Econometrica, Vol. 73, No. 5 (September, 2005), 1401–1430

ADAPTIVE HEURISTICS

BY SERGIU HART[1]

We exhibit a large class of simple rules of behavior, which we call *adaptive heuristics*, and show that they generate rational behavior in the long run. These adaptive heuristics are based on natural regret measures, and may be viewed as a bridge between rational and behavioral viewpoints. Taken together, the results presented here establish a solid connection between the dynamic approach of adaptive heuristics and the static approach of correlated equilibria.

KEYWORDS: Dynamics, heuristics, adaptive, correlated equilibrium, regret, regret-matching, uncoupled dynamics, joint distribution of play, bounded rationality, behavioral, calibration, fictitious play, approachability.

1. INTRODUCTION

CONSIDER DYNAMIC SETTINGS where a number of decision-makers interact repeatedly. We call a rule of behavior in such situations an *adaptive heuristic* if, on the one hand, it is simple, unsophisticated, simplistic, and myopic (a so-called "rule of thumb"), and, on the other, it leads to movement in seemingly "good" directions (like stimulus-response or reinforcement). One example of adaptive heuristic is to always choose a best reply to the actions of the other players in the previous period—or, for that matter, to the frequency of their actions in the past (essentially, the well-known "fictitious play").

Adaptive heuristics are boundedly rational strategies (in fact, highly "bounded away" from full rationality). The main question of interest is whether such simple strategies may in the long run yield behavior that is nevertheless highly sophisticated and rational.

This paper is based mainly on the work of Hart and Mas-Colell (2000, 2001a, 2001b, 2003a, 2003b), which we try to present here in a simple and elementary form (see Section 10 and the pointers there for the more general results). Significantly, when the results are viewed together new insights emerge—in particular, into the relations of adaptive heuristics to rationality on the one hand, and to behavioral approaches on the other. See Section 9, which may well be read immediately.

The paper is organized as follows. In Section 2 we provide a rough classification of dynamic models. The setting and notations are introduced in Section 3,

[1]Walras–Bowley Lecture 2003, delivered at the North American Meeting of the Econometric Society in Evanston, Illinois. A presentation is available at *http://www.ma.huji.ac.il/hart/abs/adaptdyn.html*. It is a great pleasure to acknowledge the joint work with Andreu Mas-Colell over the years, upon which this paper is based. I also thank Ken Arrow, Bob Aumann, Maya Bar-Hillel, Avraham Beja, Elchanan Ben-Porath, Gary Bornstein, Toni Bosch, Ido Erev, Drew Fudenberg, Josef Hofbauer, Danny Kahneman, Yaakov Kareev, Aniol Llorente, Yishay Mansour, Eric Maskin, Abraham Neyman, Bezalel Peleg, Motty Perry, Avi Shmida, Sorin Solomon, Menahem Yaari, and Peyton Young, as well as the editor and the anonymous referees, for useful discussions, suggestions, and comments. Research partially supported by the Israel Science Foundation.

and the leading adaptive heuristic, *regret matching*, is presented and analyzed in Section 4. Behavioral aspects of our adaptive heuristics are discussed in Section 5, and Section 6 deals with the notion of correlated equilibrium, to which play converges in the long run. Section 7 presents the large class of generalized regret matching heuristics. In Section 8 we introduce the notion of "uncoupledness" (which is naturally satisfied by adaptive heuristics) and show that uncoupled dynamics cannot be guaranteed to always lead to Nash equilibria. A summary together with the main insights of our work are provided in Section 9. Section 10 includes a variety of additional results and discussions of related topics, and the Appendix presents Blackwell's approachability theory, a basic technical tool in this area.

2. A RATIONAL CLASSIFICATION OF DYNAMICS

We consider dynamic models where the same game is played repeatedly over time. One can roughly classify dynamic models in game theory and economic theory into three classes: learning dynamics, evolutionary dynamics, and adaptive heuristics.

2.1. *Learning Dynamics*

In a (*Bayesian*) *learning dynamic*, each player starts with a prior belief on the relevant data (the "state of the world"), which usually includes the game being played and the other players' types and (repeated-game) strategies.[2] Every period, after observing the actions taken (or, more generally, some information about these actions), each player updates his beliefs (using Bayes' rule). He then plays optimally given his updated beliefs.

Such dynamics are the subject of much study; see, for example, the books of Fudenberg and Levine (1998, Chapter 8) and Young (2004, Chapter 7). Roughly speaking, conditions like "the priors contain a grain of truth" guarantee that in the long run play is close to the Nash equilibria of the repeated game; see Kalai and Lehrer (1993) and the ensuing literature.[3]

2.2. *Evolutionary Dynamics*

Here every player i is replaced by a *population* of individuals, each playing the given game in the role of player i. Each such individual always plays the same one-shot action (this fixed action is his "genotype"). The relative frequencies of the various actions in population i may be viewed as a *mixed action*

[2]To distinguish the choices of the players in the one-shot game and those in the repeated game, we refer to the former as *actions* and to the latter as *strategies*.

[3]Under weaker assumptions, Nyarko (1994) shows convergence to correlated equilibria.

of player i in the one-shot game (for instance, one third of the population having the "gene" L and two thirds the "gene" R corresponds to the mixed action $(1/3, 2/3)$ on (L, R)); one may think of the mixed action as the action of a randomly chosen individual.

Evolutionary dynamics are based on two main "forces": selection and mutation. *Selection* is a process whereby better strategies prevail; in contrast, *mutation*, which is rare relative to selection, generates actions at random, whether better or worse. It is the combination of the two that allows for natural adaptation: new mutants undergo selection, and only the better ones survive. Of course, selection includes many possible mechanisms: biological (the payoff determines the number of descendants, and thus the share of better strategies increases), social (imitation, learning), individual (experimentation, stimulus-response), and so on. What matters is that selection is "adaptive" or "improving," in the sense that the proportion of better strategies is likely to increase.

Dynamic evolutionary models have been studied extensively; see, for example, the books of Fudenberg and Levine (1998, Chapters 3 and 5), Hofbauer and Sigmund (1998), Weibull (1995), and Young (1998).

2.3. *Adaptive Heuristics*

We use the term *heuristics* for rules of behavior that are simple, unsophisticated, simplistic, and myopic (unlike the "learning" models of Section 2.1 above). These are "rules of thumb" that the players use to make their decisions. We call them *adaptive* if they induce behavior that reacts to what happens in the play of the game, in directions that, loosely speaking, seem "better." Thus, always making the same fixed choice, and always randomizing uniformly over all possible choices, are both heuristics. But these heuristics are not adaptive, since they are not at all responsive to the situation (i.e., to the game being played and the behavior of the other participants). In contrast, *fictitious play* is a prime example of an adaptive heuristic: at each stage one plays an action that is optimal against the frequency distribution of the past actions of the other players.

Adaptive heuristics commonly appear in behavioral models, such as reinforcement, feedback, and stimulus-response. There is a large literature, both experimental and theoretical, on various adaptive heuristics and their relative performance in different environments; see, for example, Fudenberg and Levine (1998, Chapters 2 and 4) and the literature in psychology (where this is sometimes called "learning"; also, the term "heuristics and biases" is used by Kahneman and Tversky—e.g., see Kahneman, Slovic, and Tversky (1982)).

2.4. *Degrees of Rationality*

One way to understand the distinctions between the above three classes of dynamics is in terms of the degree of rationality of the participants; see Figure 1. *Rationality* is viewed here as a process of optimization in interactive (multiplayer) environments.

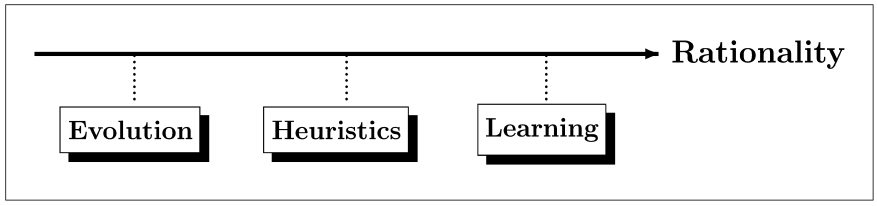

FIGURE 1.—A classification of dynamics.

Learning dynamics require high levels of rationality. Indeed, repeated-game strategies are complex objects; even more so are beliefs (i.e., probability distributions) over such objects; moreover, in every period it is necessary to update these beliefs, and, finally, to compute best replies to them.

At the other extreme are evolutionary dynamics. Here the individuals in each population do not exhibit any degree of rationality; their behavior ("phenotype") is completely mechanistic, dictated by their "genotype." They do not compute anything—they just "are there" and play their fixed actions. What may be viewed as somewhat rational is the aggregate dynamic of the population (particularly the selection component), which affects the relative proportions of the various actions.

Adaptive heuristics lie in between: on the one hand, the players do perform certain usually simple computations given the environment, and so the behavior is not fixed as in evolutionary dynamics; on the other hand, these computations are far removed from the full rationality and optimization that is carried out in learning models.

3. PRELIMINARIES

The setup is as follows. The basic game Γ is an N-person game in strategic (or normal) form. N is a positive integer, the players are $i = 1, 2, \ldots, N$, and to each player i there corresponds a set of *actions* S^i and a *payoff* (or *utility*) *function* $u^i : S \to \mathbb{R}$, where $S := S^1 \times S^2 \times \cdots \times S^N$ is the set of N-tuples of actions ("action combinations" or "action profiles") and $\mathbb{R}$ denotes the real line. When dealing with a fixed player i, it is at times convenient to take his set of actions to be[4] $S^i = \{1, 2, \ldots, m\}$.

The game Γ is repeatedly played over time. In a *discrete-time* dynamic model, the time periods are $t = 1, 2, \ldots$, and the action played by player i at time t is denoted s_t^i, with $s_t = (s_t^1, s_t^2, \ldots, s_t^N)$ standing for the N-tuple of actions at period t (these are the *actual* realized actions when randomizations are used). In a *continuous-time* dynamic model, the time t becomes a continuous variable. We will refer to Γ as the *stage* (or *one-shot*) *game.*

A standard assumption is that of *perfect monitoring*: at the end of each period t, all players observe s_t, the actions taken by everyone.

[4]The number of actions $m \equiv m^i$ may be different for different players.

Some notations: when $s = (s^1, s^2, \ldots, s^N) \in S$ is an N-tuple of actions, we write s^i for its ith coordinate and s^{-i} for the $(N-1)$-tuple of coordinates of all players except i (so $s^{-i} \in S^{-i} := S^1 \times \cdots \times S^{i-1} \times S^{i+1} \times \cdots \times S^N$ and $s = (s^i, s^{-i})$). When randomized strategies are used, σ_t^i denotes the *mixed action* of player i at time t; thus $\sigma_t^i \equiv (\sigma_t^i(1), \sigma_t^i(2), \ldots, \sigma_t^i(m)) \in \Delta(S^i) := \{x \in \mathbb{R}_+^m : \sum_{k=1}^m x(k) = 1\}$ is a probability distribution over $S^i = \{1, 2, \ldots, m\}$, with $\sigma_t^i(k)$ denoting, for each k in S^i, the probability that the action s_t^i taken by player i at time t is k.

