THE EMERGENCE OF RATIONAL BEHAVIOR IN THE PRESENCE OF STOCHASTIC PERTURBATIONS

BY PANAYOTIS MERTIKOPOULOS[1,2] AND ARIS L. MOUSTAKAS[1]

University of Athens

We study repeated games where players use an exponential learning scheme in order to adapt to an ever-changing environment. If the game's payoffs are subject to random perturbations, this scheme leads to a new stochastic version of the replicator dynamics that is quite different from the "aggregate shocks" approach of evolutionary game theory. Irrespective of the perturbations' magnitude, we find that strategies which are dominated (even iteratively) eventually become extinct and that the game's strict Nash equilibria are stochastically asymptotically stable. We complement our analysis by illustrating these results in the case of congestion games.

1. Introduction. Ever since it was introduced in [19], the notion of a Nash equilibrium and its refinements have remained among the most prominent solution concepts of noncooperative game theory. In its turn, not only has noncooperative game theory found applications in such diverse topics as economics, biology and network design, but it has also become the standard language to actually *describe* complex agent interactions in these fields.

Still, the issue of why and how players may arrive to equilibrial strategies in the first place remains an actively debated question. After all, the complexity of most games increases exponentially with the number of players and, hence, identifying a game's equilibria quickly becomes prohibitively difficult. Accordingly, as was first pointed out by Aumann in [3], a player has no incentive to play his component of a Nash equilibrium unless he is convinced that all other players will play theirs. And if the game in question has multiple Nash equilibria, this argument gains additional momentum: in that case, even players with unbounded deductive capabilities will be hard-pressed to choose a strategy.

From this point of view, rational individuals would appear to be more in tune with Aumann's notion of a correlated equilibrium where subjective beliefs are also

taken into account [3]. Nevertheless, the seminal work of Maynard Smith on ani-
mal conflicts [15] has cast Nash equilibria in a different light because it unearthed
a profound connection between evolution and rationality: roughly speaking, one
leads to the other. So, when different species contend for the limited resources
of their habitat, evolution and natural selection steer the ensuing conflict to an
equilibrial state which leaves no room for irrational behavior. As a consequence,
instinctive "fight or flight" responses that are deeply ingrained in a species can be
seen as a form of rational behavior, acquired over the species' evolutionary course.

Of course, this evolutionary approach concerns large populations of different
species which are rarely encountered outside the realm of population biology.
However, the situation is not much different in the case of a finite number of play-
ers who try to learn the game by playing again and again and who strive to do
better with the help of some learning algorithm. Therein, evolution does not occur
as part of a birth/death process; rather, it is a byproduct of the players' acquired
experience in playing the game—see [6] for a most comprehensive account.

It is also worth keeping in the back of our mind that in some applications of
game theory, "rationality" requirements precede evolution. For example, recent
applications to network design start from a set of performance aspirations (such as
robustness and efficiency) that the players (network devices) seek to attain in the
network's equilibrial state. Thus, to meet these requirements, one has to literally
reverse-engineer the process by finding the appropriate game whose equilibria will
satisfy the players—the parallel with mechanism design being obvious.

In all these approaches, a fundamental selection mechanism is that of the *repli-
cator dynamics* put forth in [23] and [22] which reinforces a strategy proportion-
ately to the difference of its payoff from the mean (taken over the species or
the player's strategies, depending on the approach). As was shown in the multi-
population setting of Samuelson and Zhang [21] (which is closer to learning than
the self-interacting single-population scenaria of [23] and [22]), these dynamics
are particularly conducive to rationality. Strategies that are suboptimal when paired
against any choice of one's adversaries rapidly become extinct, and in the long run,
only rationally admissible strategies can survive. Even more to the point, the only
attracting states of the dynamics turn out to be precisely the (strict) Nash equilibria
of the game—see [11] for a masterful survey.

We thus see that Nash equilibria arise over time as natural attractors for rational
individuals, a fact which further justifies their prominence among noncooperative
solution concepts. Yet, this behavior is also conditional on the underlying game
remaining stationary throughout the time horizon that it takes players to adapt to
it—and unfortunately, this stationarity assumption is rarely met in practical appli-
cations. In biological models, for example, the reproductive fitness of an individual
may be affected by the ever-changing weather conditions; in networks, communi-
cation channels carry time-dependent noise and interference as well as signals; and
when players try to sample their strategies, they might have to deal with erroneous
or imprecise readings.

It is thus logical to ask: *does rational behavior still emerge in the presence of stochastic perturbations that interfere with the underlying game*?

In evolutionary games, these perturbations traditionally take the form of "aggregate shocks" that are applied directly to the population of each phenotype. This approach by Fudenberg and Harris [5] has spurred quite a bit of interest and there is a number of features that differentiate it from the deterministic one. For example, Cabrales showed in [4] that dominated strategies indeed become extinct, but only if the variance of the shocks is low enough. More recently, the work of Imhof and Hofbauer [8, 10] revealed that even equilibrial play arises over time but again, conditionally on the variance of the shocks.

Be that as it may, if one looks at games with a finite number of players, it is hardly relevant to consider shocks of this type because there are no longer any populations to apply them to. Instead, the stochastic fluctuations should be reflected directly on the stimuli that incite players to change their strategies: their payoffs. This leads to a picture which is very different from the evolutionary one and is precisely the approach that we will be taking.

Outline of results. In this paper, we analyze the evolution of players in stochastically perturbed games of this sort. The particular stimulus-response model that we consider is simple enough: players keep cumulative scores of their strategies' performance and employ exponentially more often the one that scores better. After a few preliminaries in Section 2, this approach is made precise in Section 3 where we derive the stochastic replicator equation that governs the behavior of players when their learning curves are subject to random perturbations.

The replicator equation that we get is different from the "aggregate shocks" approach of [4, 5, 8, 10] and, as a result, it exhibits markedly different rationality properties as well. In stark contrast to the results of [4, 10], we show in Section 4 that dominated strategies become extinct irrespective of the noise level (Proposition 4.1) and provide an exponential bound for the rate of decay of these strategies (Proposition 4.2). In fact, by induction on the rounds of elimination of dominated strategies, we show that this is true even for *iteratively* dominated strategies: despite the noise, only rationally admissible strategies can survive in the long run (Theorem 4.3). Then, as an easy corollary of the above, we infer that players will converge to a strict equilibrium (Corollary 4.4) whenever the underlying game is dominance-solvable.

We continue with the issue of equilibrial play in Section 5 by making a suggestive detour in the land of congestion games. If the noise is relatively mild with respect to the rate with which players learn, we find that the game's potential is a Lyapunov function which ensures that strict equilibria are stochastically attracting; and if the game is dyadic (i.e., players only have two choices), this tameness assumption can be dropped altogether.

Encouraged by the results of Section 5, we attack the general case in Section 6. As it turns out, strict equilibria are *always* asymptotically stochastically stable

in the perturbed replicator dynamics that stem from exponential learning (Theorem 6.1). This begs to be compared to the results of [8, 10] where it is the equilibria of a suitably modified game that are stable, and not necessarily those of the actual game being played. Fortunately, exponential learning seems to give players a clearer picture of the original game and there is no need for similar modifications in our case.

Notational conventions. Given a finite set $S = \{s_0, \ldots, s_n\}$, we will routinely identify the set $\Delta(S)$ of probability measures on S with the standard n-dimensional simplex of $\mathbb{R}^{n+1} : \Delta(S) \equiv \{x \in \mathbb{R}^{n+1} : \sum_\alpha x_\alpha = 1 \text{ and } x_\alpha \geq 0\}$. Under this identification, we will also make no distinction between $s_\alpha \in S$ and the vertex e_α of $\Delta(S)$; in fact, to avoid an overcluttering of indices, we will frequently use α to refer to either s_α or e_α, writing, for example, "$\alpha \in S$" or "$u(\alpha)$" instead of "$s_\alpha \in S$" or "$u(e_\alpha)$," respectively.

To streamline our presentation, we will consistently employ Latin indices for players ($i, j, k, \ldots$) and Greek for their strategies ($\alpha, \beta, \mu, \ldots$), separating the two by a comma when it would have been æsthetically unpleasant not to. In like manner, when we have to discriminate between strategies, we will assume that indices from the first half of the Greek alphabet start at 0 ($\alpha, \beta = 0, 1, 2, \ldots$) while those taken from the second half start at 1 ($\mu, \nu = 1, 2, \ldots$).

Finally, if $X(t)$ is some stochastic process in $\mathbb{R}^n$ starting at $X(0) = x$, its law will be denoted by $P_{X;x}$ or simply by P_x if there is no danger of confusion; and if the context leaves no doubt as to which process we are referring to, we will employ the term "almost surely" in place of the somewhat unwieldy "P_x-almost surely."

