 &\geqslant \|E[t\bar{a}_t \mid h_{t-1}]\|^2 \\ &= (t-1)^2\, \|\bar{a}_{t-1}\|^2 + 2(t-1)\, b_t \cdot \bar{a}_{t-1} + \|b_t\|^2 \\ &\geqslant (t-1)^2\, \|\bar{a}_{t-1}\|^2 + 2(t-1)\, \beta\, \|b_t\|\, \|\bar{a}_{t-1}\| + \beta^2\, \|b_t\|^2 \\ &= ((t-1)\, \|\bar{a}_{t-1}\| + \beta\, \|b_t\|)^2.\end{aligned}$$

[21] This will provide a counterexample for any value of α, showing that it is not an isolated phenomenon.

[22] For all λ except those on the two axes (i.e., $\lambda_1 = 0$ or $\lambda_2 = 0$), the mixed action $\sigma^{\mathrm{Row}}(\lambda)$ is uniquely determined by (2.2). If the Row player were to play in these exceptional cases another mixed action satisfying (2.2), it is easy to verify that the Column player could respond appropriately so that $\mathscr{C}$ is not approached. Thus, no Λ-strategy guarantees approachability.

[23] Writing b for b_t; x for $\bar{a}_{t-1}$; and λ for $\Lambda(\bar{a}_{t-1}) = \Lambda(x)$, we have $x = (1+\alpha^2)^{-1}\, (\lambda + \alpha\hat{\lambda})$ (recall the definition of Λ and invert), thus $b \cdot x = (|\lambda_1| + |\lambda_2|)^{-1}\, \hat{\lambda} \cdot (1+\alpha^2)^{-1}\, (\lambda + \alpha\hat{\lambda}) = \alpha(1+\alpha^2)^{-1}\, (|\lambda_1| + |\lambda_2|)^{-1}\, \|\lambda\|^2$ (we have used (a2), $\lambda \cdot \hat{\lambda} = 0$, and $\|\lambda\| = \|\hat{\lambda}\|$). Now $\|b\| = (|\lambda_1| + |\lambda_2|)^{-1}\, \|\lambda\|$ and $\|x\| = (1+\alpha^2)^{-1/2}\, \|\lambda\|$, thus completing the proof.

Now $\|b_t\| \geqslant 1/\sqrt{2}$ (for instance, see (a2)), and we have thus obtained

$$E[t\ \|\bar{a}_t\| - (t-1)\ \|\bar{a}_{t-1}\| \mid h_{t-1}] \geqslant \beta/\sqrt{2} > 0,$$

from which it follows that $\liminf \|\bar{a}_t\| = \liminf (1/t) \sum_{\tau \leqslant t} [\tau\ \|\bar{a}_\tau\| - (\tau - 1)\ \|\bar{a}_{\tau-1}\|] \geqslant \beta/\sqrt{2} > 0$ a.s., again by the Strong Law of Large Numbers. Thus the distance of the average payoff vector $\bar{a}_t$ from the set $\mathscr{C} = \{(0, 0)\}$ is, with probability one, bounded away from 0 from some time on.

To get some intuition for this result, note that direction of movement from $\bar{a}_{t-1}$ to $E[\bar{a}_t \mid h_{t-1}]$ is at a fixed angle $\theta \in (0, \pi/2)$ from $\bar{a}_{t-1}$, which, if the dynamic were deterministic, would generate a counterclockwise spiral that goes away from $(0, 0)$.

EXAMPLE 2.7. *The role of (D3):* Consider the two-dimensional vector payoff matrix A

R1	(0, 1)
R2	(−1, 0)

(where i is the Row player and $-i$ has one action). The set $\mathscr{C} := \mathbb{R}^2_-$ is approachable by the Row player (by playing "R2 forever"). Consider the directional mapping Λ defined on $\mathbb{R}^2 \backslash \mathbb{R}^2_-$ by $\Lambda(x) := (1, 0)$. Then (D1) and (D2) are satisfied (with $P(x) := x_1$), but (D3) is not: $\Lambda(0, 1) \cdot (0, 1) = 0 = w(\Lambda(0, 1))$. Playing "R1 forever" is a Λ-strategy, but the payoff is $(0, 1) \notin \mathscr{C}$.

3. REGRETS

3.1. *Model and Preliminary Results*

In this section we consider standard N-person games in strategic form (with *scalar* payoffs for each player). The set of players is a finite set N, the action set of each player i is a finite set S^i, and the payoff function of i is u^i: $S \to \mathbb{R}$, where $S := \prod_{j \in N} S^j$; we will denote this game $\langle N, (S^i)_i, (u^i)_i \rangle$ by Γ.

As in the previous section, the game is played repeatedly in discrete time $t = 1, 2, \ldots$; denote by $s^i_t \in S^i$ the choice of player i at time t, and put $s_t = (s^i_t)_{i \in N} \in S$. The payoff of i in period t is $U^i_t := u^i(s_t)$, and $\bar{U}^i_t := (1/t) \sum_{\tau \leqslant t} U^i_\tau$ is his average payoff up to t.