4. REGRET MATCHING

We start by considering our basic adaptive heuristic, "regret matching," introduced in Hart and Mas-Colell (2000). While it may appear to be quite special, we will see later (in Sections 7 and 10) that the results generalize to a wide class of heuristics—and our comments and interpretations in Sections 4 and 5 apply to all of them.

Regret matching is defined by the following rule:

> Switch next period to a different action
> with a probability that is
> *proportional* to the *regret* for that action,
> where *regret* is defined as the increase in payoff
> had such a change always been made in the past.

That is, consider player i at time $T+1$. Denote by[5] U the average payoff that i has obtained up to now, i.e.,

$$U := \frac{1}{T} \sum_{t=1}^{T} u^i(s_t), \tag{1}$$

and let $j = s_T^i$ in S^i be the action that i played in the previous period T. For each alternative action $k \neq j$ in S^i, let $V(k)$ be the average payoff i would have obtained had he played k instead of j every time in the past that he actually played j; i.e.,

$$V(k) := \frac{1}{T} \sum_{t=1}^{T} v_t, \tag{2}$$

where

$$v_t := \begin{cases} u^i(k, s_t^{-i}), & \text{if } s_t^i = j, \\ u^i(s_t^i, s_t^{-i}) \equiv u^i(s_t), & \text{if } s_t^i \neq j. \end{cases} \tag{3}$$

[5]We drop the indices for readability (thus U_T^i is written U, and similarly for $V_T^i(k)$, $v_t^i(j,k)$, $R_T^i(k)$, and so on).

The *regret* $R(k)$ for action k is defined as the amount, if any, by which $V(k)$ exceeds the actual payoff U:

$$R(k) := [V(k) - U]_+, \tag{4}$$

where $[x]_+ := \max\{x, 0\}$ is the "positive part" of x. (The actual computation of the regrets consists of a simple updating from one period to the next; see Section 10.7.)

Regret matching stipulates that each action k different from the previous period's action j is played with a probability that is proportional to its regret $R(k)$, and, with the remaining probability, the same action j as last period is played again. That is, let $c > 0$ be a fixed constant;[6] then the probability of playing the action k at time $T + 1$ is given by

$$\sigma_{T+1}(k) := \begin{cases} cR(k), & \text{if } k \neq j, \\ 1 - \sum_{k:\, k \neq j} cR(k), & \text{if } k = j. \end{cases} \tag{5}$$

There are no requirements on the action in the first period: σ_1 is arbitrary.

Thus, player i considers whether to continue to play next period the same action j as in the previous period, or to switch to a different action k. Specifically, he looks at what would have happened to his average payoff had he always replaced j by k in the past:[7] he compares what he got, U, to what he would have gotten, $V(k)$. If the alternative payoff is no higher, i.e., if $V(k) \leq U$, then he has no regret for k (the regret $R(k)$ equals 0) and he does not switch to action k. If the alternative payoff is higher, i.e., if $V(k) > U$, then the regret for k is positive ($R(k)$ equals the increase $V(k) - U$) and i switches to action k with a probability that is proportional to this regret.

In particular, if i has no regret (i.e., all regrets $R(k)$ equal 0), then i plays for sure the same action j as last period. If some regrets are positive, then the higher the regret for some action, the higher the probability of switching to that action next period.

The main result of Hart and Mas-Colell (2000) is:

THEOREM 1—Regret Matching: *Let each player play regret matching. Then the joint distribution of play converges to the set of correlated equilibria of the stage game.*

[6]For instance, any c less than $1/(2mM)$ will do, where m is the number of actions of i and $M = \max_{s \in S} |u^i(s)|$ bounds the possible payoffs. Such a c guarantees that (5) yields a probability distribution over S^i and, moreover, that the probability of j is *strictly* positive.

[7]Since one is looking at the long-run average payoff, it makes sense to consider replacing j by k not just in the previous period, but also in all other periods in the past when j was played; after all, the effect of one period goes to zero as T increases. (Interestingly, one-period-regret matching yields essentially the evolutionary "replicator dynamic"; see Schlag (1998).)

The "joint distribution of play" (also known as the "empirical distribution" or "sample distribution") measures the relative frequency of each N-tuple of actions being played; i.e., the *joint distribution of play* for the first T periods is a probability distribution z_T on S, where, for each s in S,

$$z_T(s) := \frac{1}{T}\big|\{1 \le t \le T : s_t = s\}\big|$$

is the proportion of periods up to T in which the combination of actions s has been played.[8] See Section 4.1 for a discussion of the role of the joint distribution of play.

The concept of *correlated equilibrium* was introduced by Aumann (1974); it is a Nash equilibrium where the players may receive payoff-irrelevant signals before playing the game. This is discussed in Section 6.

The *Regret Matching Theorem* says that, for almost every history of play, the sequence of joint distributions of play $z_1, z_2, \ldots, z_T, \ldots$ converges to the set of correlated equilibria CE of Γ. This means that z_T is close to a correlated equilibrium, or, equivalently, is a correlated approximate equilibrium, for all T large enough. The convergence is to the *set* CE, *not* necessarily to a point in that set. See Section 10.8 for formal statements.

Note that it is the empirical distributions that become essentially correlated equilibria—not the actual play. Our results imply that the long-run *statistics of play* of heuristics-playing players (such as "regret-matchers") and of fully rational players (who play a correlated equilibrium each period) are indistinguishable.

The proof of the Regret Matching Theorem consists in showing, first, that all regrets vanish in the limit (this uses arguments suggested by Blackwell's approachability; see the Appendix) and, second, that such "no regret" situations precisely correspond to correlated equilibria; see Section 10.2 for further details.

It is interesting to note that as the regrets become small, so does the probability of switching (see (5)). Therefore, regret matching leads to longer and longer stretches of time in which the action is constant, and the play exhibits much "inertia" and infrequent switches. (This is similar to the behavior of fictitious play in the classic 3×3 example of Shapley (1964), where the play cycles among six outcomes with increasingly longer time intervals in each cycle.)

Finally, where does the "correlation" come from? The answer is, of course, from the commonly observed history of play. Indeed, each player's action is determined by his regrets, which are in turn determined by the history.[9]

[8]For a finite set A, we denote by $|A|$ the number of elements of A.

[9]At this point one may be tempted to conclude that, since the signal is common (all players observe the history of play), the convergence is in fact to *publicly* correlated equilibria (cf. Section 6). That is not so however: our players are *not* fully rational; they apply heuristics, whereby

4.1. *Joint Distribution of Play*

At each stage the players randomize independently of one another. This however does *not* imply that the joint distribution of play should be independent across players (i.e., the product of its marginal distributions) or that it should become independent in the long run.[10] The reason is that the probabilities the players use may change over time. To take a simple example, assume that in odd periods player 1 chooses T or B with probabilities $(3/4, 1/4)$ and, independently, player 2 chooses L or R with probabilities $(3/4, 1/4)$, whereas in even periods these probabilities become $(1/4, 3/4)$ for each player. The joint distribution of play will then converge almost surely to $(5/16, 3/16, 3/16, 5/16)$ (for TL, TR, BL, and BR, respectively)—which is *not* the product of its marginals, $(1/2, 1/2)$ on (T, B) and $(1/2, 1/2)$ on (L, R).

The joint distribution of play is fully determined by the history of play, which players standardly observe. So having players determine their actions based on the joint distribution of play (rather than just the marginal distributions) does not go beyond the "standard monitoring" assumption that is commonly used. It is information that the players possess anyway.

Finally—and this is a behavioral observation—people do react to the joint distribution. Think of a two-person Matching Pennies game, where, say, half the time they play HH, and half the time TT. The players will very quickly notice this, and at least one of the players (the "mismatching" player in this case) will change his behavior; but, if he were to look only at the marginal distributions of play, he would see $(1/2, 1/2)$ for each player, and have no reason to change. In general, people are very much aware of coincidences, signals, communications, and so on (even to the point of interpreting random phenomena as meaningful)—which just goes to show that they look at the joint distribution, and not only at the marginals.

To summarize: reasonable models of play can—and should—take into account the joint distribution of play.

5. BEHAVIORAL ASPECTS

Regret matching, as well as its generalizations below, embodies commonly used rules of behavior. For instance, if all the regrets are zero ("there is no regret"), then a regret matching player will continue to play the same action of the previous period. This is similar to common behavior, as expressed in the saying "Never change a winning team."

When some regrets are positive, actions may change—with probabilities that are proportional to the regrets: the higher the payoff would have been from

---

each one uses the history to determine only his *own* regrets. Since all regrets are based on the common history, they are correlated among the players—but except for special cases they are far from being fully correlated.

[10]This point is, of course, not new; see, for instance, Fudenberg and Kreps (1988, 1993).

switching to another action in the past, the higher the tendency is to switch to that action now. Again, this seems to fit standard behavior; we have all seen ads of the sort "Had you invested in A rather than B, you would have gained X more by now. So switch to A now!" (and, the larger X is, the larger the size of the ad).

It has been observed that people tend to have too much "inertia" in their decisions: they stick to their current state for a disproportionately long time (as in the "status quo bias"; see Samuelson and Zeckhauser (1988) and, recently, Moshinsky (2003)). Regret matching has "built-in" inertia: the probability of not switching (i.e., repeating the previous period's action) is always strictly positive (see footnote 6). Moreover, as we saw in Section 4, regret matching leads to behavior where the same action is played over and over again for long time intervals.

Regret matching is not very sophisticated; players neither develop beliefs nor reply optimally (as in the learning dynamics of Section 2.1). Rather, their rule of behavior is simple and defined directly on actions; "propensities" of play are adjusted over time. In the learning, experimental, and behavioral literature there are various models that bear a likeness to regret matching; see Bush and Mosteller (1955), Roth and Erev (1995), Erev and Roth (1998), Camerer and Ho (1998, 1999), and others; probably the closest are the models of Erev–Roth. Also, incorporating regret measures into the utility function has been used to provide alternative theories of decision-making under uncertainty; see Bell (1982) and Loomes and Sugden (1982).

Recently, the study of Camille et al. (2004) has shown that certain measures of regret influence choices, and that the orbitofrontal cortex is involved in experiencing regret.

In summary, while we have arrived at regret matching and the other heuristics of this paper from purely theoretical considerations, it turns out that they have much in common with actual rules of behavior that are frequently used in real decisions.