2. Preliminaries.

2.1. *Basic facts and definitions from game theory.* As is customary, our starting point will be a (finite) set of N *players*, indexed by $i \in \mathcal{N} = \{1, \ldots, N\}$. The players' possible actions are drawn from their *strategy sets* $\mathcal{S}_i = \{s_{i\alpha} : \alpha = 0, \ldots, S_i - 1\}$ and they can combine them by choosing their α_ith (pure) strategy with probability $p_{i\alpha_i}$. In that case, the players' *mixed strategies* will be described by the points $p_i = (p_{i,0}, p_{i,1}, \ldots) \in \Delta_i := \Delta(\mathcal{S}_i)$ or, more succinctly, by the *strategy profile* $p = (p_1, \ldots, p_N) \in \Delta := \prod_i \Delta_i$.

In particular, if $e_{i\alpha}$ denotes the αth vertex of the ith component simplex $\Delta_i \hookrightarrow \Delta$, the (pure) profile $q = (e_{1,\alpha_1}, \ldots, e_{N,\alpha_N})$ simply corresponds to player i playing $\alpha_i \in \mathcal{S}_i$. On the other hand, if we wish to focus on the strategy of a particular player $i \in \mathcal{N}$ against that of his *opponents* $\mathcal{N}_{-i} := \mathcal{N} \setminus \{i\}$, we will employ the shorthand notation $(p_{-i}; q_i) = (p_1 \cdots q_i \cdots p_N)$ to denote the profile where i plays $q_i \in \Delta_i$ against his opponents' strategy $p_{-i} \in \Delta_{-i} := \prod_{j \neq i} \Delta_j$.

So, once players have made their strategic choices, let $u_{i,\alpha_1,\ldots,\alpha_N}$ be the reward of player i in the profile $(\alpha_1, \ldots, \alpha_N) \in \mathcal{S} = \prod_i \mathcal{S}_i$, that is, the payoff that strategy

$\alpha_i \in \mathcal{S}_i$ yields to player i against the strategy $\alpha_{-i} \in \mathcal{S}_{-i} = \prod_{j \neq i} \mathcal{S}_j$ of i's opponents. Then, if players mix their strategies, their expected reward will be given by the (multilinear) *payoff functions* $u_i : \Delta \to \mathbb{R}$:

$$(2.1) \qquad u_i(p) = \sum_{\alpha_1 \in \mathcal{S}_1} \cdots \sum_{\alpha_N \in \mathcal{S}_N} u_{i,\alpha_1 \cdots \alpha_N} \, p_{1,\alpha_1} \cdots p_{N,\alpha_N}.$$

Under this light, the payoff that a player receives when playing a pure strategy $\alpha \in \mathcal{S}_i$ deserves special mention and will be denoted by

$$(2.2) \qquad u_{i\alpha}(p) := u_i(p_{-i}; \alpha) \equiv u_i(p_1 \cdots \alpha \cdots p_N).$$

This collection of *players* $i \in \mathcal{N}$, their *strategies* $\alpha_i \in \mathcal{S}_i$ and their *payoffs* u_i will be our working definition for a *game in normal form*, usually denoted by $\mathfrak{G}$—or $\mathfrak{G}(\mathcal{N}, \mathcal{S}, u)$ if we need to keep track of more data.

Needless to say, rational players who seek to maximize their individual payoffs will avoid strategies that always lead to diminished payoffs against any play of their opponents. We will thus say that the strategy $q_i \in \Delta_i$ is (*strictly*) *dominated* by $q_i' \in \Delta_i$ and we will write $q_i \prec q_i'$ when

$$(2.3) \qquad u_i(p_{-i}; q_i) < u_i(p_{-i}; q_i')$$

for all strategies $p_{-i} \in \Delta_{-i}$ of i's opponents $\mathcal{N}_{-i}$.

With this in mind, dominated strategies can be effectively removed from the analysis of a game because rational players will have no incentive to ever use them. However, by deleting such a strategy, another strategy (perhaps of another player) might become dominated and further deletions of *iteratively dominated* strategies might be in order (see Section 4 for more details). Proceeding ad infinitum, we will say that a strategy is *rationally admissible* if it survives every round of elimination of dominated strategies. If the set of rationally admissible strategies is a singleton (e.g., as in the Prisoner's Dilemma), the game will be called *dominance-solvable* and the sole surviving strategy will be the game's *rational solution*.

Then again, not all games can be solved in this way and it is natural to look for strategies which are stable at least under unilateral deviations. Hence, we will say that a strategy profile $p \in \Delta$ is a *Nash equilibrium* of the game $\mathfrak{G}$ when

$$(2.4) \qquad u_i(p) \geq u_i(p_{-i}; q) \qquad \text{for all } q \in \Delta_i, i \in \mathcal{N}.$$

If the equilibrium profile p only contains pure strategies $\alpha_i \in \mathcal{S}_i$, we will refer to it as a *pure equilibrium*; and if the inequality (2.4) is strict for all $q \neq p_i \in \Delta_i, i \in \mathcal{N}$, the equilibrium p will carry instead the characterization *strict*.

Clearly, if two pure strategies $\alpha, \beta \in \mathcal{S}_i$ are present with positive probability in an equilibrial strategy $p_i \in \Delta_i$, then we must have $u_{i\alpha}(p) = u_{i\beta}(p)$ as a result of u_i being linear in p_i. Consequently, only pure profiles can satisfy the strict version of (2.4) so that strict equilibria must also be pure. The converse implication is false but only barely so: a pure equilibrium fails to be strict only if a player has more

than one pure strategies that return the same rewards. Since this is almost always true (in the sense that the degenerate case can be resolved by an arbitrarily small perturbation of the payoff functions), we will relax our terminology somewhat and use the two terms interchangeably.

To recover the connection of equilibrial play with strategic dominance, note that if a game is solvable by iterated elimination of dominated strategies, the single rationally admissible strategy that survives will be the game's unique strict equilibrium. But the significance of strict equilibria is not exhausted here: strict equilibria are exactly the evolutionarily stable strategies of multi-population evolutionary games—Proposition 5.1 in [11]. Moreover, as we shall see a bit later, they are the only asymptotically stable states of the multi-population replicator dynamics—again, see Chapter 5, pages 216 and 217 of [11].

Unfortunately, strict equilibria do not always exist, Rock-Paper-Scissors being the typical counterexample. Nevertheless, pure equilibria do exist in many large and interesting classes of games, even when we leave out dominance-solvable ones. Perhaps the most noteworthy such class is that of *congestion games*.

DEFINITION 2.1. A game $\mathfrak{G} \equiv \mathfrak{G}(\mathcal{N}, \mathcal{S}, u)$ will be called a *congestion game* when:

1. all players $i \in \mathcal{N}$ share a common set of *facilities* $\mathcal{F}$ as their strategy set: $\mathcal{S}_i = \mathcal{F}$ for all $i \in \mathcal{N}$;
2. the payoffs are functions of the number of players sharing a particular facility: $u_{i,\alpha_1 \cdots \alpha \cdots \alpha_N} \equiv u_\alpha(N_\alpha)$ where N_α is the number of players choosing the same facility as i.

Amazingly enough, Monderer and Shapley made the remarkable discovery in [18] that these games are actually equivalent to the class of *potential games*.

DEFINITION 2.2. A game $\mathfrak{G} \equiv \mathfrak{G}(\mathcal{N}, \mathcal{S}, u)$ will be called a *potential game* if there exists a function $V : \Delta \to \mathbb{R}$ such that

$$(2.5) \qquad u_i(p_{-i}; q_i) - u_i(p_{-i}; q_i') = -\big(V(p_{-i}; q_i) - V(p_{-i}; q_i')\big)$$

for all players $i \in \mathcal{N}$ and all strategies $p_{-i} \in \Delta_{-i}, q_i, q_i' \in \Delta_i$.

This equivalence reveals that both classes of games possess equilibria in pure strategies: it suffices to look at the vertices of the face of Δ where the (necessarily multilinear) potential function V is minimized.

2.2. *Learning, evolution and the replicator dynamics.* As one would expect, locating the Nash equilibria of a game is a rather complicated problem that requires a great deal of global calculations, even in the case of potential games (where it reduces to minimizing a multilinear function over a convex polytope). Consequently,

it is of interest to see whether there are simple and distributed learning schemes that allow players to arrive at a reasonably stable solution.