Fix a player $i \in N$. Following Hannan [15], we consider the *regrets* of player i; namely, for each one of his actions $k \in S^i$, the change in his average payoff if he were always to choose k (while no one else makes any change in his realized actions):

$$D_t^i(k) := \frac{1}{t} \sum_{\tau=1}^{t} u^i(k, s_\tau^{-i}) - \bar{U}_t^i = u^i(k, z_t^{-i}) - \bar{U}_t^i,$$

where $z_t^{-i} \in \Delta(S^{-i})$ is the empirical distribution of the actions chosen by the other players in the past.[24] A strategy of player i is called *Hannan-consistent* if, as t increases, all regrets are guaranteed—no matter what the other players do—to become almost surely nonpositive in the limit; that is, with probability one, $\limsup_{t \to \infty} D_t^i(k) \leqslant 0$ for all $k \in S^i$.

Following Hart and Mas-Colell [16], it is useful to view the regrets of i as an m-dimensional vector payoff, where $m := |S^i|$. We thus define $A \equiv A^i \colon S \to \mathbb{R}^m$, the *i-regret vector-payoff game associated to Γ*, by

$$A_k(s^i, s^{-i}) := u^i(k, s^{-i}) - u^i(s^i, s^{-i}) \qquad \text{for all} \quad k \in S^i,$$

and

$$A(s^i, s^{-i}) := (A_k(s^i, s^{-i}))_{k \in S^i},$$

for all $s = (s^i, s^{-i}) \in S^i \times S^{-i} = S$. Rewriting the regret as

$$D_t^i(k) = \frac{1}{t} \sum_{\tau \leqslant t} [u^i(k, s_\tau^{-i}) - u^i(s_\tau^i, s_\tau^{-i})]$$

shows that the vector of regrets at time t is just the average of the A vector payoffs in the first t periods: $D_t^i = (1/t) \sum_{\tau \leqslant t} A(s_\tau)$. The existence of a Hannan-consistent strategy in Γ is thus equivalent to the approachability by player i of the nonpositive orthant $\mathbb{R}_-^{S^i}$ in the vector-payoff game A, and a strategy is Hannan-consistent if and only if it guarantees that $\mathbb{R}_-^{S^i}$ is approached.

We now present two important results that apply in all generality to the regret setup.

Proposition 3.1. *For any (finite) N-person game Γ, the nonpositive orthant $\mathbb{R}_-^{S^i}$ is approachable by player i in the i-regret vector-payoff associated game.*

This proposition follows immediately from the next one. Observe that the approachability of $\mathbb{R}_-^{S^i}$ is equivalent, by the Blackwell condition (2.1),

[24] That is, $z_t^{-i}(s^{-i}) := |\{\tau \leqslant t : s_\tau^{-i} = s^{-i}\}|/t$ for each $s^{-i} \in S^{-i}$.

to the following: For every $\lambda \in \Delta(S^i)$ there exists $\sigma^i(\lambda) \in \Delta(S^i)$, a mixed action of player i, such that

$$\lambda \cdot A(\sigma^i(\lambda), s^{-i}) \leqslant 0 \qquad \text{for all} \quad s^{-i} \in S^{-i} \tag{3.1}$$

(indeed, $w(\lambda)$ equals 0 for $\lambda \geqslant 0$ and it is infinite otherwise). That is, the expected regret obtained by playing $\sigma^i(\lambda)$ lies in the half-space (through the origin) with normal λ. In this regret setup, the mixture $\sigma^i(\lambda)$ may actually be chosen in a simple manner:

PROPOSITION 3.2. *For any (finite) N-person game Γ and every $\lambda \in \Delta(S^i)$, condition* (3.1) *is satisfied by $\sigma^i(\lambda) = \lambda$.*

Proof. Given $\lambda \in \Delta(S^i)$, a $\sigma^i \equiv (\sigma^i_k)_{k \in S^i} \in \Delta(S^i)$ satisfies (3.1) if and only if

$$\sum_{k \in S^i} \lambda_k \sum_{j \in S^i} \sigma^i_j \, [u^i(k, s^{-i}) - u^i(j, s^{-i})] \leqslant 0 \tag{3.2}$$

for all $s^{-i} \in S^{-i}$. This may be rewritten as

$$\sum_{k \in S^i} u^i(k, s^{-i}) \left(\lambda_k \sum_{j \in S^i} \sigma^i_j - \sigma^i_k \sum_{j \in S^i} \lambda_j \right) = \sum_{k \in S^i} u^i(k, s^{-i})(\lambda_k - \sigma^i_k) \leqslant 0.$$

Therefore, by choosing σ^i so that all coefficients in the square brackets vanish—that is, by choosing $\sigma^i_k = \lambda_k$—we guarantee (3.2) and thus (3.1) for all s^{-i}. ▮

3.2. *Regret-Based Strategies*

The general theory of Section 2 is now applied to the regret situation. A *stationary regret-based* strategy for player i is a strategy of i such that the choices depend only on i's regret vector; that is, for every history h_{t-1}, the mixed action of i at time t is a function[25] of D^i_{t-1} only: $\sigma^i_t = \sigma^i(D^i_{t-1}) \in \Delta(S^i)$. The main result of this section is

THEOREM 3.3. *Consider a stationary regret-based strategy of player i given by a mapping $\sigma^i \colon \mathbb{R}^{S^i} \to \Delta(S^i)$ that satisfies the following:*

(R1) *There exists a continuously differentiable function $P \colon \mathbb{R}^{S^i} \to \mathbb{R}$ such that $\sigma^i(x)$ is positively proportional to $\nabla P(x)$ for every $x \notin \mathbb{R}^{S^i}_-$; and*

(R2) $\sigma^i(x) \cdot x > 0$ *for every $x \notin \mathbb{R}^{S^i}_-$.*

Then this strategy is Hannan-consistent for any (finite) N-person game.

