Learning in the Presence of Noise

Panayotis Mertikopoulos and Aris L. Moustakas

***Abstract*— We investigate the emergence of rationality in repeated games where, at each iteration, the players' payoffs are randombly perturbed (to account e.g. for the effects of fading or errors in the reading of one's throughput). We see that even if players start out completely uneducated about the game, there is a simple learning scheme that enables them to eventually weed out the noise and identify suboptimal choices, regardless of the noise level. More precisely, we show that strategies that are strictly dominated (even iteratively) become extinct in the long run, i.e. players exhibit rational behavior. As an application, we model a number of users that are able to switch dynamically between multiple wireless nodes and see that they are able to pick up which node works best for them, even in the presence of high performance fluctuations.**

I. Introduction

Ever since the seminal work of Maynard Smith on animal conflicts [1], there has been established a profound link between evolution and rationality: roughly speaking, one leads to the other. In this way, when species compete for the limited resources of their environment, evolution and natural selection steer the conflict to an equilibrium where species have essentially learnt to respect each other's boundaries and are loath to stray away from them (e.g. lest they suffer in terms of reproductive fitness). As a consequence, "fight or flight" responses that are deeply ingrained in a species can be seen as a form of rational behavior, emerging over the background of a species' evolutionary path.

By extending this analogy to networks, one sees that evolutionary schemes can yield substantial gains to interacting agents who need to adjust to their ever-changing local environment; after all, competition for the limited resources of a network is one of the most important problems faced by network designers. Still, this extension often depends on the accuracy of the data each user has about their environment and, given the finite time horizon for receiving this information in a rapidly evolving network, this data may be widely off the mark. For example, mobility in wireless networks introduces both fast and slow fading to the channel gains of each user and this, combined with overall bad channel conditions, will decrease the quality of the feedback data. If, on top of that, one adds the underlying competition between different users for the available resources of the system, the situation becomes quite complex.

In game-theoretic terms, one models this competition by introducing a suitable game whose payoffs reflect the users' reproductive fitness. Here, evolution takes the form of a selection mechanism that promotes strategies which perform better "on average" and one would hope that users are thus shepherded to a reasonable solution. This is precisely what happens in the model of *exponential learning* where players keep scores of their strategies (based on their returns) and employ the highest scoring one exponentially more often [2]. It then turns out that the evolution of the users' strategic choices is governed by the *replicator dynamics* of [3] and [4], which are an excellent conduit for rationality: in time, suboptimal strategies cease to be replicated and become extinct [5].

As a result, the many applications of evolutionary game theory to networks should not come as a surprise. To name but a few examples, in Aloha-type games the convergence properties of the replicator dynamics lead wireless users to an equilibrial state [6]. More recently, randomness has also been introduced in [7] by giving users revision opportunities and letting them switch strategies based on a Markov decision process that dictates whether to transmit and at what power. And, in the case of users who are able to switch dynamically between several wireless nodes, it was seen in [8] that if users process a broadcast signal, they actually converge to an efficient correlated equilibrium.

However, it is not at all clear if these results still hold in the presence of uncertainty which blurs the waters and may effectively "mask" the suboptimality of certain choices. Indeed, since the "state of the world" also changes as players learn to play the game, one should also take into account stochastic perturbations caused by nature's interference. In traditional evolutionary game theory, this is done by introducing "aggregate shocks" (weather-type effects) to the phenotype (species) populations, so as to account for births and deaths that are due to the fickle hand of nature. This approach of Fudenberg and Harris [9] has stirred a considerable amount of interest and many deterministic results have been translated to the stochastic setting as well. To wit, Cabrales showed in [10] that if the variance of the shocks is low enough, rational play still emerges, even if mutations are present. More recently, it was shown in [11] and [12] that even equilibrial play emerges in the long run: strict Nash equilibria are stochastically asymptotically stable in the case of a single population of players but, again, only if the hand of nature is soft enough. In high-noise environments, this is no longer the case: if the attraction of evolutionarily stable strategies is not sufficiently strong, the users' behavior becomes ergodic and the loud noise does not allow players to pick up the underlying game.

In the present paper, we take a different approach based

P. Mertikopoulos and A. L. Moustakas are with the Physics Department, National and Kapodistrean University of Athens, 157 84 Athens, Greece.

This research was supported in part by the EU projects NetReFound (EU-FET-FP6-IST-034413), Newcom++ (EU-IST-NoE-FP6-2007-216715) and by the University of Athens Research Council Project Kapodistrias (70/3/8831).

on the analogy between evolution and learning that we just outlined. So, instead of considering very large populations of distinct species, we will consider a game with a finite number of players who "evolve" thanks to their acquired experience in playing the game. As was mentioned before, if players keep a cumulative score of their strategies according to their payoff and employ more often the ones with the highest scores, this allows them to eventually drift away from dominated strategies. This then is the main question that concerns us: what happens if the players' learning curve is constantly perturbed as a result of random fluctuations to their strategies' payoffs?