One such scheme is based on an exponential learning behavior where players play the game repeatedly and keep records of their strategies' performance. In more detail, at each instance of the game all players $i \in \mathcal{N}$ update the cumulative scores $U_{i\alpha}$ of their strategies $\alpha \in \mathcal{S}_i$ as specified by the recursive formula

$$(2.6) \qquad U_{i\alpha}(t+1) = U_{i\alpha}(t) + u_{i\alpha}(p(t)),$$

where $p(t) \in \Delta$ is the players' strategy profile at the tth iteration of the game and, in the absence of initial bias, we assume that $U_{i\alpha}(0) = 0$ for all $i \in \mathcal{N}, \alpha \in \mathcal{S}_i$. These scores reinforce the perceived success of each strategy as measured by the average payoff it yields and hence, it stands to reason that players will lean towards the strategy with the highest score. The precise way in which they do that is by playing according to the namesake exponential law:

$$(2.7) \qquad p_{i\alpha}(t+1) = \frac{e^{U_{i\alpha}(t+1)}}{\sum_{\beta \in \mathcal{S}_i} e^{U_{i\beta}(t+1)}}.$$

For simplicity, we will only consider the case where players update their scores in continuous time, that is, according to the coupled equations

$$(2.8a) \qquad dU_{i\alpha}(t) = u_{i\alpha}(x(t))\,dt,$$

$$(2.8b) \qquad x_{i\alpha}(t) = \frac{e^{U_{i\alpha}(t)}}{\sum_\beta e^{U_{i\beta}(t)}}.$$

Then, if we differentiate (2.8b) to decouple it from (2.8a), we obtain the *standard (multi-population) replicator dynamics*

$$(2.9) \qquad \frac{dx_{i\alpha}}{dt} = x_{i\alpha}\left(u_{i\alpha}(x) - \sum_\beta x_{i\beta} u_{i\beta}(x)\right) = x_{i\alpha}\big(u_{i\alpha}(x) - u_i(x)\big).$$

Alternatively, if players learn at different speeds as a result of varied stimulus-response characteristics, their updating will take the form

$$(2.10) \qquad x_{i\alpha}(t) = \frac{e^{\lambda_i U_{i\alpha}(t)}}{\sum_\beta e^{\lambda_i U_{i\beta}(t)}},$$

where λ_i represents the *learning rate* of player i, that is, the "weight" which he assigns to his perceived scores $U_{i\alpha}$. In this way, the replicator equation evolves at a different time scale for each player, leading to the *rate-adjusted* dynamics

$$(2.11) \qquad \frac{dx_{i\alpha}}{dt} = \lambda_i x_{i\alpha}\big(u_{i\alpha}(x) - u_i(x)\big).$$

Naturally, the uniform dynamics (2.9) are recovered when all players learn at the "standard" rate $\lambda_i = 1$.

If we view the exponential learning model (2.7) from a stimulus-response angle, we see that that the payoff of a strategy simply represents an (exponential) propensity of employing said strategy. It is thus closely related to the algorithm of *logistic fictitious play* [6] where the strategy x_i of (2.10) can be seen as the (unique) best reply to the profile x_{-i} in some suitably modified payoffs $v_i(x) = u_i(x) + \frac{1}{\lambda_i} H(x_i)$. Interestingly enough, $H(x_i)$ turns out to be none other than the *entropy* of x_i:

$$(2.12) \qquad H(x_i) = - \sum_{\beta : x_{i\beta} > 0} x_{i\beta} \log x_{i\beta}.$$

That being so, we deduce that the learning rates λ_i act the part of (player-specific) inverse temperatures: in high temperatures (small λ_i), the players' learning curves are "soft" and the payoff differences between strategies are toned down; on the contrary, if $\lambda_i \to \infty$ the scheme "freezes" to a myopic best-reply process.

The replicator dynamics were first derived in [23] in the context of population biology, first for different phenotypes within a single species (single-population models), and then for different species altogether (multi-population models; [9] and [11] provide excellent surveys). In both these cases, one begins with large populations of individuals that are programmed to a particular behavior (e.g., fight for "hawks" or flight for "doves") and matches them randomly in a game whose payoffs directly affect the reproductive fitness of the individual players.

More precisely, let $z_{i\alpha}(t)$ be the population size of the phenotype (strategy) $\alpha \in S_i$ of species (player) $i \in \mathcal{N}$ in some multi-population model where individuals are matched to play a game $\mathfrak{G}$ with payoff functions u_i. Then, the relative frequency (share) of α will be specified by the *population state* $x = (x_1, \ldots, x_N) \in \Delta$ where $x_{i\alpha} = z_{i\alpha} / \sum_\beta z_{i\beta}$. So, if N individuals are drawn randomly from the N species, their expected payoffs will be given by $u_i(x)$, $i \in \mathcal{N}$, and if these payoffs represent a proportionate increase in the phenotype's fitness (measured as the number of offsprings in the unit of time), we will have

$$(2.13) \qquad dz_{i\alpha}(t) = z_{i\alpha}(t) u_{i\alpha}(x(t)) \, dt.$$

As a result, the population state $x(t)$ will evolve according to

$$(2.14) \qquad \frac{dx_{i\alpha}}{dt} = \frac{1}{\sum_\beta z_{i\beta}} \frac{dz_{i\alpha}}{dt} - \sum_\gamma \frac{x_{i\alpha}}{\sum_\beta z_{i\beta}} \frac{dz_{i\gamma}}{dt} = x_{i\alpha} (u_{i\alpha}(x) - u_i(x)),$$

which is exactly (2.9) viewed from an evolutionary perspective.

On the other hand, we should note here that in *single-population* models the resulting equation is cubic and not quadratic because strategies are matched against themselves. To wit, assume that individuals are randomly drawn from a large population and are matched against one another in a (symmetric) 2-player game $\mathfrak{G}$ with strategy space $S = \{1, \ldots, S\}$ and payoff matrix $u = \{u_{\alpha\beta}\}$. Then, if x_α denotes the population share of individuals that are programmed to the strategy $\alpha \in S$, their expected payoff in a random match will be given by $u_\alpha(x) := \sum_\beta u_{\alpha\beta} x_\beta \equiv u(\alpha, x)$;

similarly, the population average payoff will be $u(x, x) = \sum_\alpha x_\alpha u_\alpha(x)$. Hence, by following the same procedure as above, we end up with the single-population replicator dynamics

$$(2.15) \qquad \frac{dx_\alpha}{dt} = x_\alpha\big(u_\alpha(x) - u(x, x)\big),$$

which behave quite differently than their multi-population counterpart (2.14).

As far as rational behavior is concerned, the replicator dynamics have some far-reaching ramifications. If we focus on multi-population models, Samuelson and Zhang showed in [21] that the share $x_{i\alpha}(t)$ of a strategy $\alpha \in S_i$ which is strictly dominated (even iteratively) converges to zero along any interior solution path of (2.9); in other words, *dominated strategies become extinct in the long run*. Additionally, there is a remarkable equivalence between the game's Nash equilibria and the stationary points of the replicator dynamics: *the asymptotically stable states of* (2.9) *coincide precisely with the strict Nash equilibria of the underlying game* [11].

2.3. Elements of stability analysis.

A large part of our work will be focused on examining whether the rationality properties of exponential learning (elimination of dominated strategies and asymptotic stability of strict equilibria) remain true in a stochastic setting. However, since asymptotic stability is (usually) too stringent an expectation for stochastic dynamical systems, we must instead consider its stochastic analogue.

That being the case, let $W(t) = (W_1(t), \ldots, W_n(t))$ be a standard Wiener process in $\mathbb{R}^n$ and consider the stochastic differential equation (SDE)

$$(2.16) \qquad dX_\alpha(t) = b_\alpha(X(t))\, dt + \sum_\beta \sigma_{\alpha\beta}(X(t))\, dW_\beta(t).$$

Following [1, 7], the notion of asymptotic stability in this SDE is expressed by the following.

DEFINITION 2.3. We will say that $q \in \mathbb{R}^n$ is *stochastically asymptotically stable* when, for every neighborhood U of q and every $\varepsilon > 0$, there exists a neighborhood V of q such that

$$(2.17) \qquad P_x\Big\{X(t) \in U \text{ for all } t \geq 0, \lim_{t \to \infty} X(t) = q\Big\} \geq 1 - \varepsilon$$

for all initial conditions $X(0) = x \in V$ of the SDE (2.16).

Much the same as in the deterministic case, stochastic asymptotic stability is often established by means of a Lyapunov function. In our context, this notion hinges on the second order differential operator that is associated to (2.16), namely the *generator L* of $X(t)$:

$$(2.18) \qquad L = \sum_{\alpha=1}^{n} b_\alpha(x) \frac{\partial}{\partial x_\alpha} + \frac{1}{2} \sum_{\alpha, \beta=1}^{n} (\sigma(x)\sigma^T(x))_{\alpha\beta} \frac{\partial^2}{\partial x_\alpha\, \partial x_\beta}.$$

The importance of this operator can be easily surmised from Itô's lemma; indeed, if $f : \mathbb{R}^n \to \mathbb{R}$ is sufficiently smooth, the generator L simply captures the drift of the process $Y(t) = f(X(t))$:

$$(2.19) \qquad dY(t) = Lf(X(t)) \, dt + \sum_{\alpha, \beta} \frac{\partial f}{\partial x_\alpha} \bigg|_{X(t)} \sigma_{\alpha\beta}(X(t)) \, dW_\beta(t).$$

In this way, L can be seen as the stochastic version of the time derivative $\frac{d}{dt}$; this analogy then leads to the following.