[25] Note that the time t does not matter: the strategy is "stationary."

Proof. Apply Theorem 2.1 for $\mathscr{C} = \mathbb{R}^{S^i}_-$ together with Propositions 3.1 and 3.2: (D1) and (D2) yield (R1), and (D3) yields (R2). ∎

We have thus obtained a wide class of strategies that are Hannan-consistent. It is noteworthy that these are "universal" strategies: the mapping σ^i is independent of the game (see also the "variable game" case in Section 5).

Condition (R2) says that when $D^i_{t-1} \notin \mathbb{R}^{S^i}_-$—i.e., when some regret is positive—the mixed choice σ^i_t of i satisfies $\sigma^i_t \cdot D^i_{t-1} > 0$. This is equivalent to

$$u^i(\sigma^i_t, z^{-i}_{t-1}) > \bar{U}^i_{t-1}. \tag{3.3}$$

That is, the expected payoff of i from playing σ^i_t against the empirical distribution z^{-i}_{t-1} of the actions chosen by the other players in the past is higher than his realized average payoff. Thus σ^i_t is a *better reply*, where "better" is relative to the obtained payoff. By comparison, fictitious play always chooses an action that is a *best reply* to the empirical distribution z^{-i}_{t-1}. For more on this "better vs best" issue, see Subsection 4.2 below and Hart and Mas-Colell [16, Section 4(e)].

We now describe a number of interesting special cases, in order of increasing generality.

1. *l_2-potential*: $P(x) = (\sum_{k \in S^i} ([x_k]_+)^2)^{1/2}$. This yields (after normalization) $\Lambda(x) = (1/\|[x]_+\|_1)[x]_+$ for $x \notin \mathbb{R}^{S^i}_-$, and the resulting strategy is $\sigma^i_t(k) = [D^i_{t-1}(k)]_+ / \sum_{k' \in S^i} [D^i_{t-1}(k')]_+$ when $D^i_{t-1} \notin \mathbb{R}^{S^i}_-$. This is the Hannan-consistent strategy introduced in Hart and Mas-Colell [16, Theorem B], where the play probabilities are proportional to the positive regrets.

2. *l_p-potential*: $P(x) = (\sum_{k \in S^i} ([x_k]_+)^p)^{1/p}$ for some $1 < p < \infty$. This yields $\sigma^i_t(k) = ([D^i_{t-1}(k)]_+)^{p-1} / \sum_{k' \in S^i} ([D^i_{t-1}(k')]_+)^{p-1}$, i.e., play probabilities that are proportional to a fixed positive power $(p-1>0)$ of the positive regrets.

3. *Separable potential*: A *separable* strategy is one where σ^i_t is proportional to a vector whose kth coordinate depends *only* on the kth regret; i.e., σ^i_t is proportional to a vector of the form $(\psi_k(D^i_{t-1}(k)))_{k \in S^i}$. Conditions (R1) and (R2) result in the following requirements:[26] For each k in S^i, the function $\psi_k : \mathbb{R} \to \mathbb{R}$ is continuous; $\psi_k(x_k) > 0$ for $x_k > 0$; and $\psi_k(x_k) = 0$ for $x_k \leqslant 0$. The corresponding potential is $P(x) = \sum_{k \in S^i} \Psi_k(x_k)$, where $\Psi_k(x) := \int_{-\infty}^{x} \psi_k(y)\, dy$. Note that, unlike the previous two cases, the functions ψ_k may differ for different k, and they need not be monotonic (thus a higher regret may not lead to a higher probability).

[26] Consider points x with $x_j = \pm\varepsilon$ for all $j \neq k$.

Finally, observe that in all of the above cases, actions with negative or zero regret are never chosen. This need no longer be true in the general (nonseparable) case; see Subsection 4.2 below.

3.3. *Counterexamples*

The counterexamples of Subsection 2.3 translate easily into the regret setup.

- *The role of "better"* (R2). Consider the one-person game

R1	0
R2	1

The resulting regret game is given in Example 2.7. The strategy of playing "R1 forever" satisfies condition (R1) but not condition (R2) (or (3.3)), and it is indeed not Hannan-consistent.

- *The role of continuity in* (R1). Consider the simplest two-person coordination game (a well-known stumbling block for many strategies)

	C1	C2
R1	(1, 1)	(0, 0)
R2	(0, 0)	(1, 1)

The resulting regret game for the Row player is precisely the vector-payoff game of Examples 2.4 and 2.5, where we looked at the approachability question for the nonpositive orthant. The two strategies we considered there —which we have shown not to be Hannan-consistent—are not continuous. They correspond to the l_∞- and the l_1-potentials, respectively, which are not differentiable. (Note in particular that the l_∞-case yields "fictitious play," which is further discussed in Subsection 4.1 below.)