Even though this approach seems closely related to the evolutionary one, the landscape actually changes dramatically. Indeed, we end up with a different stochastic replicator equation that is so robust as to allow rationality to emerge unconditionally and in complete generality: *irrespective of the noise level, only rationally admissible strategies survive in the long run*. As an immediate corollary, this shows that if players employ exponential learning in a game which can be solved by iterated elimination of dominated strategies (e.g. the Prisoner's Dilemma or its multiplayer variants), then *these players will converge to the game's (unique) Nash equilibrium*.

Outline: We begin our presentation in section II by presenting a simple wireless congestion scenario where multiple heterogeneous users seek to connect to one of several wireless nodes (perhaps belonging to different standards). There, numerical simulations reveal something even stronger than what we have hinted at so far: users quickly become rational and actually converge to a Nash equilibrium in pure strategies, even when the noise is much louder than their average payoff.[1] In section III we review a few basic facts and definitions from game theory in order to fix notation and terminology and subsequently, in section IV, we derive the stochastic replicator equation that emerges if the users' payoffs are randomly perturbed (and which differs from its other stochastic incarnations). Our main results are derived in section V, where we proceed to show that, ultimately, only rationally admissible strategies survive.

Notation and Conventions: If A is a set with finite cardinality n, we will identify the set $\Delta(A)$ of probability measures on A with the standard $(n-1)$-dimensional simplex of $\mathbb{R}^n$: $\Delta^{n-1} = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ and } \sum_i x_i = 1\}$. Also, in the interest of preserving indicial sanity, we will not differentiate between covariant and contravariant indices and we will only use subscripts. However, we will consistently employ Latin indices $(i, j, \ldots)$ for players and Greek $(\alpha, \beta, \ldots)$ for their strategies, separating the two by a semicolon if the need arises. Finally, for a given n, we will let e_k denote the standard basis vector $e_k = (0 \ldots 1 \ldots 0)$ of $\mathbb{R}^n$; still, when it is clear from the context that e_k refers to some pure strategy and there is no danger of confusion, we will sometimes simply write k instead of e_k.

[1]We address this issue of convergence to Nash equilibria in [13] .

II. The System Model: a Motivating Example

Seeing how our results revolve around arbitrary games, they are best presented (and proven) in an abstract setting. However, for concreteness, we focus here on a specific wireless scenario which consists of N wireless users (with similar transmission characteristics) that wish to connect to one of B nodes and can switch dynamically between them. The users' (selfish) objective is to maximize their throughput $u_{i\beta}$, $\beta = 1 \ldots, B$ which, in general, depends on the actions of other users as well as on the channel variations due to fading and other uncertainties.

In this endeavor, users switch between nodes and they keep track of each node's performance, trying to identify the one that works best for them. However, since there is no regulation or communication between the users, there is the very real problem that many users could switch at once. This invariably leads to congestion and, perhaps, even to "ping-pong" effects where a large group of users is locked in a vicious cycle as they simultaneously (and, from an outside perspective, irrationally) migrate from one node to the next, unable to coordinate their actions. Even worse, this situation is further exacerbated by the interference of nature; even in the absence of other users, the SNR and throughput of a single user are still stochastically perturbed.

Despite all that, one hopes that if users employ a sufficiently robust learning scheme, all this interference will "average out" and rational play will emerge. To that end, we first describe the users' scoring system:

$$U_{i\beta}(t+1) = U_{i\beta}(t) + T_{i\beta} + \eta_{i\beta} \tag{1}$$

i.e. at the t^{th} iteration user i "awards" node β with the throughput that it would have yielded if the user had selected it, modified by an "uncertainty" term $\eta_{i\beta}$ to account for stochastic shocks (due to fading, errors in reading one's throughput, etc.). Then, as this game is played again and again, user i selects node β with probability

$$p_{i\beta}(t+1) = \frac{e^{U_{i\beta}(t+1)}}{\sum_{\beta=1}^{B} e^{U_{i\beta}(t+1)}}. \tag{2}$$

To simulate this, we use for simplicity a specific form for the throughput of each user connected to node β:

$$T_{i\beta} = \frac{y_\beta}{N_\beta} \tag{3}$$

where N_β is the number of users connected to node β and y_β is a normalized ($\sum_\beta y_\beta = 1$) parameter that describes the "strength" of node β and encompasses the node's spectral efficiency, price charged per bit, etc.[2] The advantage of using this simple model is that it has a very elegant measure of the users' satisfaction with their choices, their *frustration*:

$$R = \frac{1}{N(B-1)} \sum_{\beta=1}^{B} \frac{1}{y_\beta} (N_\beta - y_\beta N)^2 \tag{4}$$

[2]Clearly, a much more elaborate model could be used but, for now, we prefer to keep things simple and on an intuitive level. After all, despite its simplicity, this model has been shown to be of the correct form for TCP and UDP protocols in IEEE 802.11b systems if we limit ourselves to a single class of users [14]; see also [8] for a relevant discussion.