DEFINITION 2.4.   Let $q \in \mathbb{R}^n$ and let U be an open neighborhood of q. We will say that f is a (local) *stochastic Lyapunov function* for the SDE (2.16) if:

1. $f(x) \geq 0$ for all $x \in U$, with equality iff $x = q$;
2. there exists a constant $k > 0$ such that $Lf(x) \leq -kf(x)$ for all $x \in U$.

Whenever such a Lyapunov function exists, it is known that the point $q \in \mathbb{R}^n$ where f attains its minimum will be stochastically asymptotically stable—for example, see Theorem 4 in pages 314 and 315 of [7].

A final point that should be mentioned here is that our analysis will be constrained on the compact polytope $\Delta = \prod_i \Delta_i$ instead of all of $\prod_i \mathbb{R}^{S_i}$. Accordingly, the "neighborhoods" of Definitions 2.3 and 2.4 should be taken to mean "neighborhoods in Δ," that is, neighborhoods in the subspace topology of $\Delta \hookrightarrow \prod_i \mathbb{R}^{S_i}$. This minor point should always be clear from the context and will only be raised in cases of ambiguity.

3. Learning in the presence of noise.   Of course, it could be argued that the rationality properties of the exponential learning scheme are a direct consequence of the players' receiving accurate information about the game when they update their scores. However, this is a requirement that cannot always be met: the interference of nature in the game or imperfect readings of one's utility invariably introduce fluctuations in (2.8a), and in their turn, these lead to a perturbed version of the replicator dynamics (2.9).

To account for these random perturbations, we will assume that the players' scores are now governed instead by the *stochastic* differential equation

$$(3.1) \qquad dU_{i\alpha}(t) = u_{i\alpha}(X(t)) \, dt + \eta_{i\alpha}(X(t)) \, dW_{i\alpha}(t),$$

where, as before, the strategy profile $X(t) \in \Delta$ is given by the logistic law

$$(3.2) \qquad X_{i\alpha}(t) = \frac{e^{U_{i\alpha}(t)}}{\sum_\beta e^{U_{i\beta}(t)}}.$$

In this last equation, $W(t)$ is a standard Wiener process living in $\prod_i \mathbb{R}^{S_i}$ and the coefficients $\eta_{i\alpha}$ measure the impact of the noise on the players' scoring systems.

Of course, these coefficients need not be constant: after all, the effect of the noise on the payoffs might depend on the state of the game in some typically continuous way. For this reason, we will assume that the functions $\eta_{i\alpha}$ are continuous on Δ, and we will only note en passant that our results still hold for essentially bounded coefficients $\eta_{i\alpha}$ (we will only need to replace min and max with ess inf and ess sup, respectively, in all expressions involving $\eta_{i\alpha}$).

A very important instance of this dependence can be seen if $\eta_{i\alpha}(x_{-i}; \alpha) = 0$ for all $i \in \mathcal{N}, \alpha \in \mathcal{S}_i, x_{-i} \in \Delta_{-i}$, in which case equation (3.1) becomes a convincing model for the case of insufficient information. It states that when a player actually uses a strategy, his payoff observations are accurate enough; but with regards to strategies he rarely employs, his readings could be arbitrarily off the mark.

Now, to decouple (3.1) and (3.2), we may simply apply Itô's lemma to the process $X(t)$. To that end, recall that $W(t)$ has independent components across players and strategies, so that $dW_{j\beta} \cdot dW_{k\gamma} = \delta_{jk}\delta_{\beta\gamma}\,dt$ (the Kronecker symbols $\delta_{\beta\gamma}$ being 0 for $\beta \neq \gamma$ and 1, otherwise). Then, Itô's formula gives

$$
\begin{aligned}
dX_{i\alpha} &= \sum_j \sum_\beta \frac{\partial X_{i\alpha}}{\partial U_{j\beta}}\, dU_{j\beta} \\[2mm]
&\quad + \frac{1}{2} \sum_{j,k} \sum_{\beta,\gamma} \frac{\partial^2 X_{i\alpha}}{\partial U_{j\beta}\, \partial U_{k\gamma}}\, dU_{j\beta} \cdot dU_{k\gamma} \\[2mm]
&= \sum_\beta \left(u_{i\beta}(X) \frac{\partial X_{i\alpha}}{\partial U_{i\beta}} + \frac{1}{2} \eta_{i\beta}^2(X) \frac{\partial^2 X_{i\alpha}}{\partial U_{i\beta}^2} \right) dt \\[2mm]
&\quad + \sum_\beta \eta_{i\beta}(X) \frac{\partial X_{i\alpha}}{\partial U_{i\beta}}\, dW_{i\beta}.
\end{aligned}
$$

(3.3)

On the other hand, a simple differentiation of (3.2) yields

(3.4a)
$$
\frac{\partial X_{i\alpha}}{\partial U_{i\beta}} = X_{i\alpha}(\delta_{\alpha\beta} - X_{i\beta}),
$$

(3.4b)
$$
\frac{\partial^2 X_{i\alpha}}{\partial U_{i\beta}^2} = X_{i\alpha}(\delta_{\alpha\beta} - X_{i\beta})(1 - 2X_{i\beta})
$$

and by plugging these expressions back into (3.3), we get

(3.5)
$$
\begin{aligned}
dX_{i\alpha} &= X_{i\alpha}[u_{i\alpha}(X) - u_i(X)]\, dt \\[2mm]
&\quad + X_{i\alpha}\left[\frac{1}{2}\eta_{i\alpha}^2(X)(1 - 2X_{i\alpha}) - \frac{1}{2}\sum_\beta \eta_{i\beta}^2(X)X_{i\beta}(1 - 2X_{i\beta}) \right] dt \\[2mm]
&\quad + X_{i\alpha}\left[\eta_{i\alpha}(X)\, dW_{i\alpha} - \sum_\beta \eta_{i\beta}(X)X_{i\beta}\, dW_{i\beta} \right].
\end{aligned}
$$

Alternatively, if players update their strategies with different learning rates λ_i, we should instead apply Itô's formula to (2.10). In so doing, we obtain

$$
\begin{aligned}
dX_{i\alpha} &= \lambda_i X_{i\alpha}[u_{i\alpha}(X) - u_i(X)]\,dt \\
&\quad + \frac{\lambda_i^2}{2} X_{i\alpha}\left[\eta_{i\alpha}^2(X)(1 - 2X_{i\alpha}) - \sum_\beta \eta_{i\beta}^2(X)X_{i\beta}(1 - 2X_{i\beta})\right]dt \\
&\quad + \lambda_i X_{i\alpha}\left[\eta_{i\alpha}(X)\,dW_{i\alpha} - \sum \eta_{i\beta}(X)X_{i\beta}\,dW_{i\beta}\right] \\
&= b_{i\alpha}(X)\,dt + \sum_\beta \sigma_{i,\alpha\beta}(X)\,dW_{i\beta},
\end{aligned}
$$

(3.5′)

where, in obvious notation, $b_{i\alpha}(x)$ and $\sigma_{i,\alpha\beta}(x)$ are, respectively, the drift and diffusion coefficients of the diffusion $X(t)$. Obviously, when $\lambda_i = 1$, we recover the uniform dynamics (3.5); equivalently (and this is an interpretation that is well worth keeping in mind), the rates λ_i can simply be regarded as a commensurate inflation of the payoffs and noise coefficients of player $i \in \mathcal{N}$ in the uniform logistic model (3.2).

Equation (3.5) and its rate-adjusted sibling (3.5′) will constitute our stochastic version of the replicator dynamics and thus merit some discussion in and by themselves. First, note that these dynamics admit a (unique) strong solution for any initial state $X(0) = x \in \Delta$, even though they do not satisfy the linear growth condition $|b(x)| + |\sigma(x)| \leq C(1 + |x|)$ that is required for the existence and uniqueness theorem for SDEs (e.g., Theorem 5.2.1 in [20]). Instead, an addition over $\alpha \in \mathcal{S}_i$ reveals that every simplex $\Delta_i \subseteq \Delta$ remains invariant under (3.5): if $X_i(0) = x_i \in \Delta_i$, then $d(\sum_\alpha X_{i\alpha}) = 0$ and hence, $X_i(t)$ will stay in Δ_i for all $t \geq 0$—actually, it is not harder to see that every face of Δ is a trap for $X(t)$.

So, if ϕ is a smooth bump function that is equal to 1 on some open neighborhood of $U \supseteq \Delta$ and which vanishes outside some compact set $K \supseteq U$, the SDE

$$
(3.6) \qquad dX_{i\alpha} = \phi(X)\left(b_{i\alpha}(X)\,dt + \sum_\beta \sigma_{i,\alpha\beta}(X)\,dW_{i\beta}\right)
$$

will have bounded diffusion and drift coefficients and will thus admit a unique strong solution. But since this last equation agrees with (3.5) on Δ and any solution of (3.5) always stays in Δ, we can easily conclude that our perturbed replicator dynamics admit a unique strong solution for any initial $X(0) = x \in \Delta$.