- *The role of integrability in* (R1). The vector-payoff game of our Example 2.6 can easily be seen to be a regret game. However, the approachable set there was not the nonpositive orthant. In order to get a counterexample to the result of Theorem 3.3 when integrability is not satisfied, one would need to resort to additional dimensions, that is, more than two strategies; we do not do it here, although it is plain that such examples are easy—though painful—to construct.

4. FICTITIOUS PLAY AND BETTER PLAY

4.1. *Fictitious Play and Smooth Fictitious Play*

As we have already pointed out, fictitious play may be viewed as a stationary regret-based strategy, corresponding to the l_∞-mapping (the directional mapping generated by the l_∞-potential). It does not guarantee Hannan-consistency (see Example 2.4 and Subsection 3.3); the culprit for this is the lack of continuity (i.e., (D1)).

Before continuing the discussion it is useful to note a property of fictitious play: *The play at time t does not depend on the realized average payoff $\bar{U}^i_{t-1}$.* Indeed, $\max_k D^i_{t-1}(k) = \max_k u^i(k, z^{-i}_{t-1}) - \bar{U}^i_{t-1}$, so an action $k \in S^i$ maximizes regret if and only if it maximizes the payoff against the empirical distribution z^{-i}_{t-1} of the actions of $-i$. In the general approachability setup of Section 2 (with $\mathscr{C} = \mathbb{R}^{S^i}_-$), this observation translates into the requirement that the directional mapping Λ be invariant to adding the same constant to all coordinates. That is, writing $e := (1, 1, ..., 1) \in \mathbb{R}^{S^i}$,

$$\Lambda(x) = \Lambda(y) \qquad \text{for any } x, y \notin \mathbb{R}^{S^i}_- \text{ with } x - y = \alpha e \text{ for some scalar } \alpha. \tag{4.1}$$

Note that, as it should be, the l_∞-mapping satisfies this property (4.1).

PROPOSITION 4.1. *A directional mapping Λ satisfies* (D2), (D3), *and* (4.1) *for $\mathscr{C} = \mathbb{R}^m_-$ if and only if it is equivalent to the l_∞-mapping, i.e., its potential P satisfies $P(x) = \phi(\max_k x_k)$ for some strictly increasing function ϕ.*

Proof. Since $\mathscr{C} = \mathbb{R}^m_-$, the allowable directions are $\lambda \geqslant 0$, $\lambda \neq 0$. Thus $\nabla P(x) \geqslant 0$ for a.e. $x \notin \mathbb{R}^m_-$ by (D2), implying that the limit of $\nabla P(x) \cdot x$ is $\leqslant 0$ as x approaches the boundary of $\mathbb{R}^m_-$. But $\nabla P(x) \cdot x > 0$ for a.e. $x \notin \mathbb{R}^m_-$ by (D3), implying that the limit of $\nabla P(x) \cdot x$ is in fact 0 as x approaches bd $\mathbb{R}^m_-$. Because P is Lipschitz, it follows that P is constant on bd $\mathbb{R}^m_-$, i.e., $P(x) = P(0)$ for every $x \in \text{bd } \mathbb{R}^m_-$. By (D3) again we have $P(x) > P(0)$ for all $x \notin \mathbb{R}^m_-$. Adding to this the invariance condition (4.1) implies that the level sets of P are all translates by multiples of e of bd $\mathbb{R}^m_- = \{x \in \mathbb{R}^m : \max_k x_k = 0\}$. ∎

Since the l_∞-mapping does not guarantee that $\mathscr{C} = \mathbb{R}^m_-$ is approached (again, see Example 2.4 and Subsection 3.3), we have

COROLLARY 4.2. *There is no stationary regret-based strategy that satisfies* (R1) *and* (R2) *and is independent of realized average payoff.*