This is just a (normalized) version of the distance of the users' distribution N_β from the Nash allocation of $y_\beta N$ users to node β. So, if the users experience no frustration ($R = 0$) their choices will not only be *rational* (in the sense that they are avoiding suboptimal nodes), but also socially stable: no user will have reason to deviate unilaterally [8].

Indeed, in figure 1 we see that the users' frustration eventually vanishes and users reach a Nash equilibrium. Clearly, in such a steady state suboptimal strategies cannot survive; hence, encouraged by these observations, we will devote the rest of this paper to show that, in any game, only rational strategies survive, regardless of the noise.

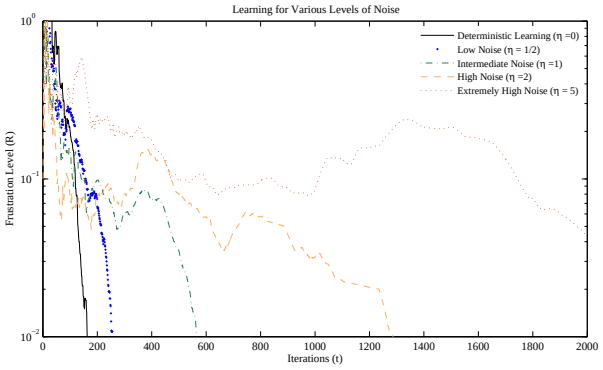

Fig. 1. Simulation of a wireless scenario where $N = 30$ users try to connect to one of $B = 3$ nodes with the help of the exponential learning scheme of equations (1) and (2). The users' throughput is perturbed by Gaussian white noise with variance η^2 and we plot the temporal evolution of the users' frustration (4) for $\eta = 0, 0.5, 1, 2, 5$ (the users' average throughput is normalized to 1). Even for very high noise levels, the users' frustration is minimized as they approach an equilibirum (for comparison, uneducated users would end up with an average frustration of $R = 1$).

III. Preliminaries

A. Basic Facts and Definitions from Game Theory

As is typical in game theory, our starting point will be a finite set of N *players*, indexed by $i \in \mathcal{N} := \{1, \dots N\}$. Each player comes with a (finite) set of *(pure) strategies* $\alpha \in \mathcal{S}_i := \{1, \dots S_i\}$, representing their possible actions when paired against one another. Naturally, players can "mix" these strategies by assigning different probabilities $p_{i\alpha}$ to every $\alpha \in \mathcal{S}_i$; in that case, their *mixed strategies* will be represented by the points $p_i = (p_{i;1} \dots p_{i;N}) \in \Delta_i := \Delta(\mathcal{S}_i)$ or, more succinctly, by the *strategy profile* $p = (p_1 \dots p_N) \in \Delta := \prod_i \Delta_i$. Alternatively, if we wish to focus on the strategy of a particular player i against his *opponents* $\mathcal{N}_{-i} := \mathcal{N} \backslash \{i\}$, we will use the standard shorthand $(p_{-i}; q)$ to stand for the profile $(p_1 \dots q \dots p_N)$ where i plays $q \in \Delta_i$ against his opponents' strategy $p_{-i} \in \Delta_{-i} := \prod_{j \neq i} \Delta_i$.

So, when the game is actually played, the players' choices are rewarded according to the *payoff functions* $u_i : \Delta \to \mathbb{R}$:

$$u_i(p) = \sum_{\alpha_1 \in \mathcal{S}_1} \cdots \sum_{\alpha_N \in \mathcal{S}_N} u_{i;\alpha_1 \dots \alpha_N} p_{1;\alpha_1} \cdots p_{N;\alpha_N} \tag{5}$$

where $u_{i;\alpha_1 \dots \alpha_N}$ is the payoff that the (pure) strategy $\alpha_i \in \mathcal{S}_i$ yields to player i when paired against the strategy $\alpha_{-i} \in \mathcal{S}_{-i} := \prod_{j \neq i} \mathcal{S}_j$ of i's opponents. Under this light, the payoff that a player receives when playing some pure strategy $\alpha \in \mathcal{S}_i$ deseves special mention and we will denote it by:

$$u_{i\alpha}(p) = u_i(p_{-i}; \alpha) = u_i(p_1 \dots \alpha \dots p_N) \tag{6}$$

This collection of players $i \in \mathcal{N}$, their strategies $\alpha \in \mathcal{S}_i$ and their payoffs u_i will be our working definition for a *game in normal form*, usually denoted by $\mathfrak{G}$ (or $\mathfrak{G}(\mathcal{N}, \mathcal{S}, u)$ when there is danger of confusion).