It is also important to compare the dynamics (3.5), (3.5′) to the "aggregate shocks" approach of Fudenberg and Harris [5] that has become the principal incarnation of the replicator dynamics in a stochastic environment. So, let us first recall how aggregate shocks enter the replicator dynamics in the first place. The main idea is that the reproductive fitness of an individual is not only affected by deterministic factors but is also subject to stochastic shocks due to the "weather" and the interference of nature with the game. More precisely, if $Z_{i\alpha}(t)$ denotes the

population size of phenotype $\alpha \in \mathcal{S}_i$ of the species $i \in \mathcal{N}$ in some multi-population evolutionary game $\mathfrak{G}$, its growth will be determined by

$$(3.7) \qquad dZ_{i\alpha}(t) = Z_{i\alpha}(t)\big(u_{i\alpha}(X(t))\, dt + \eta_{i\alpha}\, dW_{i\alpha}(t)\big),$$

where, as in (2.13), $X(t) \in \Delta$ denotes the population shares $X_{i\alpha} = Z_{i\alpha}/\sum_\beta Z_{i\beta}$. In this way, Itô's lemma yields the *replicator dynamics with aggregate shocks*:

$$
\begin{aligned}
dX_{i\alpha} = X_{i\alpha}&\left[\big(u_{i\alpha}(X) - u_i(X)\big) - \left(\eta_{i\alpha}^2 X_{i\alpha} - \sum_\beta \eta_{i\beta}^2 X_{i\beta}^2\right)\right] dt \\
(3.8) \\
&+ X_{i\alpha}\left[\eta_{i\alpha}\, dW_{i\alpha} - \sum \eta_{i\beta} X_{i\beta}\, dW_{i\beta}\right].
\end{aligned}
$$

We thus see that the effects of noise propagate differently in the case of exponential learning and in the case of evolution. Indeed, if we compare equations (3.5) and (3.8) term by term, we see that the drifts are not quite the same: even though the payoff adjustment $u_{i\alpha} - u_i$ ties both equations back together in the deterministic setting ($\eta = 0$), the two expressions differ by

$$(3.9) \qquad X_{i\alpha}\left[\frac{1}{2}\eta_{i\alpha}^2 - \frac{1}{2}\sum_\beta \eta_{i\beta}^2 X_{i\beta}\right] dt.$$

Innocuous as this term might seem, it is actually crucial for the rationality properties of exponential learning in games with randomly perturbed payoffs. As we shall see in the next sections, it leads to some miraculous cancellations that allow rationality to emerge in all noise levels.

This difference further suggests that we can pass from (3.5) to (3.8) simply by modifying the game's payoffs to $\widetilde{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2}\eta_{i\alpha}^2$. Of course, this presumes that the noise coefficients $\eta_{i\alpha}$ be constant—the general case would require us to allow for games whose payoffs may not be multilinear. This apparent lack of generality does not really change things but we prefer to keep things simple and for the time being, it suffices to point out that this modified game was precisely the one that came up in the analysis of [8, 10]. As a result, this modification appears to play a pivotal role in setting apart learning and evolution in a stochastic setting: whereas the modified game is deeply ingrained in the process of natural selection, exponential learning seems to give players a clearer picture of the actual underlying game.

4. Extinction of dominated strategies. Thereby armed with the stochastic replicator equations (3.5), (3.5′) to model exponential learning in noisy environments, the logical next step is to see if the rationality properties of the deterministic dynamics carry over to this stochastic setting. In this direction, we will first show that dominated strategies always become extinct in the long run and that only the rationally admissible ones survive.

As in [4] (implicitly) and [10] (explicitly), the key ingredient of our approach will be the *cross entropy* between two mixed strategies $q_i, x_i \in \Delta_i$ of player $i \in \mathcal{N}$:

$$(4.1) \qquad H(q_i, x_i) := -\sum_{\alpha \,:\, q_{i\alpha} > 0} q_{i\alpha} \log(x_{i\alpha}) \equiv H(q_i) + d_{\mathrm{KL}}(q_i, x_i),$$

where $H(q_i) = -\sum_\alpha q_{i\alpha} \log q_{i\alpha}$ is the *entropy* of q_i and d_{KL} is the intimately related *Kullback–Leibler divergence* (or *relative entropy*):

$$(4.2) \qquad d_{\mathrm{KL}}(q_i, x_i) := H(q_i, x_i) - H(q_i) = \sum_{\alpha \,:\, q_{i\alpha} > 0} q_{i\alpha} \log \frac{q_{i\alpha}}{x_{i\alpha}}.$$

This divergence function is central in the stability analysis of the (deterministic) replicator dynamics because it serves as a distance measure in probability space [11]. As it stands however, d_{KL} is not a distance function per se: neither is it symmetric, nor does it satisfy the triangle inequality. Still, it has the very useful property that $d_{\mathrm{KL}}(q_i, x_i) < \infty$ iff x_i employs with positive probability all pure strategies $\alpha \in \mathcal{S}_i$ that are present in q_i [i.e., iff $\mathrm{supp}(q_i) \subseteq \mathrm{supp}(x_i)$ or iff q_i is absolutely continuous w.r.t. x_i]. Therefore, if $d_{\mathrm{KL}}(q_i, x_i) = \infty$ for all dominated strategies q_i of player i, it immediately follows that x_i cannot be dominated itself. In this vein, we have the following.

PROPOSITION 4.1. *Let $X(t)$ be a solution of the stochastic replicator dynamics* (3.5) *for some interior initial condition $X(0) = x \in \mathrm{Int}(\Delta)$. Then, if $q_i \in \Delta_i$ is* (*strictly*) *dominated,*

$$(4.3) \qquad \lim_{t \to \infty} d_{\mathrm{KL}}(q_i, X_i(t)) = \infty \qquad \textit{almost surely.}$$

In particular, if $q_i = \alpha \in \mathcal{S}_i$ is pure, we will have $\lim_{t \to \infty} X_{i\alpha}(t) = 0$ (a.s.): strictly dominated strategies do not survive in the long run.

PROOF. Note first that $X(0) = x \in \mathrm{Int}(\Delta)$ and hence, $X_i(t)$ will almost surely stay in $\mathrm{Int}(\Delta_i)$ for all $t \geq 0$; this is a simple consequence of the uniqueness of strong solutions and the invariance of the faces of Δ_i under the dynamics (3.5).

Let us now consider the cross entropy $G_{q_i}(t)$ between q_i and $X_i(t)$:

$$(4.4) \qquad G_{q_i}(t) \equiv H(q_i, X_i(t)) = -\sum_\alpha q_{i\alpha} \log X_{i\alpha}(t).$$

As a result of $X_i(t)$ being an interior path, $G_{q_i}(t)$ will remain finite for all $t \geq 0$ (a.s.). So, by applying Itô's lemma we get

$$
\begin{aligned}
(4.5) \qquad dG_{q_i} &= \sum_\beta \frac{\partial G_{q_i}}{\partial X_{i\beta}} dX_{i\beta} + \frac{1}{2} \sum_{\beta, \gamma} \frac{\partial^2 G_{q_i}}{\partial X_{i\gamma} \, \partial X_{i\beta}} dX_{i\beta} \cdot dX_{i\gamma} \\
&= -\sum_\beta \frac{q_{i\beta}}{X_{i\beta}} dX_{i\beta} + \frac{1}{2} \sum_\beta \frac{q_{i\beta}}{X_{i\beta}^2} (dX_{i\beta})^2
\end{aligned}
$$

and, after substituting $dX_{i\beta}$ from the dynamics (3.5), this last equation becomes

$$
dG_{q_i} = \sum_\beta q_{i\beta} \left[u_i(X) - u_{i\beta}(X) + \frac{1}{2} \sum_\gamma \eta_{i\gamma}^2(X) X_{i\gamma} (1 - X_{i\gamma}) \right] dt
$$

(4.6)

$$
+ \sum_\beta q_{i\beta} \sum_\gamma (X_{i\gamma} - \delta_{\beta\gamma}) \eta_{i\gamma}(X) \, dW_{i\gamma}.
$$

Accordingly, if $q_i' \in \Delta_i$ is another mixed strategy of player i, we readily obtain

(4.7)
$$
dG_{q_i} - dG_{q_i'} = \left(u_i(X_{-i}; q_i') - u_i(X_{-i}; q_i) \right) dt
$$
$$
+ \sum_\beta (q_{i\beta}' - q_{i\beta}) \eta_{i\beta}(X) \, dW_{i\beta}
$$

and, after integrating,

(4.8)
$$
G_{q_i - q_i'}(t) = H(q_i - q_i', x) + \int_0^t u_i\left(X_{-i}(s); q_i' - q_i \right) ds
$$
$$
+ \sum_\beta (q_{i\beta}' - q_{i\beta}) \int_0^t \eta_{i\beta}(X(s)) \, dW_{i\beta}(s).
$$

Suppose then that $q_i \prec q_i'$ and let $v_i = \inf\{u_i(x_{-i}; q_i' - q_i) : x_{-i} \in \Delta_{-i}\}$. With Δ_{-i} compact, it easily follows that $v_i > 0$ and the first term of (4.8) will be bounded from below by $v_i t$.