6. CORRELATED EQUILIBRIA

In this section we leave the dynamic framework and discuss the notion of correlated equilibrium. Its presentation (which may well be skipped by the expert reader) is followed by a number of comments showing that this concept, belonging to the realm of full rationality, is particularly natural and useful.

We thus consider the one-shot game Γ. Assume that, before playing the game Γ, each player i receives a signal θ^i. These signals may be correlated: the combination of signals $\theta = (\theta^1, \theta^2, \ldots, \theta^N)$ occurs according to a joint probability distribution F, commonly known to all players. Moreover, the signals do not affect the payoffs of the game. Can this affect the outcome?

Indeed, it can: the players may use these signals to correlate their choices. For a simple example, take the Battle of the Sexes game (see Figure 2). Consider a public coin toss (i.e., the common signal $\theta^1 = \theta^2$ is either H or T, with

	HOCKEY	THEATER
HOCKEY	$2,1$	$0,0$
THEATER	$0,0$	$1,2$

	HOCKEY	THEATER
HOCKEY	$1/2$	0
THEATER	0	$1/2$

FIGURE 2.—The Battle of the Sexes game (left) and a (publicly) correlated equilibrium (right).

probabilities $(1/2, 1/2)$), after which both go to the hockey game if H and to the theater if T; this constitutes a Nash equilibrium of the extended game (with the signals)—which cannot be achieved in the original game.

Formally, a *correlated equilibrium* (introduced by Aumann (1974)) of the game Γ is a Nash equilibrium of a "pregame signals extension" of Γ. Clearly, only the probability distribution of the signals matters. Let thus ψ be the induced probability distribution on the N-tuples of actions; i.e., for each action combination s in S, let $\psi(s)$ be the probability of all those signal combinations after which the players choose[11] s. The conditions for ψ to be a correlated equilibrium are that

$$\sum_{s^{-i}\in S^{-i}} \psi(j, s^{-i}) u^i(j, s^{-i}) \geq \sum_{s^{-i}\in S^{-i}} \psi(j, s^{-i}) u^i(k, s^{-i}) \tag{6}$$

for all players i and all actions j, k in S^i. Indeed, if there are no deviations, then the expected payoff of player i when he chooses action j is the expression on the left-hand side; if i were the only one to deviate and choose instead of j some other action k, then his expected payoff would be the expression on the right-hand side. As a *canonical* setup, think of a "referee" who chooses, according to the distribution ψ, an N-tuple of actions $s = (s^1, s^2, \ldots, s^N)$ for all players, and then sends to each player i the message "s^i" (a "recommendation to play s^i"); a correlated equilibrium ensues if for each player it is always best to follow the recommendation (i.e., to play the recommended s^i), assuming that all other players also do so.

We denote by CE the set of correlated equilibria; it is a subset of $\Delta(S)$, the set of probability distributions on S. If the inequalities (6) hold only within some $\varepsilon > 0$, we will say that ψ is a *correlated approximate equilibrium*; more precisely, a *correlated ε-equilibrium*.

In the special case where the signals are *independent* across players (i.e., when the joint distribution ψ satisfies $\psi(s) = \psi^1(s^1) \cdot \psi^2(s^2) \cdot \ldots \cdot \psi^N(s^N)$ for all s, with ψ^i denoting the ith marginal of ψ), a correlated equilibrium is just a Nash equilibrium of Γ. At the other extreme, when the signals are *fully correlated* (i.e., common, or public—like "sunspots"), each signal must necessarily be followed by a Nash equilibrium play; hence such correlated equilibria—

[11] In general, since players may randomize their choices, $\psi(s)$ is the corresponding total probability of s.

	LEAVE	STAY
LEAVE	$5,5$	$3,6$
STAY	$6,3$	$0,0$

	LEAVE	STAY
LEAVE	$1/3$	$1/3$
STAY	$1/3$	0

FIGURE 3.—A correlated equilibrium in the Chicken game.

called *publicly correlated equilibria*—correspond to weighted averages (convex combinations) of Nash equilibria of Γ, as in the Battle of the Sexes example of Figure 2.

In general, when the signals are neither independent nor fully correlated, new equilibria arise. For example, in the Chicken game, there is a correlated equilibrium that yields equal probabilities of $1/3$ to each action combination except (STAY, STAY) (see Figure 3). Indeed, let the signal to each player be L or S; think of this as a recommendation to play LEAVE or STAY, respectively. When the row player gets the signal L, he assigns a (conditional) probability of $1/2$ to each one of the two pairs of signals (L, L) and (L, S); so, if the column player follows his recommendation, then the row player gets an expected payoff of $4 = (1/2)5 + (1/2)3$ from playing LEAVE, and only $3 = (1/2)\,6 + (1/2)\,0$ from deviating to STAY. When the row player gets the signal S, he deduces that the pair of signals is necessarily (S, L), so if the column player indeed plays LEAVE then the row player is better off choosing STAY. Similarly for the column player.

For examples of correlated equilibria in biology, see Hammerstein and Selten (1994, Section 8.2 and the references there) and Shmida and Peleg (1997) (speckled wood butterflies, studied by Davies (1978), "play" the Chicken game). Other examples can be found in team sports, like basketball and football. Teams that are successful due to their so-called "team play" develop signals that allow correlation among their members but yield no information to their opponents.[12] For a stylized example, consider a signal that tells the team members whether to attack on the left or on the right—but is unknown (or unobservable) to the opposing team.

In fact, signals are all around us—whether public, private, or mixed. These signals are mostly irrelevant to the payoffs of the game that is being played. Nevertheless, it is hard to exclude the possibility that they may find their way into the equilibrium—so the notion of correlated equilibrium becomes all the more relevant.

Finally, Aumann (1987) shows that all players always being "Bayesian rational" is equivalent to their playing a correlated equilibrium (under the "com-

[12]These signals may well be largely inexplicit and unconscious; they are recognized due to the many repeated plays of the team.

mon prior" or "consistency" assumption of Harsanyi (1967–1968)).[13] Correlated equilibrium is thus a concept that embodies full rationality.

7. GENERALIZED REGRET MATCHING

The regret matching strategy of Section 4 appears to be very specific: the play probabilities are directly proportional to the regrets. It is natural to enquire whether this is necessary for our result of Theorem 1. What would happen were the probabilities proportional to, say, the square of the regrets? Another issue is the connection to other dynamics leading to correlated equilibria, particularly variants of conditional smooth fictitious play (Fudenberg and Levine (1999a); see Section 10.4 below).

This leads us to consider a large class of adaptive heuristics that are based on regrets. Specifically, instead of $cR(k)$ in (5), we now allow functions $f(R(k))$ of the regret $R(k)$ that are *sign-preserving*, i.e., $f(x) > 0$ for $x > 0$ and $f(0) = 0$.

A strategy σ of player i is called a *generalized regret matching* strategy if the action at time $T+1$ is chosen according to probabilities

$$\sigma_{T+1}(k) := \begin{cases} f(R(k)), & \text{if } k \neq j, \\ 1 - \sum_{k:\, k \neq j} f(R(k)), & \text{if } k = j, \end{cases} \tag{7}$$

where f is a Lipschitz continuous sign-preserving real function,[14] $j = s_T^i$ is the previous period's action, and $R(k)$ is the regret for action k (as given in Section 4); again, the play in the first period is arbitrary. The following result is based on Hart and Mas-Colell (2001a, Sections 3.2 and 5.1) and proved in Cahn (2004, Theorem 4.1):

THEOREM 2—Generalized Regret Matching: *Let each player play a generalized regret matching strategy*.[15] *Then the joint distribution of play converges to the set of correlated equilibria of the stage game*.

In fact, the full class of generalized regret matching strategies (for which Theorem 2 holds) is even larger; see Section 10.3. In particular, one may use a different $f_{k,j}$ for each pair $k \neq j$, or allow $f_{k,j}$ to depend on the whole vector of regrets and not just on the kth regret.

As a special case, consider the family of functions $f(x) = cx^r$, where $r \geq 1$ and $c > 0$ is an appropriate constant.[16] At one extreme, when $r = 1$, this is

[13] In this light, our result may appear even more surprising: non-Bayesian and far-from-rational behavior leads in the long run to outcomes that embody full Bayesian rationality and common prior.

[14] There is L such that $|f(x) - f(y)| \leq L|x - y|$ for all x, y, and also $\sum_k f(R(k)) < 1$.

[15] Different players may use different such strategies.

[16] The condition $r \geq 1$ is needed for Lipschitz continuity.

regret matching. At the other extreme, the limit as $r \to \infty$ is such that the switching probability $1 - \sigma(j)$ is equally divided among those actions $k \neq j$ with maximal regret (i.e., with $R(k) = \max_{\ell \neq j} R(\ell)$). This yields a variant of fictitious play, which however no longer satisfies the continuity requirement; therefore this strategy does not belong to our class and, indeed, the result of Theorem 2 does not hold for it (see Section 10.4). To regain continuity one needs to smooth it out, which leads to *smooth conditional fictitious play*; see Cahn (2004, Section 5).

8. UNCOUPLED DYNAMICS

At this point it is natural to ask whether there are adaptive heuristics that lead to *Nash equilibria* (the set of Nash equilibria being, in general, a strict subset of the set of correlated equilibria).

The answer is positive for *special* classes of games. For instance, two-person zero-sum games, two-person potential games, dominance-solvable games, and supermodular games are classes of games where fictitious play or general regret-based strategies make the marginal distributions of play converge to the set of Nash equilibria of the game (see Hofbauer and Sandholm (2002) and Hart and Mas-Colell (2003a) for some recent work). But what about general games? Short of variants of exhaustive search (deterministic or stochastic),[17] there are no general results in the literature. Why is that so?

A natural requirement for adaptive heuristics (and adaptive dynamics in general) is that each player's strategy not depend on the payoff functions of the other players; this condition was introduced in Hart and Mas-Colell (2003b) and called *uncoupledness*. Thus, the strategy may depend on the actions of the other players—what they *do*—but not on their preferences—*why* they do it. This is an *informational* requirement: actions are observable, utilities are not. Almost all dynamics in the literature are indeed uncoupled: best-reply, better-reply, payoff-improving, monotonic, fictitious play, regret-based, replicator dynamics, and so on.[18] They all use the history of actions and determine the play as some sort of "good" reply to it, using only the player's own utility function.

Formally, we consider here dynamic systems in continuous time,[19] of the general form

$$\dot{x}(t) = F(x(t); \Gamma), \tag{8}$$

[17]See Foster and Young (2003a, 2003b), Kakade and Foster (2004), Young (2004), Hart and Mas-Colell (2004), and Germano and Lugosi (2004).