Needless to say, rational players will seek to maximize their individual payoff and, in so doing, will avoid those strategies that always lead to diminished payoffs against any play of their opponents. Making this idea more precise, we will say that a strategy $q \in \Delta_i$ of player i is *(strictly) dominated* by $q' \in \Delta_i$ - and we will write $q \prec q'$ - if:

$$u_i(p_{-i}; q) < u_i(p_{-i}; q') \tag{7}$$

for all choices $p_{-i} \in \Delta_{-i} = \prod_{j \neq i} \Delta_j$ of i's opponents.[3] In this way, strictly dominated strategies can be effectively "removed" from the analysis of a game because rational players will never use them. However, by deleting a strategy, another strategy (perhaps of another player) might become dominated and further deletions of *iteratively (strictly) dominated* strategies might be in order. Proceeding in this way ad infinitum, we will say that a strategy is *rationally admissible* if it survives every round of deletion of (strictly) dominated strategies.

The importance of this procedure is that many games can be solved in this way: as a simple (but important) example, if we remove the dominated strategies Cooperate from the Prisoner's Dilemma, we are left with the strategy (Defect, Defect), which is also the game's (unique) Nash equilibrium. This behavior actually occurs in a rather large class of games and we will say that a game $\mathfrak{G}$ is *dominance-solvable* when there is only one rationally admissible strategy. In that case, it should be clear that the game has a unique Nash equilibrium in pure strategies: this is the players' only rational strategy.

B. Exponential Learning and the Replicator Dynamics

Unfortunately, it is not realistic to expect users to be able to perform this elimination of dominated strategies in a timely fashion; if nothing else, this requires an exponentially large amount of information that can hardly be made available to the players. On the brighter side however, players can "learn" to play the game with the help of the so-called *logit model* of exponential learning (see [2] and [15] for its relation to time-averaged best reply dynamics). In a nutshell, players in this scenario play the game repeatedly and they keep records of their strategies' performance; then, at each step, they employ the strategy with the best track record according to an exponential probability law.

To be more precise, player $i \in \mathcal{N}$ keeps a cumulative score $U_{i\alpha}$ of his strategy $\alpha \in \mathcal{S}_i$ as specified by the recursive

[3]The adjective "strict" characterizes the (strict) inequality (7); if the inequality is not strict, q will be called *weakly dominated* by q' and we will write $q \preccurlyeq q'$.

formula:

$$U_{i\alpha}(t+1) = U_{i\alpha}(t) + u_{i\alpha}(p(t)) \tag{8}$$

where, in the absence of initial bias, players set $U_{i\alpha}(0) = 0$ for all $i \in \mathcal{N}$, $\alpha \in \mathcal{S}_i$ and $p(t) \in \Delta$ is the players' strategy profile at the t-th iteration of the game. Needless to say, these profiles are not chosen arbitrarily but in accordance with the scheme's namesake, the *exponential law*:

$$p_{i\alpha}(t+1) = \frac{e^{U_{i\alpha}(t+1)}}{\sum_{\beta\in\mathcal{S}_i} e^{U_{i\beta}(t+1)}} \tag{9}$$

Thus, if we descend to continuous time (which is much more reasonable from an evoluationary perspective), we get:

$$dU_{i\alpha}(t) = u_{i\alpha}(p(t))\,dt \tag{10}$$

where, again, $p(t)$ is the profile determined by the exponential learning model (9):

$$p_{i\alpha}(t) = \frac{e^{U_{i\alpha}(t)}}{\sum_{\beta} e^{U_{i\beta}(t)}} \tag{11}$$

Hence, by differentiating (11) in order to decouple it from (10), we obtain the *standard multi-population replicator dynamics*:

$$\frac{dp_{i\alpha}}{dt} = p_{i\alpha}\left(u_{i\alpha}(p) - \sum\nolimits_{\beta} p_{i\beta}\,u_{i\beta}(p)\right) = p_{i\alpha}(u_{i\alpha}(p) - u_i(p)) \tag{12}$$

These dynamics have been studied extensively (see e.g. [16] for an excellent survey) and one of their most important properties is that only rationally admissible strategies survive in the long run. In particular, Samuelson and Zhang proved the following theorem in [5]:

Theorem 1 (Samuelson and Zhang, 1992): If a pure strategy $\alpha \in \mathcal{S}_i$ is strictly dominated (even iteratively), then $p_{i\alpha}(t)$ converges to zero along any interior solution path of (11).

In sections IV and V, we will show that this result also holds in the stochastic setting where payoffs could be perturbed by arbitrarily loud noise.

IV. The Stochastic Replicator Dynamics

Clearly, one of the drawbacks of the logit model (11) is the latent assumption that players have perfect knowledge of their true payoffs u_i and that the game is unaffected by the stochastic fluctuations of their environment. Since we wish to relax this requirement somewhat in order to study the effect that noise has on learning, we will examine what happens when the scores of (10) are perturbed by noise (e.g. caused by imperfect readings of a player's utility, stochastic interference, fading, etc.).