However, since monotonicity fails for Itô integrals, the second term must be handled with more care. To that end, let $\xi_i(s) = \sum_\beta (q_{i\beta}' - q_{i\beta}) \eta_{i\beta}(X(s))$ and note that the Cauchy–Schwarz inequality gives

(4.9)
$$
\xi_i^2(s) \leq S_i \sum_\beta (q_{i\beta}' - q_{i\beta})^2 \eta_{i\beta}^2(X(s))
$$
$$
\leq S_i \eta_i^2 \sum_\beta (q_{i\beta}' - q_{i\beta})^2 \leq 2 S_i \eta_i^2,
$$

where $S_i = |\mathcal{S}_i|$ is the number of pure strategies available to player i and $\eta_i = \max\{|\eta_{i\beta}(x)| : x \in \Delta, \beta \in \mathcal{S}_i\}$; recall also that $q_i, q_i' \in \Delta_i$ for the last step. Therefore, if $\psi_i(t) = \sum_\beta (q_{i\beta}' - q_{i\beta}) \int_0^t \eta_{i\beta}(X(s)) \, dW_{i\beta}(s)$ denotes the martingale part of (4.7) and $\rho_i(t)$ is its quadratic variation, the previous inequality yields

(4.10)
$$
\rho_i(t) = [\psi_i, \psi_i](t) = \int_0^t \xi_i^2(s) \, ds \leq 2 S_i \eta_i^2 t.
$$

Now, if $\lim_{t \to \infty} \rho_i(t) = \infty$, it follows from the time-change theorem for martingales (e.g., Theorem 3.4.6 in [12]) that there exists a Wiener process $\widetilde{W}_i$ such

that $\psi_i(t) = \widetilde{W}_i(\rho_i(t))$. Hence, by the law of the iterated logarithm we get

$$\liminf_{t\to\infty} G_{q_i - q_i'}(t)$$

$$\geq H(q_i - q_i', x) + \liminf_{t\to\infty}(v_i t + \widetilde{W}_i(\rho_i(t)))$$

(4.11) $$\geq H(q_i - q_i', x) + \liminf_{t\to\infty}\left(v_i t - \sqrt{2\rho_i(t)\log\log\rho_i(t)}\right)$$

$$\geq H(q_i - q_i', x) + \liminf_{t\to\infty}\left(v_i t - 2\eta_i\sqrt{S_i t\log\log(2S_i\eta_i^2 t)}\right)$$

$$= \infty \qquad \text{(almost surely)}.$$

On the other hand, if $\lim_{t\to\infty}\rho_i(t) < \infty$, it is trivial to obtain $G_{q_i - q_i'}(t) \to \infty$ by letting $t \to \infty$ in (4.8). Therefore, with $G_{q_i}(t) \geq G_{q_i}(t) - G_{q_i'}(t) \to \infty$, we readily get $\lim_{t\to\infty} d_{\mathrm{KL}}(q_i, X_i(t)) = \infty$ (a.s.); and since $G_\alpha(t) = -\log X_{i\alpha}(t)$ for all pure strategies $\alpha \in \mathcal{S}_i$, our proof is complete. $\square$

As in [10], we can now obtain the following estimate for the lifespan of pure dominated strategies.

PROPOSITION 4.2. *Let $X(t)$ be a solution path of* (3.5) *with initial condition $X(0) = x \in \mathrm{Int}(\Delta)$ and let P_x denote its law. Assume further that the strategy $\alpha \in \mathcal{S}_i$ is dominated; then, for any $M > 0$ and for t large enough, we have*

(4.12) $$P_x\{X_{i\alpha}(t) < e^{-M}\} \geq \frac{1}{2}\mathrm{erfc}\left(\frac{M - h_i(x_i) - v_i t}{2\eta_i\sqrt{S_i t}}\right),$$

where $S_i = |\mathcal{S}_i|$ is the number of strategies available to player i, $\eta_i = \max\{|\eta_{i\beta}(y)| : y \in \Delta, \beta \in \mathcal{S}_i\}$ and the constants $v_i > 0$ and $h_i(x_i)$ do not depend on t.

PROOF. The proof is pretty straightforward and for the most part follows [10]. Surely enough, if $\alpha \prec p_i \in \Delta_i$ and we use the same notation as in the proof of Proposition 4.1, we have

$$-\log X_{i\alpha}(t) = G_\alpha(t) \geq G_\alpha(t) - G_{p_i}(t)$$

(4.13) $$\geq H(\alpha, x) - H(p_i, x) + v_i t + \widetilde{W}_i(\rho_i(t))$$

$$= h_i(x_i) + v_i t + \widetilde{W}_i(\rho_i(t)),$$

where $v_i := \min_{x_{-i}}\{u_i(x_{-i}; p_i) - u_i(x_{-i}; \alpha)\} > 0$ and $h_i(x_i) := \log x_{i\alpha} - \sum_\beta p_{i\beta}\log x_{i\beta}$. Then

(4.14) $$P_x(X_{i\alpha}(t) < e^{-M}) \geq P_x\{\widetilde{W}_i(\rho_i(t)) > M - h_i(x_i) - v_i t\}$$

$$= \frac{1}{2}\mathrm{erfc}\left(\frac{M - h_i(x_i) - v_i t}{\sqrt{2\rho_i(t)}}\right)$$

and, since the quadratic variation $\rho_i(t)$ is bounded above by $2S_i\eta_i^2 t$ (4.10), the estimate (4.12) holds for all sufficiently large t [i.e., such that $M < h_i(x_i) + v_i t$]. □

Some remarks are now in order: first and foremost, our results should be contrasted to those of Cabrales [4] and Imhof [10] where dominated strategies die out only if the noise coefficients (shocks) $\eta_{i\alpha}$ satisfy certain tameness conditions. The origin of this notable difference is the form of the replicator equation (3.5) and, in particular, the extra terms that are propagated there by exponential learning and which are absent from the aggregate shocks dynamics (3.8). As can be seen from the derivations in Proposition 4.1, these terms are precisely the ones that allow players to pick up on the true payoffs $u_{i\alpha}$ instead of the modified ones $\widetilde{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2}\eta_{i\alpha}^2$ that come up in [8, 10] (and, indirectly, in [4] as well).

Secondly, it turns out that the way that the noise coefficients $\eta_{i\beta}$ depend on the profile $x \in \Delta$ is not really crucial: as long as $\eta_{i\beta}(x)$ is continuous (or essentially bounded), our arguments are not affected. The only way in which a specific dependence influences the extinction of dominated strategies is seen in Proposition 4.2: a sharper estimate of the quadratic variation of $\int_0^t \eta_{i\beta}(X(s))\,ds$ could conceivably yield a more accurate estimate for the cumulative distribution function of (4.12).

Finally, it is only natural to ask if Proposition 4.1 can be extended to strategies that are only *iteratively* dominated. As it turns out, this is indeed the case.

THEOREM 4.3. *Let $X(t)$ be a solution path of* (3.5) *starting at $X(0) = x \in$* Int(Δ). *Then, if $q_i \in \Delta_i$ is iteratively dominated,*

$$(4.15) \qquad \lim_{t \to \infty} d_{\mathrm{KL}}(q_i, X_i(t)) = \infty \qquad almost\ surely,$$

that is, only rationally admissible strategies survive in the long run.

PROOF. As in the deterministic case [21], the main idea is that the solution path $X(t)$ gets progressively closer to the faces of Δ that are spanned by the pure strategies which have not yet been eliminated. Following [4], we will prove this by induction on the rounds of elimination of dominated strategies; Proposition 4.1 is simply the case $n = 1$.

To wit, let $A_i \subseteq \Delta_i$, $A_{-i} \subseteq \Delta_{-i}$ and denote by $\mathrm{Adm}(A_i, A_{-i})$ the set of strategies $q_i \in A_i$ that are admissible (i.e., not dominated) with respect to any strategy $q_{-i} \in A_{-i}$. So, if we start with $\mathcal{A}_i^0 = \Delta_i$ and $\mathcal{A}_{-i}^0 = \prod_{j \neq i} \mathcal{A}_j^0$, we may define inductively the set of strategies that remain admissible after n elimination rounds by $\mathcal{A}_i^n := \mathrm{Adm}(\mathcal{A}_i^{n-1}, \mathcal{A}_{-i}^{n-1})$ where $\mathcal{A}_i^{n-1} := \prod_{j \neq i} \mathcal{A}_j^{n-1}$; similarly, the pure strategies that have survived after n such rounds will be denoted by $\mathcal{S}_i^n := \mathcal{S}_i \cap \mathcal{A}_i^n$. Clearly, this sequence forms a descending chain $\mathcal{A}_i^0 \supseteq \mathcal{A}_i^1 \supseteq \cdots$ and the set $\mathcal{A}_i^\infty := \bigcap_0^\infty \mathcal{A}_i^n$ will consist precisely of the strategies of player i that are rationally admissible.