[18]One example of a "non-uncoupled" dynamic is to compute a Nash equilibrium $\bar{x} = (\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^N)$ and then to let each player i converge to $\bar{x}^i$; of course, the determination of $\bar{x}$ generally requires knowing *all* payoff functions.

[19]The results up to now on regret matching and generalized regret matching carry over to the continuous-time setup—see Section 10.6. For a discrete-time treatment of "uncoupledness," see Hart and Mas-Colell (2004).

where Γ is the game and the state variable is $x(t) = (x^1(t), x^2(t), \ldots, x^N(t))$, an N-tuple of (mixed) actions in $\Delta(S^1) \times \Delta(S^2) \times \cdots \times \Delta(S^N)$; equivalently, this may be written as

$$\dot{x}^i(t) = F^i(x(t); \Gamma) \quad \text{for each } i \in N, \tag{9}$$

where $F = (F^1, F^2, \ldots, F^N)$. Various dynamics can be represented in this way; the variable $x^i(t)$ may be, for instance, the choice of i at time t, or the long-run average of his choices up to time t (see Hart and Mas-Colell (2003b, footnote 3)).

To state the condition of "uncoupledness," fix the set of players N and the action spaces $S^1, S^2, \ldots, S^N$; a game Γ is thus given by its payoff functions $u^1, u^2, \ldots, u^N$. We consider a family of games $\mathcal{U}$ (formally, a family of N-tuples of payoff functions $(u^1, u^2, \ldots, u^N)$). A general dynamic (9) is thus

$$\dot{x}^i(t) = F^i\big(x(t); u^1, u^2, \ldots, u^N\big) \quad \text{for each } i \in N.$$

We will call the dynamic $F = (F^1, F^2, \ldots, F^N)$ *uncoupled* on $\mathcal{U}$ if each F^i depends on the game Γ only through the payoff function u^i of player i, i.e.,

$$\dot{x}^i(t) = F^i(x(t); u^i) \quad \text{for each } i \in N.$$

That is, let Γ and Γ' be two games in the family $\mathcal{U}$ for which the payoff function of player i is the same (i.e., $u^i(\Gamma) = u^i(\Gamma')$); uncoupledness requires that, if the current state of play of all players, $x(t)$, is the same, then player i will adapt his action in the same way in the two games Γ and Γ'.

To study the impact of uncoupledness, we will deal with games that have *unique* Nash equilibria; this eliminates difficulties of coordination (different players may converge to different Nash equilibria). If there are dynamics that always converge to the set of Nash equilibria, we can apply them in particular when there is a unique Nash equilibrium.

We thus consider families of games $\mathcal{U}$ such that each game Γ in $\mathcal{U}$ possesses a unique Nash equilibrium, which we denote $\bar{x}(\Gamma)$. A dynamic F is *Nash-convergent* on $\mathcal{U}$ if, for each game Γ in $\mathcal{U}$, the unique Nash equilibrium $\bar{x}(\Gamma)$ is a rest-point of the dynamic (i.e., $F(\bar{x}(\Gamma); \Gamma) = 0$), which is moreover stable for the dynamic (i.e., $\lim_{t\to\infty} x(t) = \bar{x}(\Gamma)$ for any solution $x(t)$ of (8); some regularity assumptions are used here to facilitate the analysis).

The result of Hart and Mas-Colell (2003b) is:

THEOREM 3—Uncoupled Dynamics: *There exist no uncoupled dynamics that guarantee Nash convergence*.

It is shown that there are simple families of games $\mathcal{U}$ (in fact, arbitrarily small neighborhoods of a single game), such that every uncoupled dynamic on $\mathcal{U}$ is *not* Nash-convergent; i.e., the unique Nash equilibrium is unstable for every

uncoupled dynamic. The properties of uncoupledness and Nash-convergence are thus incompatible, even on simple families of games (and thus, a fortiori, on any larger families).

It follows that there can be no uncoupled dynamics that always converge to the set of Nash equilibria, or, for that matter, to the convex hull of Nash equilibria (which is the set of publicly correlated equilibria), since, in our games, both sets consist of the single Nash equilibrium.

The result of Theorem 3 indicates why dynamics that are to some extent "adaptive" or "rational" *cannot* always lead to Nash equilibria (see the references in Hart and Mas-Colell (2003b, Section IV(d))). In contrast, correlated equilibria may be obtained by uncoupled dynamics, such as regret matching and the other adaptive heuristics of this paper.[20]

9. SUMMARY

Our results can be summarized as follows:

1. *There* are *simple adaptive heuristics that always lead to correlated equilibria* (the Regret Matching Theorem in Section 4).

2. *There is a* large *class of adaptive heuristics that always lead to correlated equilibria* (the Generalized Regret Matching Theorem in Section 7).

3. *There can be* no *adaptive heuristics that always lead to Nash equilibria, or to the convex hull of Nash equilibria* (the Uncoupled Dynamics Theorem in Section 8).

Taken together, these results establish a solid connection between the dynamic approach of adaptive heuristics and the static approach of correlated equilibria.

From a more general viewpoint, the results show how simple and far-from-rational behavior in the short run may well lead to fully rational outcomes in the long run. Adaptive heuristics are closely related to behavioral models of what people do, whereas correlated equilibria embody fully rational considerations (see Sections 5 and 6, respectively). Our results show that rational behavior, which has been at times quite elusive and difficult to exhibit in single acts, may nevertheless be obtained in the long run.[21]

In short, adaptive heuristics may serve as a natural bridge connecting "behavioral" and "rational" approaches.

[20]This suggests a "Coordination Conservation Law": some form of coordination must be present, either in the equilibrium concept (such as correlated equilibrium) or, if not (as in the case of Nash equilibrium), then in the dynamics leading to it (see Hart and Mas-Colell (2003b)). As a further illustration, consider the learning dynamics of Section 2.1; while they are usually uncoupled, the convergence to Nash equilibria is obtained there only under certain initial conditions—such as "beliefs that contain a grain of truth"—which are in fact a form of coordination.

[21]Aumann, in various lectures since the late nineties (e.g., Aumann (1997); see also his interview in Hart (2005)), has argued that rationality should be examined in the context of *rules* rather

9.1. *Directions of Research*

There are many interesting questions that arise in connection with this research. We will mention a few.

First, we need to further understand the relations between dynamics and equilibria. Which equilibria are obtained from adaptive heuristics and which are not? At this point, we only know that the joint distribution of play converges to the *set* of correlated equilibria, and that it converges to a *point* only when it is a pure Nash equilibrium.[22] We do not know if *all* correlated equilibria are obtained from adaptive heuristics, or if only a strict subset of them are; recall that the Uncoupled Dynamics Theorem implies that the set of limit points can be neither the set of Nash equilibria nor its convex hull. A more refined question is to characterize which dynamics lead to which equilibria. Finally, one needs to understand the behavior of these dynamics not just in the limit, but also along the way.

Second, alternative notions of regret—in particular, those that are obtained by different ways of time-averaging, like discounting, and finite recall or memory—should be analyzed. For such an analysis of approachability, see Lehrer and Solan (2003).

Third, one needs to strengthen the ties between the behavioral, experimental, and empirical approaches on the one hand, and the theoretical and rational approaches on the other. Adaptive heuristics that arise from theoretical work may be tested in practice, and theoretical work may be based on the empirical findings.

Fourth, some of the focus needs to be shifted from Nash equilibria to the more general class of correlated equilibria—in both static and dynamic setups. Problems of coordination, correlation, and communication have to be studied extensively.

Finally, we emphasize again (see Section 4.1) that looking at what each player does separately—i.e., considering the mixed actions independently—misses much relevant information; one needs to look at the *joint* distribution of play.[23]

10. ADDITIONAL RESULTS

This final section is devoted to a number of additional results and discussions of related issues.

---

than *acts*; "rational rules" (i.e., rules of behavior that are best when compared to other rules) may well lead to single acts that are not rational. Here, we argue that rationality should be examined also *in the long run*; single acts that are not rational may nevertheless generate long-run behavior that is rational.

[22] See Hart and Mas-Colell (2000, p. 1132, comment (4)).

[23] Unfortunately, almost all the experimental and behavioral literature deals only with the marginal distributions. (Two books where the joint distribution appears are Suppes and Atkinson (1960) and Rapoport, Guyer, and Gordon (1976).)

10.1. *Hannan Consistency and the Hannan Set*

The regret for action k has been defined relative to the previous period's action j. One may consider a rougher measure instead: the increase in average payoff, if any, were one to replace *all* past plays, and not just the j-plays, by k. We thus define the *unconditional regret* for action k as

$$\widetilde{R}(k) := [\widetilde{V}(k) - U]_+, \tag{10}$$

where

$$\widetilde{V}(k) := \frac{1}{T}\sum_{t=1}^{T} u^i(k, s_t^{-i}) \tag{11}$$

and U is the average payoff (see (1)). *Unconditional regret matching* prescribes play probabilities at each period that are directly proportional to the vector of unconditional regrets; i.e.,

$$\sigma_{T+1}(k) := \frac{\widetilde{R}(k)}{\sum_{\ell=1}^{m} \widetilde{R}(\ell)} \quad \text{for each } k = 1, 2, \ldots, m$$

(of course, this applies only when there is some positive unconditional regret, i.e., $\widetilde{R}(k) > 0$ for some k; σ_{T+1} is arbitrary otherwise). Unlike with regret matching, here we do not use a constant proportionality factor c, but rather normalize the vector of unconditional regrets to get a probability vector.

A strategy of player i is said to be *Hannan-consistent* (following Hannan (1957)[24]) if it guarantees, for any strategies of the other players, that all the unconditional regrets of i become nonnegative in the limit, i.e., $\widetilde{R}(k) \to 0$ (almost surely) as $T \to \infty$ for all $k = 1, 2, \ldots, m$. We have:

PROPOSITION 4: *Unconditional regret matching is Hannan-consistent. Moreover, if all players play unconditional regret matching, then the joint distribution of play converges to the Hannan set of the stage game.*

Proposition 4 is Theorem B in Hart and Mas-Colell (2000). The proof applies Blackwell's Approachability Theorem 5 (see the Appendix) to the m-dimensional vector of unconditional regrets $(\widetilde{R}(1), \widetilde{R}(2), \ldots, \widetilde{R}(m))$: the negative orthant is shown to be approachable, and unconditional regret matching is the corresponding Blackwell strategy.

The *Hannan set* (see Hart and Mas-Colell (2003a) and Moulin and Vial (1978)), like the set of correlated equilibria, consists of joint distributions of

[24] Fudenberg and Levine (1995) call this "universal consistency."

play (i.e., it is a subset of $\Delta(S)$).[25] In contrast to correlated equilibria, the requirement now is that no player can gain unilaterally by playing a *constant* action (regardless of his signal). The set of correlated equilibria is contained in the Hannan set (and the two sets coincide when every player has at most two strategies); moreover, the Hannan distributions that are independent across players are precisely the Nash equilibria of the game.