To do that, we first need to note that players' choices can only depend on their past performance and that they cannot see into nature's future. In other words, their scores should be modelled by Itô stochastic processes (as opposed to future-correlated Stratonovich ones) satisfying the perturbed version of (10):

$$dU_{i\alpha}(t) = u_{i\alpha}(X(t))\,dt + \eta_{i\alpha}(X(t))\,dW_{i\alpha}(t). \tag{13}$$

Here, as in the deterministic setting, $X(t) \in \Delta$ is the strategy profile:

$$X_{i\alpha}(t) = \frac{e^{U_{i\alpha}(t)}}{\sum_{\beta} e^{U_{i\beta}(t)}} \tag{14}$$

and $W_{i\alpha}(t)$ is a Wiener process (Brownian motion) that lives in $\prod_i \mathbb{R}^{S_i}$ and whose components are independent both across players $i \in \mathcal{N}$ and across a particular player's individual strategies $\alpha \in \mathcal{S}_i$.[4] The intensity of the Wiener process is controlled by the diffusion coefficients $\eta_{i\alpha}$ which, conceivably, could depend on the player's actions (for example, this is what happens when there is fading). Somewhat surprisingly, it turns out that this extra degree of generality does not affect our results in any way: as long as the coefficients $\eta_{i\alpha}$ are bounded on Δ (which is only reasonable from a physical perspective), they might as well be constant.

Now, to obtain the noisy analogue of (12), we must decouple the stochastic processes U and X, and this can be done by applying Itô's lemma (see e.g. [17]) to (13). Indeed, with $dW_{j\beta}\cdot dW_{k\gamma} = \delta_{jk}\delta_{\beta\gamma}\,dt$, we easily get:[5]

$$\begin{aligned} dX_{i\alpha} &= \sum\nolimits_j \sum\nolimits_\beta \frac{\partial X_{i\alpha}}{\partial U_{j\beta}}\,dU_{j\beta} \\ &+ \frac{1}{2}\sum\nolimits_{j,k}\sum\nolimits_{\beta,\gamma} \frac{\partial^2 X_{i\alpha}}{\partial U_{j\beta}\partial U_{k\gamma}}\,dU_{j\beta}\cdot dU_{k\gamma} \\ &= \sum\nolimits_\beta \left(u_{i\beta}(X)\frac{\partial X_{i\alpha}}{\partial U_{i\beta}} + \frac{1}{2}\eta_{i\beta}\sum\nolimits_\gamma \eta_{i\gamma}\frac{\partial^2 X_{i\alpha}}{\partial U_{i\beta}\partial U_{i\gamma}}\right) dt \\ &+ \sum\nolimits_\beta \eta_{i\beta}\frac{\partial X_{i\alpha}}{\partial U_{i\beta}}\,dW_{i\beta} \end{aligned} \tag{15}$$

Then, differentiation of (14) finally yields:

$$\begin{aligned} dX_{i\alpha} &= X_{i\alpha}\Big[\Big(u_{i\alpha}(X) - u_i(X)\Big) \\ &+ \frac{1}{2}\left(\eta_{i\alpha}^2(1-2X_{i\alpha}) - \sum\nolimits_\beta \eta_{i\beta}^2 X_{i\beta}(1-2X_{i\beta})\right)\Big]\,dt \\ &+ X_{i\alpha}\left(\eta_{i\alpha}\,dW_{i\alpha} - \sum \eta_{i\beta}X_{i\beta}\,dW_{i\beta}\right) \\ &= X_{i\alpha}\rho_{i\alpha}(X)\,dt + X_{i\alpha}\sum\nolimits_\beta \tau_{i;\alpha\beta(X)}\,dW_{i\beta} \end{aligned} \tag{16}$$

where $\rho_{i\alpha}$ is the relative drift in the brackets of (16) and $\tau_{i;\alpha\beta}(x) = \eta_{i\beta}(x)(\delta_{\alpha\beta} - x_\beta)$, $x \in \Delta$ is the relative diffusion.

Equation (16) will be our stochastic version of the standard replicator dynamics and thus merits some discussion in and by itself. The first important question that needs to be answered is whether the above dynamics admit a (unique) solution $X(t)$ for all $t \in \mathbb{R}$. At first sight, the situation appears to be a bit problematic since the drift $\mu = X\rho$ and diffusion $\sigma = X\tau$ of (13) do not satisfy the linear growth condition $|\mu(x)| + |\sigma(x)| \leq C(1+|x|)$ that is usually necessary for solutions to exist (and to be unique). Fortunately, this problem can be easily circumvented since a simple addition in $\alpha \in \mathcal{S}_i$ reveals that the simplices $\Delta_i \subseteq \Delta$ remain invariant under (16): if $X_i(0) \in \Delta_i$ then $d\left(\sum_\alpha X_{i\alpha}\right) = 0$ and, hence,

[4] In other words, the quadratic covariation of W satisfies $[W_{i\alpha}, W_{j\beta}](t) = \delta_{ij}\delta_{\alpha\beta}t$, for all $i, j \in \mathcal{N}$, $\alpha, \beta \in \mathcal{S}_i$.

[5] Recall here the formal rules of stochastic calculus: $dt\cdot dt = 0 = dt\cdot dW_{i\alpha}$.