Assume then that the cross entropy $G_{q_i}(t) = H(q_i, X_i(t)) = -\sum_\alpha q_{i\alpha} \times \log X_{i\alpha}(t)$ diverges as $t \to \infty$ for all strategies $q_i \notin \mathcal{A}_i^k$ that die out within the first k rounds; in particular, if $\alpha \notin \mathcal{S}_i^k$ this implies that $X_{i\alpha}(t) \to 0$ as $t \to \infty$. We will show that the same is true if q_i survives for k rounds but is eliminated in the subsequent one.

Indeed, if $q_i \in \mathcal{A}_i^k$ but $q_i \notin \mathcal{A}_i^{k+1}$, there will exist some $q_i' \in \mathcal{A}_i^{k+1}$ such that

$$(4.16) \qquad u_i(x_{-i}; q_i') > u_i(x_{-i}; q_i) \qquad \text{for all } x_{-i} \in \mathcal{A}_{-i}^k.$$

Now, note that any $x_{-i} \in \Delta_{-i}$ can be decomposed as $x_{-i} = x_{-i}^{\mathrm{adm}} + x_{-i}^{\mathrm{dom}}$ where x_{-i}^{adm} is the "admissible" part of x_{-i}, that is, the projection of x_{-i} on the subspace spanned by the surviving vertices $\mathcal{S}_{-i}^k = \prod_{j \neq i} \mathcal{S}_i^k$. Hence, if $v_i = \min\{u_i(\alpha_{-i}; q_i') - u_i(\alpha_{-i}; q_i) : \alpha_{-i} \in \mathcal{S}_{-i}^k\}$, we will have $v_i > 0$ and, by linearity,

$$(4.17) \qquad u_i(x_{-i}^{\mathrm{adm}}; q_i') - u_i(x_{-i}^{\mathrm{adm}}; q_i) \geq v_i > 0 \qquad \text{for all } x_{-i} \in \Delta_{-i}.$$

Moreover, by the induction hypothesis, we also have $X_{-i}^{\mathrm{dom}}(t) \to 0$ as $t \to \infty$. Thus, there exists some t_0 such that

$$(4.18) \qquad |u_i(X_{-i}^{\mathrm{dom}}(t), q_i') - u_i(X_{-i}^{\mathrm{dom}}(t), q_i)| < v_i/2$$

for all $t \geq t_0$ [recall that $X_{-i}^{\mathrm{dom}}(t)$ is spanned by already eliminated strategies].

Therefore, as in the proof of Proposition 4.1, we obtain for $t \geq t_0$

$$(4.19) \quad G_{q_i}(t) - G_{q_i'}(t) \geq M + \frac{1}{2} v_i t + \sum_\beta (q_{i\beta}' - q_{i\beta}) \int_0^t \eta_{i\beta}(X(s)) \, dW_{i\beta}(s),$$

where M is a constant depending only on t_0. In this way, the same reasoning as before gives $\lim_{t \to \infty} G_{q_i}(t) = \infty$ and the theorem follows. $\square$

As a result, if there exists only one rationally admissible strategy, we get the following.

COROLLARY 4.4.   *Let $X(t)$ be an interior solution path of the replicator equation* (3.5) *for some dominance-solvable game $\mathfrak{G}$ and let $x_0 \in \mathcal{S}$ be the* (*unique*) *strict equilibrium of $\mathfrak{G}$. Then*

$$(4.20) \qquad \lim_{t \to \infty} X(t) = x_0 \qquad \textit{almost surely},$$

that is, players converge to the game's strict equilibrium (a.s.).

In concluding this section, it is important to note that all our results on the extinction of dominated strategies remain true in the adjusted dynamics (3.5′) as well: this is just a matter of rescaling. The only difference in using different learning rates λ_i comes about in Proposition 4.2 where the estimate (4.12) becomes

$$(4.21) \qquad P_x\{X_{i\alpha}(t) < e^{-M}\} \geq \frac{1}{2} \operatorname{erfc}\left(\frac{M - h_i(x_i) - \lambda_i v_i t}{2\lambda_i \eta_i \sqrt{S_i t}}\right).$$

As it stands, this is not a significant difference in itself because the two estimates are asymptotically equal for large times. Nonetheless, it is this very lack of contrast that clashes with the deterministic setting where faster learning rates accelerate the emergence of rationality. The reason for this gap is that an increased learning rate λ_i also carries a commensurate increase in the noise coefficients η_i, and thus deflates the benefits of accentuating payoff differences. In fact, as we shall see in the next sections, the learning rates do not really allow players to learn any faster as much as they help diminish their shortsightedness: by effectively being lazy, it turns out that players are better able to average out the noise.

5. Congestion games: A suggestive digression. Having established that irrational choices die out in the long run, we turn now to the question of whether equilibrial play is stable in the stochastic replicator dynamics of exponential learning. However, before tackling this issue in complete generality, it will be quite illustrative to pay a visit to the class of congestion games where the presence of a potential simplifies things considerably. In this way, the results we obtain here should be considered as a motivating precursor to the general case analyzed in Section 6.

5.1. *Congestion games.* To begin with, it is easy to see that the potential V of Definition 2.2 is a Lyapunov function for the deterministic replicator dynamics. Indeed, assume that player $i \in \mathcal{N}$ is learning at a rate $\lambda_i > 0$ and let $x(t)$ be a solution path of the rate-adjusted dynamics (2.11). Then, a simple differentiation of $V(x(t))$ gives

(5.1)
$$\frac{dV}{dt} = \sum_{i,\alpha} \frac{\partial V}{\partial x_{i\alpha}} \frac{dx_{i\alpha}}{dt} = -\sum_{i,\alpha} u_{i\alpha}(x)\lambda_i x_{i\alpha}\big(u_{i\alpha}(x) - u_i(x)\big)$$
$$= -\sum_i \lambda_i \left(\sum_\alpha x_{i\alpha} u_{i\alpha}^2(x) - u_i^2(x)\right) \le 0,$$

the last step following from Jensen's inequality—recall that $\frac{\partial V}{\partial x_{i\alpha}} = -u_{i\alpha}(x)$ on account of (2.5) and also that $u_i(x) = \sum_\alpha x_{i\alpha} u_{i\alpha}(x)$. In particular, this implies that the trajectories $x(t)$ are attracted to the local minima of V, and since these minima coincide with the strict equilibria of the game, we painlessly infer that strict equilibrial play is asymptotically stable in (2.11)—as mentioned before, we plead guilty to a slight abuse of terminology in assuming that all equilibria in pure strategies are also strict.

It is therefore reasonable to ask whether similar conclusions can be drawn in the noisy setting of (3.5′). Mirroring the deterministic case, a promising way to go about this question is to consider again the potential function V of the game and try to show that it is stochastically Lyapunov in the sense of Definition 2.4. Indeed, if $q_0 = (e_{1,0}, \ldots, e_{N,0}) \in \Delta$ is a local minimum of V (and hence, a strict

equilibrium of the underlying game), we may assume without loss of generality that $V(q_0) = 0$ so that $V(x) > 0$ in a neighborhood of q_0. We are thus left to examine the negativity condition of Definition 2.4, that is, whether there exists some $k > 0$ such that $LV(x) \le -kV(x)$ for all x sufficiently close to q_0.

To that end, recall that $\frac{\partial V}{\partial x_{i\alpha}} = -u_{i\alpha}$ and that $\frac{\partial^2 V}{\partial x_{i\alpha}^2} = 0$. Then, the generator L of the rate-adjusted dynamics (3.5$'$) applied to V produces

(5.2)
$$
\begin{aligned}
LV(x) = &-\sum_{i,\alpha} \lambda_i x_{i\alpha} u_{i\alpha}(x)\big(u_{i\alpha}(x) - u_i(x)\big) \\
&- \sum_{i,\alpha} \frac{\lambda_i^2}{2} x_{i\alpha} u_{i\alpha}(x)\Big(\eta_{i\alpha}^2(1 - 2x_{i\alpha}) - \sum_\beta \eta_{i\beta}^2 x_{i\beta}(1 - 2x_{i\beta})\Big),
\end{aligned}
$$

where, for simplicity, we have assumed that the noise coefficients $\eta_{i\alpha}$ are constant.