Hannan-consistent strategies have been constructed by Hannan (1957), Blackwell (1956b) (see also Luce and Raiffa (1957, pp. 482–483)), Foster and Vohra (1993, 1998), Fudenberg and Levine (1995), and Freund and Schapire (1999).[26] Many of these strategies are smoothed-out variants of fictitious play, which, by itself, is not Hannan-consistent; see Section 10.4.[27] For a general class of Hannan-consistent strategies, which includes the unconditional regret matching of Proposition 4—apparently the simplest Hannan-consistent strategy—as well as smooth fictitious play, see Section 10.3.

10.2. *Regret Eigenvector Strategies*

Returning to our ("conditional") setup where regrets are defined relative to the previous period's action, Blackwell's approachability (see the Appendix) leads to the following construction (see Hart and Mas-Colell (2000, Section 3)).[28] Start by defining the *regret* $R(j,k)$ *from* j *to* k using the same formulas (1)–(4) of Section 4 for *every* pair $j \neq k$ (i.e., j need no longer be the previous period's action). Take as payoff vector the $m(m-1)$-dimensional vector of *signed regrets* (i.e., *before* taking the positive part $[\cdot]_+$ in (4)). The negative orthant turns out to be approachable, and the Blackwell strategy translates into playing at each stage a randomized action σ that satisfies

$$\sum_{k:k\neq j} \sigma(j)R(j,k) = \sum_{k:k\neq j} \sigma(k)R(k,j) \quad \text{for all } j = 1,2,\ldots,m. \tag{12}$$

That is, σ is a "regret-invariant vector": for all j, the average regret *from* j equals the average regret *to* j. Equivalently, put $q(j,k) := cR(j,k)$ for $j \neq k$ and $q(j,j) := 1 - \sum_{k:k\neq j} q(j,k)$, where $c > 0$ is large enough to guarantee that $q(j,j) > 0$ for all j; then $Q = (q(j,k))_{j,k=1,2,\ldots,m}$ is a stochastic (or Markov) matrix and (12) is easily seen to be equivalent to $\sigma = \sigma Q$. Thus σ is a left eigenvector of Q (corresponding to the eigenvalue 1); or, regarding Q as the one-step transition probability matrix of a Markov chain, σ is an invariant vector of Q.

[25] Hannan-consistency is a property of strategies in the *repeated* game, whereas the Hannan set is a concept defined for the *one-shot* game.

[26] See also the references in Hart and Mas-Colell (2001a, footnote 6) for related work.

[27] The original strategy of Hannan (1957) essentially uses at each stage an average of the best replies to a small neighborhood of the distribution of the opponents' past play.

[28] Interestingly, similar "regrets" (on probabilistic forecasts) as well as formula (12) appear in the earlier work—not based on approachability—of Foster and Vohra (1998); see Section 10.9.

We will call a strategy satisfying (12) a *regret eigenvector* strategy. By comparison, regret matching is defined by $\sigma(k) = q(j,k)$, which amounts to using Q as a Markov one-step transition probability matrix (from the previous period's action j to the current period's action k).[29]

Hart and Mas-Colell (2000, Corollary to Theorem A) show that if every player plays a regret eigenvector strategy, then, again, the joint distribution of play converges almost surely to the set of correlated equilibria. While the proof of this result is much simpler than that of Theorem 1, we do not regard the regret eigenvector strategies as heuristics (since they require computing each period an eigenvector of a matrix Q that moreover changes over time).

10.3. *Generalized Regret Matching Strategies*

It is convenient to consider first the "unconditional" Hannan setup of Section 10.1. Hart and Mas-Colell (2001a) characterize the class of *generalized unconditional regret* strategies, which are Hannan-consistent, as follows. For each $k = 1, 2, \ldots, m$ there is a function f_k defined on the m-dimensional vector of signed regrets $x = (\widetilde{V}(1) - U, \widetilde{V}(2) - U, \ldots, \widetilde{V}(m) - U)$, such that f_k is continuous, $\sum_{k=1}^{m} x_k f_k(x) > 0$ for all $x \not\leq 0$, and the vector of functions $f = (f_1, f_2, \ldots, f_m)$ is integrable (i.e., there exists a continuously differentiable function $P : \mathbb{R}^m \to \mathbb{R}$, a "potential function," such that $f_k = \partial P / \partial x_k$ for all k, or f is the gradient ∇P of P). Finally, the strategy is given by $\sigma(k) = f_k(x) / \sum_{\ell=1}^{m} f_\ell(x)$ for each $k = 1, 2, \ldots, m$. (Unconditional regret matching is obtained when $P(x) = \sum_{k=1}^{m} ([x_k]_+)^2$.)[30] The proof is based on characterizing *universal* approachability strategies.

Next, in the "conditional" setup, the generalization proceeds in two steps. The first step yields *generalized regret eigenvector* strategies, which are obtained by replacing each $R(j,k)$ in (12) with a function $f_{j,k}(R)$ defined on the $m(m-1)$-dimensional vector of (signed) regrets R, such that the vector of $m(m-1)$ functions $(f_{j,k})_{j \neq k}$ is the gradient ∇P of a continuously differentiable potential function P with $x \cdot \nabla P(x) > 0$ for all $x \not\leq 0$ (again, these strategies are too complex to be viewed as heuristics). In the second step, we dispose of the computation of eigenvectors in (12) and get the full class of *generalized regret matching* strategies: $\sigma(k) = c f_{j,k}(R)$ for $k \neq j$ and $\sigma(j) = 1 - \sum_{k\,:\,k \neq j} \sigma(k)$, where j is the previous period's action. (The strategies given by (7) in Section 7 correspond to the "separable" special case where $P(x) = \sum \sum_{j \neq k} F(x_{j,k})$ and $F = \int f$.) If every player uses such a strategy, convergence to the set of correlated equilibria is obtained. For precise statements and proofs, see Hart and Mas-Colell (2001a, Section 5.1) and Cahn (2004, Theorem 4.1).

[29] If Q were constant over time, the Ergodic Theorem for Markov chains would imply that regret matching and regret eigenvector strategies lead to essentially the same long-run distribution of play. The proof of the Regret Matching Theorem in Hart and Mas-Colell (2000) shows that this also holds when Q changes over time as a function of the regrets.

[30] See Sandholm (2004) for related work in an evolutionary setup.

10.4. *Fictitious Play and Variants*

Fictitious play is an extensively studied adaptive heuristic; it prescribes playing at each period a best reply to the distribution of the past play of the opponents. Now the action k is such a best reply if and only if k maximizes $\widetilde{V}(k)$ or, equivalently, the signed unconditional regret $\widetilde{V}(k) - U$ (see (11) and (10)). Thus fictitious play turns out to be a function of the regrets too; however, since choosing a maximizer does not yield a continuous function, fictitious play does not belong to the class of unconditional regret-based strategies of Section 10.3—and it is indeed not Hannan-consistent. Therefore some smoothing out is needed, as in the strategies mentioned at the end of Section 10.1.

Similarly, *conditional fictitious play* consists of playing at each period a best reply to the distribution of the play of the opponents in those periods where i played the same action j as in the previous period. Smoothing this out yields *smooth fictitious play eigenvector* strategies (see Fudenberg and Levine (1998, 1999a)) and *smooth conditional fictitious play* (see Cahn (2004, Section 4.5)), which lead to the set of correlated approximate equilibria; for a discussion of the reason that one gets only approximate equilibria, see Hart and Mas-Colell (2001a, Section 4.1).

10.5. *The Case of the Unknown Game*

Consider now the case where player i knows initially only his set of actions S^i, and is informed, after each period of play, of his realized payoff.[31] He does not know what game he is playing: how many players there are and what their actions and payoffs are. In particular, he does not know his own payoff function—but only the payoffs he did actually receive every period. Thus at time $T+1$ he knows the T numbers $u^i(s_1), u^i(s_2), \ldots, u^i(s_T)$; in addition, he recalls what he did in the past (i.e., his actual actions, $s_1^i, s_2^i, \ldots, s_T^i$ in S^i, and the probabilities that he used, $\sigma_1^i, \sigma_2^i, \ldots, \sigma_T^i$ in $\Delta(S^i)$). This is essentially a standard stimulus-response setup.

At each period the player can compute his realized average payoff U, but he cannot compute his regrets $R(k)$ (see (2) and (4)): he knows neither what the other players did (i.e., s_t^{-i}) nor what his payoff would have been had he played k instead (i.e., $u^i(k, s_t^{-i})$). We therefore define the *proxy regret* $\widehat{R}(k)$ for action k by using the payoffs he got when he did actually play k:

$$\widehat{R}(k) := \left[\frac{1}{T} \sum_{t \le T : s_t^i = k} \frac{\sigma_t^i(j)}{\sigma_t^i(k)} u^i(s_t) - \frac{1}{T} \sum_{t \le T : s_t^i = j} u^i(s_t) \right]_+$$

(the normalizing factor $\sigma_t^i(j)/\sigma_t^i(k)$ is needed, roughly speaking, to offset the possibly unequal frequencies of j and k being played in the past).

[31] Following a suggestion of Dean Foster; see Foster and Vohra (1993).

In Hart and Mas-Colell (2001b) it is shown that convergence to correlated approximate equilibria is obtained also for *proxy regret matching* strategies.

10.6. *Continuous Time*

The regret-based dynamics up to this point have been *discrete-time* dynamics: the time periods were $t = 1, 2, \ldots$. It is natural to study also *continuous-time* models, where the time t is a continuous variable and the change in the players' actions is governed by appropriate differential equations. It turns out that the results carry over to this framework (in fact, some of the proofs become simpler). See Hart and Mas-Colell (2003a) for details.

10.7. *Computing Regrets*

The regrets, despite depending on the whole history, are easy to compute. A player needs to keep record only of his $m(m-1)$ signed regrets $D_T(j,k)$ (one for each $j \neq k$) and the "calendar" time T. The updating is simply $D_T(j,k) = (1-1/T)D_{T-1}(j,k) + (1/T)(u^i(k, s_T^{-i}) - u^i(s_T))$ for $j = s_T^i$ and $D_T(j,k) = (1-1/T)D_{T-1}(j,k)$ for $j \neq s_T^i$, and the regrets at time T are $R_T(k) = [D_T(s_T^i, k)]_+$.