$X_i(t) \in \Delta_i$ for all t.[6] So, if ϕ is a smooth bump function that is equal to 1 on some compact set $K \supset \Delta$ and vanishes outside a compact neighborhood K' of K, the stochastic differential equation:

$$dX_{i\alpha} = \phi(X)X_{i\alpha}\left[\rho_{i\alpha}(X)\,dt + \sum\nolimits_{\beta} \tau_{i;\alpha\beta}(X)\,dW_{i\beta}\right] \tag{17}$$

will have smooth and bounded drift and diffusion coefficients and, by the general theory [17], it will admit a unique solution. Since the latter equation agrees with (16) on $K \supset \Delta$ and since any solution of (16) that starts in Δ will always stay in Δ, we conclude that (16) *admits a unique solution for any initial condition* $X(0) = x \in \Delta$.

With all this said and done, it is also important to compare these dynamics to the traditional "aggregate-shocks" of Fudenberg and Harris where most rationality analysis has been taking place (e.g. as in [10]–[12]). To that end, note that the Fudenberg-Harris dynamics take the following form in our notation:

$$\begin{aligned} dX_{i\alpha} &= X_{i\alpha}\left[\left(u_{i\alpha}(X) - u_i(X)\right) - \left(\eta_{i\alpha}^2 X_{i\alpha} - \sum\nolimits_{\beta} \eta_{i\beta}^2 X_{i\beta}^2\right)\right] dt \\ &+ X_{i\alpha}\left(\eta_{i\alpha}\,dW_{i\alpha} - \sum \eta_{i\beta} X_{i\beta}\,dW_{i\beta}\right). \end{aligned} \tag{18}$$

It can be seen immediately that the first and last terms are the same as in our case: they correspond to the drift incurred by the underlying game and the direct effect of the noise on the players' strategy profile respectively. The difference lies in the second term which describes the propagation of noise in the drift. There, the two versions differ by a term of $\eta_{i\alpha}^2$ per strategy $\alpha \in \mathcal{S}_i$ and, innocuous as this difference might appear, we will see that this term has some extremely important ramifications: in stark contrast with [10] or [11], the dynamics (16) lead to rational behavior in all noise levels.

V. The Emergence of Rationality

Thereby, armed with a stochastic differential equation for the evolution of players in a noisy environment, we now seek to show that players eventually do become rational, despite all these disturbances. To that end, we will first need some more technical machinery; motivated by [10], let $p_i, q_i \in \Delta_i$ be two strategies of player i and define:

$$V_{q_i}(p_i) \coloneqq \prod\nolimits_{\alpha} (p_{i\alpha})^{q_{i\alpha}} \tag{19}$$

with the standard convention that $0^0 = 1$ (this amounts to taking the product over the pure strategies that have positive weight in q_i). In effect, this is just another guise of the *Kullback-Leibler relative entropy*:

$$H(q_i, p_i) \coloneqq \sum_{\alpha: q_{i\alpha} > 0} q_{i\alpha} \log \frac{q_{i\alpha}}{p_{i\alpha}} = \sum_{\alpha: q_{i\alpha} > 0} q_{i\alpha} \log(q_{i\alpha}) - L_{q_i}(p_i) \tag{20}$$

where $L_{q_i}(p_i) = \log(V_{q_i}(p_i))$. In particular, we can easily see that $V_{q_i}(p_i) = 0$ if and only if p_i is not using at least one pure strategy that has positive weight under q_i. So, if $V_{q_i}(p_i) = 0$ for all dominated strategies q_i of player i, it immediately follows that p_i cannot be dominated itself.

[6]This should come as no surprise; after all, we did begin with the premise that X is a strategy profile in Δ.

We can now give a precise formulation of our first important result:

Theorem 2: Let $X(t)$ be an interior solution path of the stochastic replicator equation (16) for some game $\mathfrak{G}$. Then, if $q_i \in \Delta_i$ is (strictly) dominated:

$$\lim_{t\to\infty} V_{q_i}(X_i(t)) = 0 \quad \text{almost surely.} \tag{21}$$

In other words, *strictly dominated strategies do not survive in the long run (a.s.).*

Proof: To prove our assertion, we will first need to estimate the temporal evolution of $V_q(X(t))$ (henceforward, we will be dropping the index i when there is no fear of confusion). Thus, if we keep in mind that we can keep all indices in the sums of (20) (on account of X being an interior path) and we apply Itô's lemma to $L_q = \log V_q$, we will get:

$$\begin{aligned} dL_q &= \sum\nolimits_{\beta} \frac{\partial L_q}{\partial x_\beta}\,dX_\beta + \frac{1}{2}\sum\nolimits_{\beta,\gamma} \frac{\partial^2 L_q}{\partial X_\gamma\,\partial X_\beta}\,dX_\beta \cdot dX_\gamma \\ &= \sum\nolimits_{\beta} \frac{q_\beta}{X_\beta}\,dX_\beta - \frac{1}{2}\sum\nolimits_{\beta} \frac{q_\beta}{X_\beta^2}\left(dX_\beta\right)^2 \end{aligned} \tag{22}$$