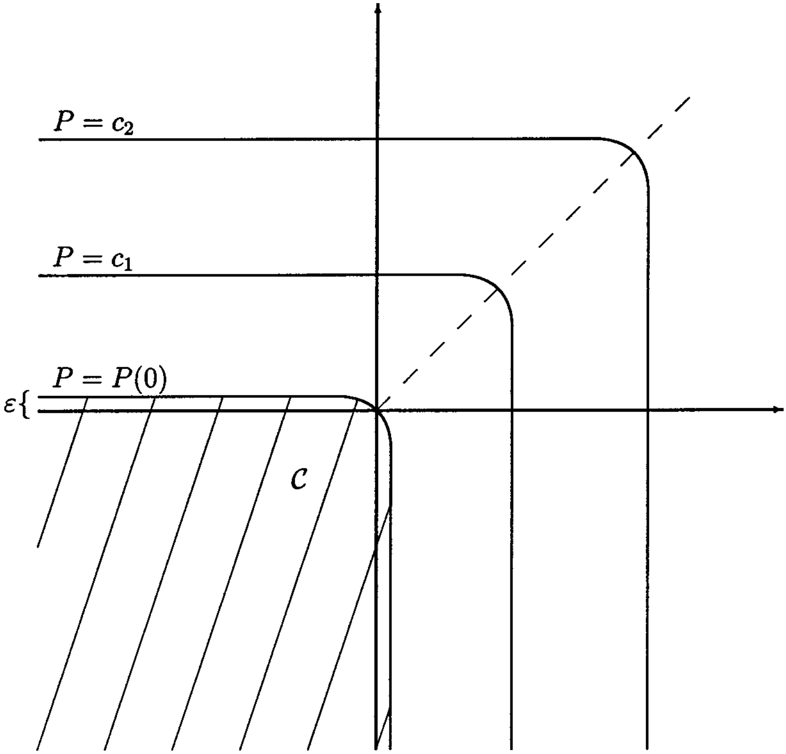

FIG. 6. The level sets of the potential of smooth fictitious play.

The import of the corollary (together with the indispensability of conditions (R1) and (R2), as shown by the counterexamples in Subsection 3.3) is that one cannot simultaneously have independence of realized payoffs and guarantee Hannan-consistency in every game.

We must weaken one of the two properties. One possibility is to weaken the consistency requirement to *ε-consistency*, $\limsup_t D_t^i(k) \leqslant \varepsilon$ for all k. Fudenberg and Levine [12] propose a smoothing of fictitious play that—like fictitious play itself—is independent of realized payoffs. In essence, their function P is convex, smooth, and satisfies the property that its level sets are obtained from each other by translations along the $e = (1, ..., 1)$ direction[27] (see Fig. 6). The level set of P through 0 is therefore smooth; it is very close to the boundary of the negative orthant but unavoidably distinct from it. The resulting strategy approaches $\mathscr{C} = \{x : P(x) \leqslant P(0)\}$ (recall Remark 5 in Subsection 2.1: a set of the form

[27] Specifically, $P(x) = \max_{\sigma^i \in \Delta(S^i)} \{\sigma^i \cdot x + \varepsilon v(\sigma^i)\}$, where $\varepsilon > 0$ is small and v is a strictly differentiably concave function, with gradient vector approaching infinite length as one approaches the boundary of $\Delta(S^i)$.

$\{x : P(x) \leqslant c\}$, for constant c, is approachable when $c \geqslant P(0)$—since it contains $\mathbb{R}_{-}^{S^i}$—and is not approachable when $c < P(0)$—since it does not contain 0). The set $\mathscr{C}$ is strictly larger than $\mathbb{R}_{-}^{S^i}$; it is an ε-neighborhood of the negative orthant $\mathbb{R}_{-}^{S^i}$. Thus one obtains only ε-consistency.[28] [29]

The other possibility is to allow the strategy to depend also on the realized payoffs. Then there are strategies that are close to fictitious play and guarantee Hannan-consistency in any game. Take, for instance, the l_p-potential strategy for large enough p (see Subsection 3.2).[30]

4.2. *Better Play*

All the examples presented until now satisfy an additional natural requirement, namely, that only actions with positive regret are played (provided, of course, that there are such actions). Formally, consider a stationary regret-based strategy of player i that is given by a mapping σ^i: $\mathbb{R}^{S^i} \to \Delta(S^i)$ (see Theorem 3.3); we add to (R1) and (R2) the following condition:[31]

(R3) For every $x \notin \mathbb{R}^{S^i}_{-\,-}$, if $x_k < 0$ then $[\sigma^i(x)]_k = 0$.

Since x is the i-regret vector, (R3) means that σ^i gives probability 1 to the set of actions with nonnegative regret (unless all regrets are negative, in which case there is no requirement[32]). This may be rewritten as

$$[\sigma^i_t]_k > 0 \qquad \text{only if} \quad u^i(k, z^{-i}_{t-1}) \geqslant \bar{U}^i_{t-1}. \tag{4.2}$$

That is, only those actions k are played whose payoff against the empirical distribution z^{-i}_{t-1} of the opponents' actions is at least as large as the actual realized average payoff $\bar{U}^i_{t-1}$; in short, only the "better actions."[33] For an example where (R3) is *not* satisfied, see Fig. 7.

[28] Other smoothings have been proposed, including Hannan [15], Foster and Vohra [7, 9], and Freund and Schapire [11] (in the latter—which corresponds to exponential smoothing—the strategy is nonstationary, i.e., it depends not only on the point in regret space but also on the time t; of course, nonstationary strategies where ε decreases with t may yield exact consistency).

[29] Smooth fictitious play may be equivalently viewed, in our framework, as first taking a set $\mathscr{C}$ that is close to the negative orthant and has smooth boundary, and then using the l_∞-distance from $\mathscr{C}$ as a potential (recall Remark 6 in Subsection 2.1).

[30] This amounts to smoothing the norm and keeping $\mathscr{C}$ equal to the negative orthant, whereas the previous construction smoothed the boundary of $\mathscr{C}$ and kept the l_∞-norm. These are the two "dual" ways of generating a smooth potential (again, see Remark 6 in Subsection 2.1).