10.8. *Convergence*

The convergence of the joint distributions play z_T to the set of correlated equilibria CE, i.e., $z_T \to \mathrm{CE}$ (a.s.) as $T \to \infty$, means that[32]

$$\mathrm{dist}(z_T, \mathrm{CE}) \underset{T\to\infty}{\longrightarrow} 0 \quad \text{(a.s.)}.$$

That is, the sequence z_T eventually enters any neighborhood of the set CE, and stays there forever: for every $\varepsilon > 0$ there is a time $T_0 \equiv T_0(\varepsilon)$ such that for each $T > T_0$ there is a correlated equilibrium within ε of z_T; i.e., there is $\psi_T \in \mathrm{CE}$ with $\|z_T - \psi_T\| < \varepsilon$. Since the players randomize, all of the above are random variables, and all statements hold with probability 1 (i.e., for almost every history); in particular, T_0 and ψ_T depend on the history.

An equivalent way of stating this is as follows. Given $\varepsilon > 0$, there is a time $T_1 \equiv T_1(\varepsilon)$ after which the joint distribution of play is always a correlated ε-equilibrium; i.e., z_T satisfies the correlated equilibrium constraints (see (6)) within ε, for all $T > T_1$.

As for the rate of convergence, it is essentially of the order of $1/\sqrt{T}$; see the Proof of the Approachability Theorem in the Appendix and, for more precise recent bounds, Cesa-Bianchi and Lugosi (2003) and Blum and Mansour (2005).

[32]The distance between a point x and a set A is $\mathrm{dist}(x, A) := \inf_{a\in A} \|x - a\|$.

10.9. *A Summary of Strategies*

At this point the reader may well be confused by the plethora of regret-based strategies that have been presented above. We therefore provide a summary, with precise references, in Table I.

An important additional dynamic leading to correlated equilibria is the *calibrated learning* of Foster and Vohra (1997). Here each player computes "calibrated forecasts" on the behavior of the other players, and then plays a best reply to these forecasts. Forecasts are *calibrated* if, roughly speaking, the probabilistic forecasts and the long-run frequencies are close: for example, it must have rained on approximately 75% of all days for which the forecast was "a 75% chance of rain" (and the same holds when replacing 75% with any other percentage). There are various ways to generate calibrated forecasts; see

TABLE I

REGRET-BASED STRATEGIES

Correlated Equilibria (Conditional Setup)	Hannan Consistency (Unconditional Setup)
Regret matching[a] Regret eigenvector[b]	Unconditional regret matching[c]
Generalized regret matching[d] Generalized regret eigenvector[e]	Generalized unconditional regret matching[f]
Conditional fictitious play[g]	Fictitious play[h]
Smooth conditional fictitious play[i] Smooth fictitious play eigenvector[j]	Smooth fictitious play[k]
Proxy regret matching[l]	Proxy unconditional regret matching[m]
Continuous-time regret matching[n]	Continuous-time unconditional regret matching[o]

[a] Section 4; Hart and Mas-Colell (2000, Main Theorem).
[b] Section 10.2; Hart and Mas-Colell (2000, Theorem A).
[c] Section 10.1; Hart and Mas-Colell (2000, Theorem B).
[d] Sections 7 and 10.3; Hart and Mas-Colell (2001a, Section 5.1), Cahn (2004, Theorem 4.1).
[e] Section 4; Hart and Mas-Colell (2001a, Section 5.1).
[f] Section 10.3; Hart and Mas-Colell (2001b, Theorem 3.3).
[g] Section 10.4; it does *not* converge to the set of correlated equilibria.
[h] Section 10.4; it is *not* Hannan-consistent.
[i] Section 10.4; Cahn (2004, Proposition 4.3).
[j] Section 10.4; Fudenberg and Levine (1998, 1999a).
[k] Section 10.4; Fudenberg and Levine (1995).
[l] Section 10.5; Hart and Mas-Colell (2001b).
[m] Section 10.5; Hart and Mas-Colell (2000, Section 4(j); 2001a, Section 5.3).
[n] Section 10.6; Hart and Mas-Colell (2003a).
[o] Section 10.6; Hart and Mas-Colell (2003a).

Foster and Vohra (1997, 1998, 1999), Foster (1999), Fudenberg and Levine (1999b), and Kakade and Foster (2004).[33]

There is also a significant body of work in the computer science literature (where conditional regrets are called "internal regrets" and unconditional ones, "external"), with connections to machine learning, on-line prediction, experts, classification, perceptrons, and so on. For a recent study (and earlier references), see Cesa-Bianchi and Lugosi (2003).

10.10. *Variable Game*

Even if the one-shot game changes every period (as in stochastic games), our results—that the regrets converge to zero—continue to hold, provided that the payoffs are uniformly bounded and the players are told at the end of each period which game has been played. This follows easily from our proofs (replace u^i by u^i_t throughout) and is related to the "universality" of the regret-based strategies; see Fudenberg and Levine (1998, Chapter 4, footnote 19) and Hart and Mas-Colell (2001a, Section 5.2).

10.11. *The Set of Correlated Equilibria*

Correlated equilibria always exist in finite games. This follows from the existence of Nash equilibria (which requires fixed-point arguments), or directly (by linear duality arguments; see Hart and Schmeidler (1989)[34] and Nau and McCardle (1990)).

A natural question is, how large (or small) is the set of correlated equilibria? An interesting result in this direction is provided by Keiding and Peleg (2000). Fix the number of players N, the action sets $S^1, S^2, \ldots, S^N$, and a bound on the possible payoffs M. If one chooses at random (uniformly) a game Γ and a joint distribution of play $z \in \Delta(S)$, then the probability that z is a correlated equilibrium of Γ is at most $1/2^N$ (which goes to zero as N increases).

There is also work on the structure of the set of correlated equilibria; see Evangelista and Raghavan (1996), Myerson (1997), Nau, Gomes Canovas, and Hansen (2004), Calvó-Armengol (2004), and Nitzan (2005). For some recent results on the computation of correlated equilibria, see Kakade, Kearns, Langford, and Ortiz (2003) and Papadimitriou (2005).

[33]One construction, due to Hart and Mas-Colell, uses approachability to prove "no regret"; see Foster and Vohra (1999, Section 2). Interestingly, calibration is also closely related to the "merging of opinions" that arises in the study of the learning dynamics of Section 2.1; see Kalai, Lehrer, and Smorodinsky (1999).

[34]The starting point for the research presented here was the application of fictitious play to the auxiliary two-person zero-sum game of Hart and Schmeidler (1989); see Hart and Mas-Colell (2000, Section 4(i)).

Center for the Study of Rationality, Dept. of Economics, and Institute of Mathematics, The Hebrew University of Jerusalem, Feldman Building, Givat Ram Campus, 91904 Jerusalem, Israel; hart@huji.ac.il; http://www.ma.huji.ac.il/hart.

Manuscript received October, 2004; final revision received May, 2005.

APPENDIX: APPROACHABILITY

A most useful technical tool in this area is the approachability theory originally introduced by Blackwell (1956a). The setup is that of games where the payoffs are *vectors* (rather than, as in standard games, scalar real numbers). For instance, the coordinates may represent different commodities; or contingent payoffs in different states of the world (when there is incomplete information—see Aumann and Maschler (1995, Section I.6 and postscript I.d)); or, as in the current setup, regrets for the various actions in a standard game.

Let thus $A : S \equiv S^i \times S^{-i} \to \mathbb{R}^m$ be the payoff function of player i, where $\mathbb{R}^m$ denotes the m-dimensional Euclidean space (thus $A(s^i, s^{-i}) \in \mathbb{R}^m$ is the payoff vector when player i chooses s^i and the other players, $-i$, choose s^{-i}), which is extended bilinearly to mixed actions, i.e., $A : \Delta(S^i) \times \Delta(S^{-i}) \to \mathbb{R}^m$. The time is discrete, $t = 1, 2, \ldots$, and let $s_t = (s_t^i, s_t^{-i}) \in S^i \times S^{-i}$ be the actions chosen by i and $-i$, respectively, at time t, with payoff vector $a_t := A(s_t)$; put $\overline{a}_T := (1/T) \sum_{t=1}^{T} a_t$ for the average payoff vector up to T.

Let $C \subset \mathbb{R}^m$ be a convex and closed set.[35] We define:

- The set C is *approachable* by player i (cf. Blackwell (1956a)[36]) if there exists a strategy of i such that, no matter what the opponents $-i$ do, $\operatorname{dist}(\overline{a}_T, C) \to 0$ almost surely as $T \to \infty$.

- The set C is *enforceable* by player i if there exists a mixed action σ^i in $\Delta(S^i)$ such that, no matter what the opponents $-i$ do, the one-shot vector payoff is guaranteed to lie in C; i.e., $A(\sigma^i, s^{-i}) \in C$ for all s^{-i} in S^{-i} (and so also $A(\sigma^i, \sigma^{-i}) \in C$ for all σ^{-i} in $\Delta(S^{-i})$).

Approachability is a notion in the long-run repeated game, whereas enforceability is a notion in the one-shot game.

We restate the result of Blackwell (1956a) as follows:

THEOREM 5—Approachability:
(i) *A half-space H is approachable if and only if it is enforceable.*

[35] For nonconvex sets, see Blackwell (1956a), Vieille (1992), and Spinat (2002).

[36] Blackwell's definition requires in addition that the approachability be *uniform* over the strategies of the opponents; namely, for every $\varepsilon > 0$ there is $T_0 \equiv T_0(\varepsilon)$ such that $E[\operatorname{dist}(\overline{a}_T, C)] < \varepsilon$ for all $T > T_0$ and all strategies of $-i$ (i.e., T_0 is independent of the strategy of $-i$). It turns out that for convex sets C this strengthening is always satisfied.

T	$(1,0)$
B	$(0,1)$

FIGURE 4.—A game with vector payoffs.

(ii) *A convex set C is approachable if and only if every half-space H containing C is approachable*.

The statement of Theorem 5 seems like a standard convexity result (based on the fact that a closed convex set is the intersection of the half-spaces containing it). This is however not so, since the intersection of approachable sets need not be approachable. For a simple example, consider the game of Figure 4, where player i has two actions: T yields the payoff vector $(1,0)$ and B yields $(0,1)$ (the opponent $-i$ has only one action). The half-space $\{x=(x_1,x_2)\in\mathbb{R}^2 : x_1 \geq 1\}$ is approachable (by playing T), and so is the half-space $\{x : x_2 \geq 1\}$ (by playing B)—whereas their intersection $\{x : x \geq (1,1)\}$ is clearly not approachable. What the Approachability Theorem says is that if *all* half-spaces containing the convex set C are approachable, then, and only then, their intersection C is approachable.