Then, with dX_β given by (16), this last equation becomes:

$$\begin{aligned} dL_q &= \sum\nolimits_{\beta} q_\beta(u_\beta(X) - u(X)) \\ &- \sum\nolimits_{\beta} q_\beta \cdot \frac{1}{2}\sum\nolimits_{\mu} X_\mu(1 - X_\mu)\eta_\mu^2(X)\,dt \\ &+ \sum\nolimits_{\beta} q_\beta \sum\nolimits_{\mu} (\delta_{\beta\mu} - X_\mu)\eta_\mu(X)\,dW_\mu \end{aligned} \tag{23}$$

Now, if q is strictly dominated by some $q' \in \Delta_i$ (i.e. $u_i(p_{-i}; q) < u_i(p_{-i}; q')$ for all $p_{-i} \in \Delta_{-i}$), we will also have:

$$\begin{aligned} dL_{q-q'} \equiv d\left(L_q - L_{q'}\right) &= \left(u(X_{-i}; q) - u(X_{-i}; q')\right) dt \\ &+ \sum\nolimits_{\beta} (q_\beta - q'_\beta)\,\eta_\beta(X)\,dW_\beta \end{aligned} \tag{24}$$

which represents the integral equation:

$$\begin{aligned} L_{q-q'}(X(t)) &= \int_0^t \left(u(X_{-i}(s); q) - u(X_{-i}(s); q')\right) ds \\ &+ \sum\nolimits_{\beta} (q_\beta - q'_\beta) \int_0^t \eta_\beta(X(s))\,dW_\beta(s) \end{aligned} \tag{25}$$

Since $q \prec q'$, the first term will be bounded above by vt where $v = \max_{x_{-i}}\{u(x_{-i}; q) - u(x_{-i}; q')\} < 0$.[7] However, since monotonicity fails for Itô integrals,[8] the second term must be handled with more care. The trick here is to recall that η_β is bounded on Δ, say by some constant $\eta > 0$. Therefore, the integral $\int_0^t \eta_\beta(X(s))\,dW(s)$ can be almost surely replaced by another Wiener process $\overline{W}$ that evolves at a time scale no faster than ηt (thm 3.4.6 in [18]). Hence, by the law of the iterated logarithm ($\limsup_{t\to\infty} \frac{|\overline{W}(t)|}{\sqrt{2t\log\log t}} = 1$), we will have $\lim_{t\to\infty} L_{q-q'}(X(t)) = -\infty$ almost surely and, with $L_{q'} \leq 0$, the theorem follows. ∎

[7]This is where the proof breaks down if the dominance is weak; in that case, players could experience an ergodic oscillation between pure strategies that yield the same payoff.

[8]For example, $\int_0^1 1\,dW(t) = W(1)$ and the latter is positive or negative with equal probability.

In fact, by induction on the rounds of elimination of dominated strategies, we can now show that the result of Samuelson and Zhang on the extinction of dominated strategies (theorem 1) carries over to our stochastic environment:

Theorem 3: Let $X(t)$ be an interior solution path of the stochastic replicator equation (16) for some game $\mathfrak{G}$. Then, if $q_i \in \Delta_i$ is iteratively (strictly) dominated:

$$\lim_{t\to\infty} V_{q_i}(X_i(t)) = 0 \quad \text{almost surely.} \tag{26}$$

In other words, *only rationally admissible strategies survive in the long run (a.s.).*

Then, as an immediate consequence of this, we have:

Corollary 3.1: Let $\mathfrak{G}$ be a dominance-solvable game and let $x_0 \in \Delta$ be the (unique) Nash equilibrium of $\mathfrak{G}$. Then, *every interior solution path $X(t)$ of the replicator dynamics (16) will converge to x_0 (a.s.)*; more precisely:

$$\lim_{t\to\infty} X(t) = x_0 \quad \text{almost surely.} \tag{27}$$

Intuitively, what happens is similar to the deterministic case [5]. Since dominated strategies die out after some time, we can approximate (16) by a "reduced" dynamical system that lives on the faces of Δ that do not contain dominated vertices. Then, by repeating this argument, further vertices can be eliminated until we reach a point where no dominated strategies remain. Essentially, the proof works just like in [10] and for this reason we will not present it here (see [13] instead); the point where our proofs diverge is that we need no bounds on the noise level η thanks to the form of (25).