[31] Note that the condition needs to be satisfied not only for $x \notin \mathbb{R}_{-}^{S^i}$, but also for $x \in \operatorname{bd} \mathbb{R}_{-}^{S^i}$.

[32] See Footnote 34 below.

[33] Observe that (R2) (or (3.3)) is a requirement on the average over all played actions k, whereas (R3) (or (4.2)) applies to each such k separately.

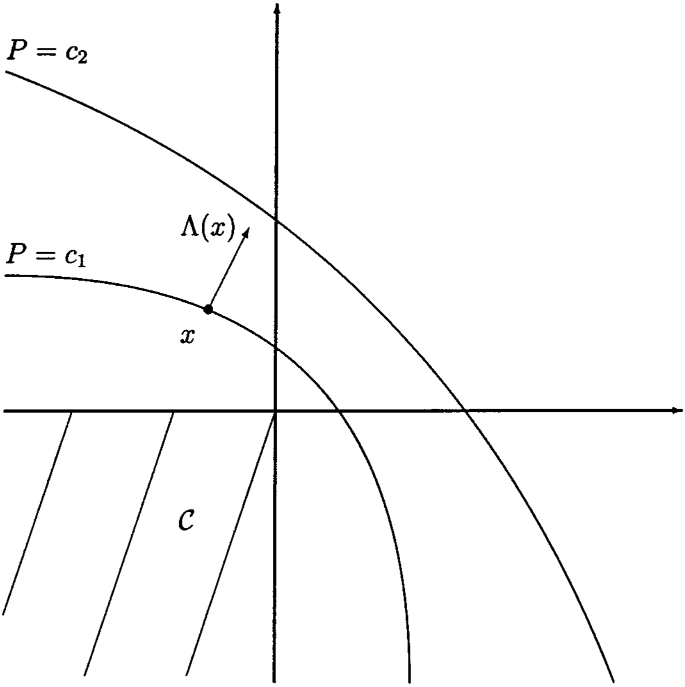

FIG. 7. (R3) is not satisfied.

The l_p-potential strategies, for $1 < p < \infty$, and in fact all separable strategies (see Subsection 3.2) essentially[34] satisfy (R3). Fictitious play (with the l_∞-potential) also satisfies (R3): the action chosen is a "best" one (rather than just "better"). At the other extreme, the l_1-potential strategy gives equal probability to *all* better actions, so it also satisfies (R3). (However, these last two do not satisfy (R1).)

Using condition (R3) yields the following result, which is a generalization of Theorem A of Monderer *et al.* [22] for fictitious play:

PROPOSITION 4.3. *Consider a stationary regret-based strategy of player i given by a mapping* $\sigma^i\colon \mathbb{R}^{S^i} \to \Delta(S^i)$ *that satisfies*[35] (R3). *Then, in any (finite) N-person game, the maximal regret of i is always nonnegative*:

$$\max_{k \in S^i} D_t^i(k) \geqslant 0 \qquad \textit{for all } t.$$

[34] The condition in (R3) is automatically satisfied in these cases for $x \notin \mathbb{R}^{S^i}_-$; one needs to impose it explicitly for $x \in \operatorname{bd} \mathbb{R}^{S^i}_-$ (where, until now, we had no requirements).

[35] Note that (R1) and (R2) are not assumed.

Proof. The proof is by induction, starting with $D_0^i(k)=0$ for all k. Assume that $\max_k D_{t-1}^i(k) \geqslant 0$ (or $D_{t-1}^i \notin \mathbb{R}_{--}^{S^i}$). By (R3), $D_{t-1}^i(k) \geqslant 0$ for any k chosen at time t; since $A_k(k, s_t^{-i})=0$ it follows that $D_t^i(k) = (1/t)((t-1)\, D_{t-1}^i(k) + t0) \geqslant 0$. ▮

Thus, the vector of regrets never enters the negative orthant.[36] Recall that the result of Theorem 3.3 is that the vector of regrets approaches the nonpositive orthant. To combine the two, we define *better play* as any stationary regret-based strategy of player i that is given by a mapping σ^i satisfying (R1)–(R3). We thus have

Corollary 4.4. *In any (finite) N-person game, if player i uses a better play strategy, then his maximal regret converges to* 0 *a.s.*

$$\lim_{t\to\infty} \max_{k\in S^i} D_t^i(k) = \lim_{t\to\infty} \left(\max_{k\in S^i} u^i(k, z_t^{-i}) - \bar{U}_t^i\right) = 0 \qquad a.s.$$

That is, the average payoff $\bar{U}_t^i$ of player i up to time t is close, as $t \to \infty$, to i's best-reply payoff against the empirical distribution of the other players' actions. In particular, in a two-person zero-sum game we obtain the following.