Let $H=\{x\in\mathbb{R}^m : \lambda\cdot x \geq \rho\}$ be a half-space, where $\lambda\neq 0$ is a vector in $\mathbb{R}^m$ and ρ is a real number. Consider the game $\lambda\cdot A$ with scalar payoffs given by $\lambda\cdot A(s^i, s^{-i})$ (i.e., the linear combination of the m coordinates with coefficients λ). Then H is enforceable if and only if the minimax value of this game, $\mathrm{val}(\lambda\cdot A)=\max\min \lambda\cdot A(\sigma^i,\sigma^{-i})=\min\max \lambda\cdot A(\sigma^i,\sigma^{-i})$, where the max is over $\sigma^i\in\Delta(S^i)$ and the min over $\sigma^{-i}\in\Delta(S^{-i})$, satisfies $\mathrm{val}(\lambda\cdot A)\geq\rho$. Given a convex set C, let φ be the "support function" of C, namely, $\varphi(\lambda):=\inf\{\lambda\cdot c : c\in C\}$ for all $\lambda\in\mathbb{R}^m$. Since, for every direction λ, only the minimal half-space containing C, i.e., $\{x : \lambda\cdot x\geq\varphi(\lambda)\}$, matters in (ii),[37] the Approachability Theorem may be restated as follows: C is approachable if and only if

$$\mathrm{val}(\lambda\cdot A)\geq\varphi(\lambda) \quad \text{for all } \lambda\in\mathbb{R}^m.$$

PROOF OF THEOREM 5: (i) If the half-space $H=\{x\in\mathbb{R}^m : \lambda\cdot x\geq\rho\}$ is enforceable then there exists a mixed action $\sigma^i\in\Delta(S^i)$ such that $\lambda\cdot A(\sigma^i,\sigma^{-i})\geq\rho$ for all $\sigma^{-i}\in\Delta(S^{-i})$. Player i, by playing σ^i every period, guarantees that $b_t:=E[a_t|h_{t-1}]$, the expected vector payoff conditional on the history h_{t-1} of the previous periods, satisfies $\lambda\cdot b_t\geq\rho$. Put $\overline{b}_T:=(1/T)\sum_{t=1}^T b_t$; then $\lambda\cdot\overline{b}_T\geq\rho$, or $\overline{b}_T\in H$; the Strong Law of Large Numbers[38] implies that almost

[37]Trivially, a superset of an approachable set is also approachable.

[38]The version used in this proof is: $(1/T)\sum_{t=1}^T(X_t-E[X_t|h_{t-1}])\to 0$ almost surely as $T\to\infty$, where the X_t are uniformly bounded random variables; see Loève (1978, Theorem 32.1.E). While

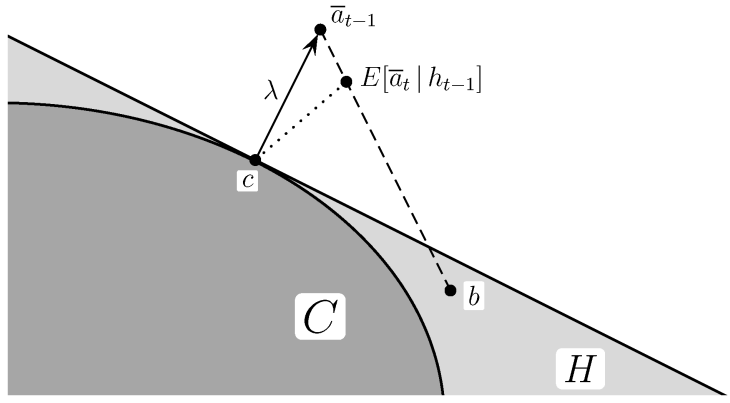

FIGURE 5.—The Blackwell strategy for the set C.

surely $\overline{a}_T - \overline{b}_T \to 0$, and therefore $\overline{a}_T \to H$, as $T \to \infty$. Conversely, if H is not enforceable then $\mathrm{val}(\lambda \cdot A) < \rho$. Therefore, given a strategy of player i, for every history h_{t-1} the other players can respond to player i's mixed action at t so that[39] $\lambda \cdot b_t \le v$; hence $\lambda \cdot \overline{b}_T \le v < \rho$ and $\overline{a}_T$ cannot converge to H.

(ii) The condition that every half-space containing C be approachable is clearly necessary. To see that it is also sufficient, for every history h_{t-1} such that $\overline{a}_{t-1} \notin C$, let $c \in C$ be the closest point to $\overline{a}_{t-1}$, and so $\delta_{t-1} := [\mathrm{dist}(\overline{a}_{t-1}, C)]^2 = \|\overline{a}_{t-1} - c\|^2$. Put $\lambda := \overline{a}_{t-1} - c$ and $H := \{x : \lambda \cdot x \le \lambda \cdot c\}$; then $C \subset H$ (since C is a convex set and λ is orthogonal to its boundary at c; see Figure 5). By assumption the half-space H is approachable, and thus enforceable; the *Blackwell strategy* prescribes to player i to play a mixed action $\sigma^i \in \Delta(S^i)$ corresponding to this[40] H—and so the expected payoff vector $b := E[a_t \mid h_{t-1}]$ satisfies $b \in H$, or[41] $\lambda \cdot b \le \lambda \cdot c$. Now

$$\delta_t = [\mathrm{dist}(\overline{a}_t, C)]^2 \le \|\overline{a}_t - c\|^2 = \left\| \left(\frac{t-1}{t} \overline{a}_{t-1} + \frac{1}{t} a_t \right) - c \right\|^2$$
$$= \frac{(t-1)^2}{t^2} \|\overline{a}_{t-1} - c\|^2 + \frac{2(t-1)}{t^2} (\overline{a}_{t-1} - c) \cdot (a_t - c) + \frac{1}{t^2} \|a_t - c\|^2.$$

player i's actions constitute an independent and identically distributed sequence, those of the other players may well depend on histories—and so we need a Strong Law of Large Numbers for *dependent* random variables (essentially, a Martingale Convergence Theorem).

[39]These responses may be chosen pure—which shows that whether or not the opponents $-i$ can correlate their actions is irrelevant for approachability; see Hart and Mas-Colell (2001a, footnote 12).

[40]When $\overline{a}_{t-1} \in C$ take $\lambda = 0$ and an arbitrary σ^i.

[41]As can be seen in Figure 5, it follows that (the conditional expectation of) $\overline{a}_t$ is closer to C than $\overline{a}_{t-1}$ is. The computation below will show that the distance to C not only decreases but, in fact, converges to zero.

Taking expectation conditional on h_{t-1} yields in the middle term $\lambda \cdot (b - c)$, which is ≤ 0 by our choice of σ^i, and so

$$E[t^2\delta_t|h_{t-1}] \leq (t-1)^2\delta_{t-1} + M^2 \tag{13}$$

for some bound[42] M. Taking overall expectation and then using induction implies that $E[t^2\delta_t] \leq M^2 t$; hence $E[\text{dist}(\overline{a}_t, C)] = E[\sqrt{\delta_t}] \leq \sqrt{E[\delta_t]} \leq M/\sqrt{t}$ and so $\overline{a}_t \to C$ in probability.

To get almost sure convergence,[43] put $\zeta_t := t\delta_t - (t-1)\delta_{t-1}$. Then $E[\zeta_t|h_{t-1}] \leq -(1 - 1/t)\delta_{t-1} + M^2/t \leq M^2/t \to 0$ (by (13)) and $|\zeta_t| \leq M$ (since $|\delta_t - \delta_{t-1}| \leq \|\overline{a}_t - \overline{a}_{t-1}\| = (1/t)\|a_t - \overline{a}_{t-1}\|$), from which it follows that $\delta_T = (1/T)\sum_{t=1}^{T} \zeta_t \to 0$ almost surely (by the Strong Law of Large Numbers; see footnote 38). *Q.E.D.*

Historically, Blackwell (1956b) used the Approachability Theorem to provide an alternative proof of the Hannan (1957) result (see Rustichini (1999) for an extension); the proof was indirect and nonconstructive. The direct application of approachability to regrets, with the vector of regrets as payoff vector and the negative orthant as the approachable set, was introduced in the 1996 preprint of Hart and Mas-Colell (2000) (see Section 3 there). This eventually led to the simple regret matching strategy of Section 4, to the universal approachability strategies and the generalized regret matching of Section 7, and to the other results presented here—as well as to various related uses of approachability (for example, Foster (1999), Foster and Vohra (1999), Lehrer (2001, 2002, 2003), Sandroni, Smorodinsky, and Vohra (2003), Cesa-Bianchi and Lugosi (2003), Greenwald and Jafari (2003), and Lehrer and Solan (2003)).

REFERENCES

AUMANN, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

——— (1997): "Rationality and Bounded Rationality," in *Cooperation: Game Theoretic Approaches*, ed. by S. Hart and A. Mas-Colell. Berlin: Springer-Verlag, 219–232; (1997), *Games and Economic Behavior*, 21, 2–14.

AUMANN, R. J., AND M. MASCHLER (1995): *Repeated Games of Incomplete Information*. Cambridge, MA: MIT Press.

BELL, D. E. (1982): "Regret in Decision Making under Uncertainty," *Operations Research*, 30, 961–981.

BLACKWELL, D. (1956a): "An Analog of the Minmax Theorem for Vector Payoffs," *Pacific Journal of Mathematics*, 6, 1–8.

[42]Take $M := \max_{b,b' \in B} \|b - b'\| + \max_{b \in B} \text{dist}(b, C)$, where B is the compact set $B := \text{conv}\{A(s) : s \in S\}$.

[43]See also Mertens, Sorin, and Zamir (1995, Proof of Theorem 4.3) or Hart and Mas-Colell (2001a, Proof of Lemma 2.2).

——— (1956b): "Controlled Random Walks," in *Proceedings of the International Congress of Mathematicians* 1954, Vol. III. Amsterdam: North-Holland, 335–338.

BLUM, A., AND Y. MANSOUR (2005): "From External Regret to Internal Regret," in *Learning Theory*, ed. by P. Auer and R. Meir. Berlin: Springer-Verlag, 621–636.

BUSH, R. R., AND F. MOSTELLER (1955): *Stochastic Models for Learning*. New York: John Wiley & Sons.

CAHN, A. (2004): "General Procedures Leading to Correlated Equilibria," *International Journal of Game Theory*, 33, 21–40.

CALVÓ-ARMENGOL, A. (2004): "The Set of Correlated Equilibria of 2×2 Games," Mimeo, Universitàt Autonoma de Barcelona.

CAMERER, C., AND T.-H. HO (1998): "Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation," *Journal of Mathematical Psychology*, 42, 305–326.

——— (1999): "Experience-Weighted Attraction Learning in Normal Form Games," *Econometrica*, 67, 827–874.

CAMILLE, N., G. CORICELLI, J. SALLET, P. PRADAT-DIEHL, J.-R. DUHAMEL, AND A. SIRIGU (2004): "The Involvement of the Orbitofrontal Cortex in the Experience of Regret," *Science*, 304, 1167–1170.