Instead, it is more interesting to note that (16) can be rewritten as:

$$\begin{aligned}\log\frac{X_\alpha(t)}{X_\alpha(0)} &= \int_0^t \Big(u_\alpha(X) - u(X) - \frac{1}{2}\sum\nolimits_\beta \eta_\beta^2 X_\beta(1-X_\beta)\Big)\,ds \\ &+ \sum\nolimits_\beta \int_0^t \eta_\beta(\delta_{\alpha\beta} - X_\beta)\,dW(s)\end{aligned} \tag{28}$$

Then, if $\alpha \prec \beta$, we obtain the estimate:

$$\frac{X_\alpha(t)}{X_\beta(t)} < \frac{X_\alpha(0)}{X_\beta(0)} \exp\left\{\nu t + \int_0^t \Big(\eta_\alpha\,dW_\alpha(s) - \eta_\beta\,dW_\beta(s)\Big)\right\} \tag{29}$$

with ν and η as before. So, by using the same trick to estimate the behavior of the Wiener process, we can see that the deterministic drift νt will become the dominant term after some time of the order of $h = \frac{\eta^2}{\nu^2}$. In other words, rationality emerges late for high noise and early for high dominance, the precise time-scale depending on the square of the ratio $\frac{\eta}{\nu}$ which roughly describes the relation between the game's payoffs and the intensity of the noise.

VI. Conclusions and Future Work

We have thus seen that the simplicity of the exponential learning scheme is complemented quite nicely by its robustness: even when the players' payoffs sustain arbitrarily large shocks, irrational behavior always becomes extinct. More to the point, if the underlying game can be solved by deletion of dominated strategies, the players' behavior actually converges to the game's (unique) Nash equilibrium. The only way that noise affects this scenario is by slowing down the players' learning rate: as can be seen by (29), this learning rate is controlled by the "payoff-to-noise" ratio.

On the other hand, corollary 3.1 also touches upon another crucial issue: what happens when a game cannot be solved by elimination of dominated strategies? In particular, does rational behavior (in the form of equilibrial play) still emerge in non-solvable games as it does in the deterministic case? Numerical simulations reveal that this seems to be the case indeed: in the congestion model of section II, users converge to the game's Nash equilibrium in time scales that agree with those predicted by (29).

A thorough analysis of equilibrial play requires a different approach because the payoff differences that allowed us to weed out the noise in (25) cannot be properly bounded if a strategy is not dominated. Still, with the aid of the stochastic Lyapunov method (inspired by [11]), we may recover a large part of the deterministic picture. As it turns out, strict Nash equilibria are stochastically asymptotically stable in the stochastic replicator dynamics (16), no matter how loud the noise is. However, due to space limitations and some technicalities that would take us too far afield, we feel that this issue is better left to be addressed in future work [13].

References

[1] J. Maynard Smith, "The theory of games and the evolution of animal conflicts," *Journal of Theoretical Biology*, vol. 47, no. 1, pp. 209–221, 1974.

[2] D. Fudenberg and D. K. Levine, "Consistency and cautious fictitious play," *Journal of Economic Dynamics and Control*, vol. 19, no. 5-7, pp. 1065–1089, 1995.

[3] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, vol. 40, no. 1-2, pp. 145–156, 1978.

[4] P. Schuster and K. Sigmund, "Replicator dynamics," *Journal of Theoretical Biology*, vol. 100, no. 3, pp. 533–538, 1983.

[5] L. Samuelson and J. Zhang, "Evolutionary stability in asymmetric games," *Journal of Economic Theory*, vol. 57, pp. 363–391, 1992.

[6] E. Altman and Y. Hayel, "A stochastic evolutionary game approach to energy management in a distributed aloha network," in *Proceedings of IEEE INFOCOM 2008*, 2008.

[7] ——, "Stochastic evolutionary games," in *Proceedings of the 13th symposium of dynamic games and applications*, Jul. 2008.

[8] P. Mertikopoulos and A. L. Moustakas, "Correlated anarchy in overlapping wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1160–1169, September 2008.

[9] D. Fudenberg and C. Harris, "Evolutionary dynamics with aggregate shocks," *Journal of Economic Theory*, vol. 57, no. 2, pp. 420–441, August 1992.

[10] A. Cabrales, "Stochastic replicator dynamics," *International Economic Review*, vol. 41, no. 2, pp. 451–81, May 2000.

[11] L. A. Imhof, "The long-run behavior of the stochastic replicator dynamics," *Annals of Applied Probability*, vol. 15, no. 1B, pp. 1019–1045, 2005.

[12] J. Hofbauer and L. A. Imhof, "Time averages, recurrence and transience in the stochastic replicator dynamics," *Annals of Applied Probability*, 2009, to appear.

[13] P. Mertikopoulos and A. L. Moustakas, "Adapting to noise: Exponential learning in randomly perturbed games," forthcoming.

[14] S. Shakkottai, E. Altman, and A. Kumar, "Multihoming of users to access points in WLANs: A population game perspective," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, p. 1207, Aug. 2007.

[15] J. Hofbauer, S. Sorin, and Y. Viossat, "Time average replicator and best reply dynamics," *Mathematics of Operations Research*, to appear.

[16] J. W. Weibull, *Evolutionary Game Theory*. The MIT Press, 1995.

[17] B. Øksendal, *Stochastic Differential Equations*. Springer-Verlag, 2006.

[18] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1998.