Corollary 4.5. *In any (finite) two-person zero-sum game, if both players use better-play strategies, then*:

(i) *For each player the empirical distribution of play converges to the set of optimal actions.*[37]

(ii) *The average payoff converges to the value of the game.*

Proof. Let 1 be the maximizer and 2 the minimizer, and denote by v the minimax value of the game. Then $\max_{k\in S^1} u^1(k, z_t^2) \geqslant v$, so by Corollary 4.4 we have $\liminf_t \bar{U}_t^1 \geqslant v$. The same argument for player 2 yields the opposite inequality, thus $\lim_t \bar{U}_t^1 = v$. Therefore $\lim_t \max_{k\in S^1} u^1(k, z_t^2) = v$ (apply the Corollary again), hence any limit point of the sequence z_t^2 must be an optimal action of player 2; similarly for player 1. ▮

Thus, better play enjoys the same properties as fictitious play in two-person zero-sum games (for fictitious play, see Robinson [24] for the convergence to the set of optimal strategies, and Monderer *et al.* [22, Theorem B] and Rivière [23] for the convergence of the average payoff).

[36] Therefore, at every period there are always actions with nonnegative regret—out of which the next action is chosen (and so the condition $x \notin \mathbb{R}_{--}^{S^i}$ in (R3) always holds).

[37] That is, the set of mixed actions that guarantee the value.

5. DISCUSSION AND EXTENSIONS

In this section we discuss a number of extensions of our results.

5.1. *Conditional Regrets*

As stated in the Introduction, we have been led to the "no regret" Hannan-consistency property from considerations of "no conditional regret" that correspond to correlated equilibria (see Hart and Mas-Colell [16]). Given two actions k and j of player i, the *conditional regret* from j to k is the change that would have occurred in the average payoff of i if he had played action k in all those periods where he did play j (and everything else is left unchanged). That is,

$$DC_t^i(j,k) := \frac{1}{t} \sum_{\tau \leqslant t : s_\tau^i = j} [u^i(k, s_\tau^{-i}) - u^i(s_\tau)]$$

for every[38] $j,\ k \in S^i$. The vector of regrets DC_t^i is now in $\mathbb{R}^L$, where $L := S^i \times S^i$, and the empirical distribution of actions up to time t constitutes a correlated equilibrium if and only if $DC_t^i \leqslant 0$. Thus, the set to be approached is the nonpositive orthant $\mathbb{R}_-^L$. The corresponding game with vector payoffs A is defined as follows: the (j,k) coordinate of the vector payoff $A(s^i, s^{-i}) \in \mathbb{R}^L$ is $u^i(k, s^{-i}) - u^i(j, s^{-i})$ when $s^i = j$, and it is 0 otherwise; hence $DC_t^i = (1/t) \sum_{\tau \leqslant t} A(s_\tau)$.

As in Propositions 3.1 and 3.2 (see Hart and Mas-Colell [16, Section 3]), it can easily be verified that:

- $\mathscr{C} = \mathbb{R}_-^L$ is always approachable.
- For every $\lambda \in \mathbb{R}_+^L$, the Blackwell approachability condition for $\mathscr{C} = \mathbb{R}_-^L$ ((2.1) or (3.1)) holds for any mixed action $\sigma^i = (\sigma_k^i)_{k \in S^i} \in \Delta(S^i)$ that satisfies

$$\sum_{j \in S^i} \sigma_j^i \lambda(j,k) = \sigma_k^i \sum_{j \in S^i} \lambda(k,j) \qquad \text{for all} \quad k \in S^i. \tag{5.1}$$

Viewing λ as an $S^i \times S^i$ matrix, condition (5.1) says that σ^i is an invariant vector for the (nonnegative) matrix λ.

- For every $\lambda \in \mathbb{R}_+^L$, there exists a $\sigma^i \in \Delta(S^i)$ satisfying (5.1).

[38] Note that each Hannan regret $D_t^i(k)$ is the sum of the conditional regrets $DC_t^i(j,k)$ over $j \neq k$. Thus the set of distributions of N-tuples of actions that satisfy the Hannan no-regret conditions includes the set of correlated equilibrium distributions. The inclusion is, in general, strict (the two sets coincide when every player has only two actions).

Applying Theorem 2.1 yields a large class of strategies. For example (as in Subsection 3.2), if P is the l_p-potential for some $1 < p < \infty$, then σ^i is an invariant vector of the matrix of the $p-1$ powers of the nonnegative regrets.[39] In the more general separable case, σ^i is an invariant vector of the matrix whose (j, k) coordinate is $\psi_{(j,k)}(DC^i_{t-1}(j, k))$, where $\psi_{(j,k)}$ is any real continuous function which vanishes for $x \leqslant 0$ and is positive for $x > 0$. As in Hart and Mas-Colell [16, Theorem A], if every player uses a strategy in this class (of course, different players may use different types of strategies), then the empirical distribution of play converges to the set of correlated equilibria of Γ.

Since finding invariant vectors is by no means a simple matter, in Hart and Mas-Colell [16] much effort is devoted to obtaining simple adaptive procedures, which use the matrix of regrets as a one-step transition matrix. To do the same here, one would use instead the matrix $\Lambda\ (DC^i_{t-1})$.