CESA-BIANCHI, N., AND G. LUGOSI (2003): "Potential-Based Algorithms in On-Line Prediction and Game Theory," *Machine Learning*, 51, 239–261.

DAVIES, N. B. (1978): "Territorial Defense in the Speckled Wood Butterfly *Pararge aegeria*: The Resident Always Wins," *Animal Behavior*, 26, 138–147.

EREV, I., AND A. E. ROTH (1998): "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review*, 88, 848–881.

EVANGELISTA, F., AND T. E. S. RAGHAVAN (1996): "A Note on Correlated Equilibrium," *International Journal of Game Theory*, 25, 35–41.

FOSTER, D. P. (1999): "A Proof of Calibration via Blackwell's Approachability Theorem," *Games and Economic Behavior*, 29, 73–78.

FOSTER, D. P., AND R. V. VOHRA (1993): "A Randomization Rule for Selecting Forecasts," *Operations Research*, 41, 704–709.

——— (1997): "Calibrated Learning and Correlated Equilibrium," *Games and Economic Behavior*, 21, 40–55.

——— (1998): "Asymptotic Calibration," *Biometrika*, 85, 379–390.

——— (1999): "Regret in the On-Line Decision Problem," *Games and Economic Behavior*, 29, 7–35.

FOSTER, D. P., AND H. P. YOUNG (2003a): "Learning, Hypothesis Testing, and Nash Equilibrium," *Games and Economic Behavior*, 45, 73–96.

——— (2003b): "Regret Testing: A Simple Payoff-Based Procedure for Learning Nash Equilibrium," Mimeo, University of Pennsylvania and Johns Hopkins University.

FREUND, Y., AND R. E. SCHAPIRE (1999): "Adaptive Game Playing Using Multiplicative Weights," *Games and Economic Behavior*, 29, 79–103.

FUDENBERG, D., AND D. KREPS (1988): "A Theory of Learning, Experimentation, and Equilibrium in Games," Mimeo, Stanford University.

——— (1993): "Learning Mixed Equilibria," *Games and Economic Behavior*, 5, 320–367.

FUDENBERG, D., AND D. LEVINE (1995): "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, 19, 1065–1090.

——— (1998): *Theory of Learning in Games*. Cambridge, MA: MIT Press.

——— (1999a): "Conditional Universal Consistency," *Games and Economic Behavior*, 29, 104–130.

——— (1999b): "An Easier Way to Calibrate," *Games and Economic Behavior*, 29, 131–137.

GERMANO, F., AND G. LUGOSI (2004): "Global Nash Convergence of Foster and Young's Regret Testing," Mimeo, Universitat Pompeu Fabra.

GREENWALD, A., AND A. JAFARI (2003): "A General Class of No-Regret Algorithms and Game-Theoretic Equilibria," in *Learning Theory and Kernel Machines*, Lecture Notes on Artificial Intelligence, Vol. 2777, ed. by J. G. Carbonell and J. Siekmann. Berlin: Springer-Verlag, 2–12.

HAMMERSTEIN, P., AND R. SELTEN (1994): "Game Theory and Evolutionary Biology," in *Handbook of Game Theory, with Economic Applications*, Vol. 2, ed. by R. J. Aumann and S. Hart. Amsterdam: Elsevier, 929–993.

HANNAN, J. (1957): "Approximation to Bayes Risk in Repeated Play," in *Contributions to the Theory of Games*, Vol. III, Annals of Mathematics Studies, Vol. 39, ed. by M. Dresher, A. W. Tucker, and P. Wolfe. Princeton, NJ: Princeton University Press, 97–139.

HARSANYI, J. C. (1967–1968): "Games with Incomplete Information Played by Bayesian Players," Parts I, II, III, *Management Science*, 14, 159–182, 320–334, 486–502.

HART, S. (2005): "An Interview with Robert Aumann," *Macroeconomic Dynamics*, forthcoming.

HART, S., AND A. MAS-COLELL (2000): "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica*, 68, 1127–1150.

——— (2001a): "A General Class of Adaptive Strategies," *Journal of Economic Theory*, 98, 26–54.

——— (2001b): "A Reinforcement Procedure Leading to Correlated Equilibrium," in *Economic Essays: A Festschrift for Werner Hildenbrand*, ed. by G. Debreu, W. Neuefeind, and W. Trockel. Berlin: Springer-Verlag, 181–200.

——— (2003a): "Regret-Based Dynamics," *Games and Economic Behavior*, 45, 375–394.

——— (2003b): "Uncoupled Dynamics Do Not Lead to Nash Equilibrium," *American Economic Review*, 93, 1830–1836.

——— (2004): "Stochastic Uncoupled Dynamics and Nash Equilibria," Mimeo, DP-371, Center for Rationality, The Hebrew University of Jerusalem.

HART, S., AND D. SCHMEIDLER (1989): "Existence of Correlated Equilibria," *Mathematics of Operations Research*, 14, 18–25.

HOFBAUER, J., AND W. H. SANDHOLM (2002): "On the Global Convergence of Stochastic Fictitious Play," *Econometrica*, 70, 2265–2294.

HOFBAUER, J., AND K. SIGMUND (1998): *Evolutionary Games and Population Dynamics*. Cambridge, U.K.: Cambridge University Press.

KAHNEMAN, D., P. SLOVIC, AND A. TVERSKY (EDS.) (1982): *Judgement under Uncertainty: Heuristics and Biases*, Cambridge, U.K.: Cambridge University Press.

KAKADE, S., AND D. P. FOSTER (2004): "Deterministic Calibration and Nash Equilibrium," in *Learning Theory*, ed. by J. Shawe-Taylor and Y. Singer. Berlin: Springer-Verlag, 33–48.

KAKADE, S., M. KEARNS, J. LANGFORD, AND L. ORTIZ (2003): "Correlated Equilibria in Graphical Games," *ACM Conference on Electronic Commerce 2003*. New York: ACM Press.

KALAI, E., AND E. LEHRER (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61, 1019–1045.

KALAI, E., E. LEHRER, AND R. SMORODINSKY (1999): "Calibrated Forecasting and Merging," *Games and Economic Behavior*, 29, 151–169.

KEIDING, H., AND B. PELEG (2000): "Correlated Equilibria of Games with Many Players," *International Journal of Game Theory*, 29, 375–389.

LEHRER, E. (2001): "Any Inspection is Manipulable," *Econometrica*, 69, 1333–1347.

——— (2002): "Approachability in Infinitely Dimensional Spaces," *International Journal of Game Theory*, 31, 255–270.

——— (2003): "A Wide Range No-Regret Theorem," *Games and Economic Behavior*, 42, 101–115.

LEHRER, E., AND E. SOLAN (2003): "No-Regret with Bounded Computational Capacity," Mimeo, Tel Aviv University.

LOÈVE, M. (1978): *Probability Theory*, Vol. II (Fourth Ed.). Berlin: Springer-Verlag.

LOOMES, G., AND R. SUGDEN (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal*, 92, 805–824.

LUCE, R. D., AND H. RAIFFA (1957): *Games and Decisions*. New York: John Wiley & Sons.
MERTENS, J.-F., S. SORIN, AND S. ZAMIR (1995): "Repeated Games," Part A, Mimeo, CORE DP-9420, Université Catholique de Louvain.
MOSHINSKY, A. (2003): "The Status-Quo Bias in Policy Judgements," Ph.D. Thesis, Mimeo, The Hebrew University of Jerusalem.
MOULIN, H., AND J. P. VIAL (1978): "Strategically Zero-Sum Games: The Class of Games whose Completely Mixed Equilibria Cannot Be Improved Upon," *International Journal of Game Theory*, 7, 201–221.
MYERSON, R. B. (1997): "Dual Reduction and Elementary Games," *Games and Economic Behavior*, 21, 183–202.
NAU, R., S. GOMEZ CANOVAS, AND P. HANSEN (2004): "On the Geometry of Nash Equilibria and Correlated Equilibria," *International Journal of Game Theory*, 32, 443–453.
NAU, R., AND K. F. MCCARDLE (1990): "Coherent Behavior in Noncooperative Games," *Journal of Economic Theory*, 50, 424–444.
NITZAN, N. (2005): "Tight Correlated Equilibrium," Mimeo, DP-394, Center for Rationality, The Hebrew University of Jerusalem.
NYARKO, Y. (1994): "Bayesian Learning Leads to Correlated Equilibria in Normal Form Games," *Economic Theory*, 4, 821–841.
PAPADIMITRIOU, C. H. (2005): "Computing Correlated Equilibria in Multi-Player Games," Mimeo, University of California at Berkeley.
RAPOPORT, A., M. J. GUYER, AND D. J. GORDON (1976): *The* 2×2 *Game*. Ann Arbor: University of Michigan Press.
ROTH, A. E., AND I. EREV (1995): "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8, 164–212.
RUSTICHINI, A. (1999): "Optimal Properties of Stimulus-Response Learning Models," *Games and Economic Behavior*, 29, 244–273.
SAMUELSON, W., AND R. ZECKHAUSER (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1, 7–59.
SANDHOLM, W. H. (2004): "Excess Payoff Dynamics, Potential Dynamics, and Stable Games," Mimeo, University of Wisconsin.
SANDRONI, A., R. SMORODINSKY, AND R. V. VOHRA (2003): "Calibration with Many Checking Rules," *Mathematics of Operations Research*, 28, 141–153.
SCHLAG, K. H. (1998): "Why Imitate, and If So, How? A Bounded Rational Approach to Multi-Armed Bandits," *Journal of Economic Theory*, 78, 130–156.
SHAPLEY, L. S. (1964): "Some Topics in Two-Person Games," in *Advances in Game Theory*, Annals of Mathematics Studies, Vol. 52, ed. by M. Dresher, L. S. Shapley, and A. W. Tucker. Princeton, NJ: Princeton University Press, 1–28.
SHMIDA, A., AND B. PELEG (1997): "Strict and Symmetric Correlated Equilibria Are the Distributions of the ESS's of Biological Conflicts with Asymmetric Roles," in *Understanding Strategic Interaction*, ed. by W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, and E. van Damme. Berlin: Springer-Verlag, 149–170.
SPINAT, X. (2002): "A Necessary and Sufficient Condition for Approachability," *Mathematics of Operations Research*, 27, 31–44.
SUPPES, P., AND R. C. ATKINSON (1960): *Markov Learning Models for Multiperson Interactions*. Palo Alto, CA: Stanford University Press.
VIEILLE, N. (1992): "Weak Approachability," *Mathematics of Operations Research*, 17, 781–791.
WEIBULL, J. W. (1995): *Evolutionary Game Theory*. Cambridge, U.K.: Cambridge University Press.
YOUNG, H. P. (1998): *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press.
——— (2004): *Strategic Learning and Its Limits*. Oxford, U.K.: Oxford University Press.