5.2. *Variable Game*

We noted in Subsection 3.2 that our strategies are game-independent. This allows us to consider the case where, at each period, a different game is being played (for example, a stochastic game). The strategy set of player i is the same set S^i in all games, but he does not know which game is currently being played. All our results—in particular, Theorem 3.3—continue to hold[40] provided player i is told, *after* each period t, which game was played at time t and what were the chosen actions s_t^{-i} of the other players. Indeed, as in Section 3, i can then compute the vector $a := A(s^i_t, s^{-i}_t) \in \mathbb{R}^{S^i}$, update his regret vector—$D^i_t = (1/t)((t-1)\, D^i_{t-1} + a)$—and then play $\sigma^i(D^i_t)$ in the next period, where σ^i is any mapping satisfying (R1) and (R2).

5.3. *Unknown Game*

When the player does not know the (fixed) game Γ that is played and is told, at each stage, only his own realized payoff (but not the choices of the other players)—in what may be referred to as a "stimulus–response" model—Hannan-consistency may nonetheless be obtained (see Foster and Vohra [7, 9], Auer *et al.* [1], Fudenberg and Levine [13, Section 4.8], Hart and Mas-Colell [16, Section 4(j); 17], and also Baños [2] and Megiddo [21] for related work). For instance, one can replace the regrets—which cannot be computed here—with appropriate estimates.[41]

[39] For $p = 2$ we get the matrix of regrets—which yields precisely Theorem A of Hart and Mas-Colell [16].

[40] Assuming the payoffs of i are uniformly bounded.

[41] Specifically, use $\hat{D}^i_t(k) := (1/t) \sum_{\tau \leqslant t : s^i_\tau = k} (1/[\sigma^i_\tau]_k)\, U^i_\tau - \bar{U}^i_t$ instead of $D^i_t(k)$ (see Hart and Mas-Colell [17, Section 3(c)]).

REFERENCES

1. P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, Gambling in a rigged casino: The adversarial multiarmed bandit problem, *in* "Proceedings of the 36th Annual Symposium on Foundations of Computer Science," pp. 322–331, IEEE, 1995.
2. A. Baños, On pseudo-games, *Ann. of Math. Statist.* **39** (1968), 1932–1945.
3. D. Blackwell, An analog of the minmax theorem for vector payoffs, *Pacific J. Math.* **6** (1956a), 1–8.
4. D. Blackwell, Controlled random walks, *in* "Proceedings of the International Congress of Mathematicians 1954," Vol. 3, pp. 335–338, North-Holland, Amsterdam, 1956.
5. A. Borodin and R. El-Yaniv, "Online Computation and Competitive Analysis," Cambridge Univ. Press, Cambridge, UK, 1998.
6. F. Clarke, "Optimization and Nonsmooth Analysis," Wiley, New York, 1983.
7. D. Foster and R. V. Vohra, A randomized rule for selecting forecasts, *Operations Res.* **41** (1993), 704–709.
8. D. Foster and R. V. Vohra, Calibrated learning and correlated equilibrium, *Games Econ. Behav.* **21** (1997), 40–55.
9. D. Foster and R. V. Vohra, Asymptotic calibration, *Biometrika* **85** (1998), 379–390.
10. D. Foster and R. V. Vohra, Regret in the on-line decision problem, *Games Econ. Behav.* **29** (1999), 7–35.
11. Y. Freund and R. E. Schapire, Adaptive game playing using multiplicative weights, *Games Econ. Behav.* **29** (1999), 79–103.
12. D. Fudenberg and D. K. Levine, Universal consistency and cautious fictitious play, *J. Econ. Dynam. Control* **19** (1995), 1065–1090.
13. D. Fudenberg and D. K. Levine, "Theory of Learning in Games," MIT Press, Cambridge, MA, 1998.
14. D. Fudenberg and D. K. Levine, Conditional universal consistency, *Games Econ. Behav.* **29** (1999), 104–130.
15. J. Hannan, Approximation to Bayes risk in repeated play, *in* "Contributions to the Theory of Games," Vol. III (M. Dresher, A. W. Tucker, and P. Wolfe, Eds.), Annals of Mathematics Studies, Vol. 39, pp. 97–139, Princeton Univ. Press, Princeton, NJ, 1957.
16. S. Hart and A. Mas-Colell, A simple adaptive procedure leading to correlated equilibrium, *Econometrica* **68** (2000), 1127–1150.
17. S. Hart and A. Mas-Colell, A reinforcement procedure leading to correlated equilibrium, Discussion Paper 224 (mimeo), Center for Rationality, The Hebrew University of Jerusalem, 2000.
18. N. Littlestone and M. K. Warmuth, The weighted majority algorithm, *Inform. Comput.* **108** (1994), 212–261.
19. M. Loève, "Probability Theory," Vol. II, 4th ed., Springer-Verlag, New York/Berlin, 1978.
20. R. D. Luce and H. Raiffa, "Games and Decisions," Wiley, New York, 1957.
21. N. Megiddo, On repeated games with incomplete information played by non-Bayesian players, *Int. J. Game Theory* **9** (1980), 157–167.
22. D. Monderer, D. Samet, and A. Sela, Belief affirming in learning processes, *J. Econ. Theory* **73** (1997), 438–452.
23. P. Rivière, "Quelques Modèles de Jeux d'Evolution," Ph.D. thesis, Université Paris 6, 1997.
24. J. Robinson, An iterative method of solving a game, *Ann. Math.* **54** (1951), 296–301.