We will study (5.2) term by term by considering the perturbed strategies $x_i = (1 - \varepsilon_i)e_{i,0} + \varepsilon_i y_i$ where y_i belongs to the face of Δ_i that lies opposite to $e_{i,0}$ (i.e., $y_{i\mu} \ge 0$, $\mu = 1, 2, \dots$ and $\sum_\mu y_{i\mu} = 1$) and $\varepsilon_i > 0$ measures the distance of player i from $e_{i,0}$. In this way, we get

(5.3)
$$
\begin{aligned}
u_i(x) = \sum_\alpha x_{i\alpha} u_{i\alpha}(x) &= (1 - \varepsilon_i)u_{i,0}(x) + \varepsilon_i \sum_\mu y_{i\mu} u_{i\mu}(x) \\
&= u_{i,0}(x) + \varepsilon_i \sum_\mu y_{i\mu}[u_{i\mu}(x) - u_{i,0}(x)] \\
&= u_{i,0}(x) - \varepsilon_i \sum_\mu y_{i\mu} \Delta u_{i\mu} + \mathcal{O}(\varepsilon_i^2),
\end{aligned}
$$

where $\Delta u_{i\mu} = u_{i,0}(q_0) - u_{i\mu}(q_0) > 0$. Then, by going back to (5.2), we obtain

(5.4)
$$
\begin{aligned}
\sum_\alpha x_{i\alpha} u_{i\alpha}&(x)[u_{i\alpha}(x) - u_i(x)] \\
&= (1 - \varepsilon_i)u_{i,0}(x)[u_{i,0}(x) - u_i(x)] \\
&\quad + \varepsilon_i \sum_\mu y_{i\mu} u_{i\mu}(x)[u_{i\mu}(x) - u_i(x)] \\
&= (1 - \varepsilon_i)u_{i,0}(x) \cdot \varepsilon_i \sum_\mu y_{i\mu} \Delta u_{i\mu} \\
&\quad - \varepsilon_i \sum_\mu y_{i\mu} u_{i\mu}(q_0) \Delta u_{i\mu} + \mathcal{O}(\varepsilon_i^2) \\
&= \varepsilon_i \sum_\mu y_{i\mu} u_{i,0}(q_0) \Delta u_{i\mu} - \varepsilon_i \sum_\mu y_{i\mu} u_{i\mu}(q_0) \Delta u_{i\mu} + \mathcal{O}(\varepsilon_i^2) \\
&= \varepsilon_i \sum_\mu y_{i\mu}(\Delta u_{i\mu})^2 + \mathcal{O}(\varepsilon_i^2).
\end{aligned}
$$

As for the second term of (5.2), some easy algebra reveals that

$$
\eta_{i,0}^2(1 - 2x_{i,0}) - \sum_\beta \eta_{i\beta}^2 x_{i\beta}(1 - 2x_{i\beta})
$$

(5.5)
$$
= -\eta_{i,0}^2(1 - 2\varepsilon_i) - \eta_{i,0}^2(1 - \varepsilon_i) - \varepsilon_i \sum_\mu \eta_{i\mu}^2 y_{i\mu}
$$
$$
+ 2(1 - \varepsilon_i)^2\eta_{i,0}^2 + 2\varepsilon_i^2 \sum_\mu \eta_{i\mu}^2 y_{i\mu}^2
$$
$$
= -\varepsilon_i \left(\eta_{i,0}^2 + \sum_\mu y_{i\mu}\eta_{i\mu}^2 \right) + \mathcal{O}(\varepsilon_i^2)
$$

and, after a (somewhat painful) series of calculations, we get

$$
\sum_\alpha x_{i\alpha}u_{i\alpha}(x)\left(\eta_{i\alpha}^2(1 - 2x_{i\alpha}) - \sum_\beta \eta_{i\beta}^2 x_{i\beta}(1 - 2x_{i\beta}) \right)
$$
$$
= (1 - \varepsilon_i)u_{i,0}(x)\left(\eta_{i,0}^2(1 - 2x_{i,0}) - \sum_\beta \eta_{i\beta}^2 x_{i\beta}(1 - 2x_{i\beta}) \right)
$$
$$
+ \varepsilon_i \sum_\mu y_{i\mu}\left(\eta_{i\mu}^2(1 - 2x_{i\mu}) - \sum_\beta \eta_{i\beta}^2 x_{i\beta}(1 - 2x_{i\beta}) \right)
$$

(5.6)
$$
= -\varepsilon_i u_{i,0}(q_0)\left(\eta_{i,0}^2 + \sum_\mu y_{i\mu}\eta_{i\mu}^2 \right)
$$
$$
+ \varepsilon_i \sum_\mu y_{i\mu}u_{i\mu}(q_0)(\eta_{i\mu}^2 + \eta_{i,0}^2) + \mathcal{O}(\varepsilon_i^2)
$$
$$
= -\varepsilon_i \sum_\mu y_{i\mu}\Delta u_{i\mu}(\eta_{i\mu}^2 + \eta_{i,0}^2) + \mathcal{O}(\varepsilon_i^2).
$$

Finally, if we assume without loss of generality that $V(q_0) = 0$ and set $\xi = x - q_0$ (i.e., $\xi_{i,0} = -\varepsilon_i$ and $\xi_{i\mu} = \varepsilon_i y_{i\mu}$ for all $i \in \mathcal{N}$, $\mu \in \mathcal{S}_i \setminus \{0\}$), we readily get

$$
V(x) = \sum_{i,\alpha} \frac{\partial V}{\partial x_{i\alpha}}\xi_{i\alpha} + \mathcal{O}(\xi^2)
$$

(5.7)
$$
= -\sum_{i,\alpha} \frac{\partial u_i}{\partial x_{i\alpha}}\bigg|_{q_0} \xi_{i\alpha} + \mathcal{O}(\varepsilon^2)
$$
$$
= -\sum_{i,\alpha} u_{i\alpha}(q_0)\xi_{i\alpha} + \mathcal{O}(\varepsilon^2)
$$
$$
= \sum_i \varepsilon_i \sum_\mu y_{i\mu}\Delta u_{i\mu} + \mathcal{O}(\varepsilon^2),
$$

where $\varepsilon^2 = \sum_i \varepsilon_i^2$. Therefore, by combining (5.4), (5.6) and (5.7), the negativity condition $LV(x) \leq -kV(x)$ becomes

(5.8)
$$\sum_i \lambda_i \varepsilon_i \sum_\mu y_{i\mu} \Delta u_{i\mu} \left[\Delta u_{i\mu} - \frac{\lambda_i}{2}(\eta_{i\mu}^2 + \eta_{i,0}^2) \right]$$
$$\geq k \sum_i \varepsilon_i \sum_\mu y_{i\mu} \Delta u_{i\mu} + \mathcal{O}(\varepsilon^2).$$

Hence, if $\Delta u_{i\mu} > \frac{\lambda_i}{2}(\eta_{i\mu}^2 + \eta_{i,0}^2)$ for all $\mu \in \mathcal{S}_i \setminus \{0\}$, this last inequality will be satisfied for some $k > 0$ whenever ε is small enough. Essentially, this proves the following.

PROPOSITION 5.1.   *Let* $q = (\alpha_1, \ldots, \alpha_N)$ *be a strict equilibrium of a congestion game* $\mathfrak{G}$ *with potential function* V *and assume that* $V(q) = 0$. *Assume further that the learning rates* λ_i *are sufficiently small so that, for all* $\mu \in \mathcal{S}_i \setminus \{\alpha_i\}$ *and all* $i \in \mathcal{N}$,

(5.9)
$$V(q_{-i}, \mu) > \frac{\lambda_i}{2}(\eta_{i\mu}^2 + \eta_{i,0}^2).$$

Then q *is stochastically asymptotically stable in the rate-adjusted dynamics* (3.5′).

We thus see that no matter how loud the noise η_i might be, stochastic stability is always guaranteed if the players choose a learning rate that is slow enough as to allow them to average out the noise (i.e., $\lambda_i < \Delta V_i / \eta_i^2$). Of course, it can be argued here that it is highly unrealistic to expect players to be able to estimate the amount of Nature's interference and choose a suitably small rate λ_i. On top of that, the very form of the condition (5.9) is strongly reminiscent of the "modified" game of [8, 10], a similarity which seems to contradict our statement that exponential learning favors rational reactions in the *original* game. The catch here is that condition (5.9) is only *sufficient* and Proposition 5.1 merely highlights the role of a potential function in a stochastic environment. As we shall see in Section 6, nothing stands in the way of choosing a different Lyapunov candidate and dropping requirement (5.9) altogether.

5.2. *The dyadic case.*   To gain some further intuition into why the condition (5.9) is redundant, it will be particularly helpful to examine the case where players compete for the resources of only two facilities (i.e., $\mathcal{S}_i = \{0, 1\}$ for all $i \in \mathcal{N}$) and try to learn the game with the help of the uniform replicator equation (3.5). This is the natural setting for the El Farol bar problem [2] and the ensuing minority game [14] where players choose to "buy" or "sell" and are rewarded when they are in the minority–buyers in a sellers' market or sellers in an abundance of buyers.

As has been shown in [17], such games always possess strict equilibria, even when players have distinct payoff functions. So, by relabeling indices if neces-

sary, let us assume that $q_0 = (e_{1,0}, \ldots, e_{N,0})$ is such a strict equilibrium and set $x_i \equiv x_{i,0}$. Then, the generator of the replicator equation (3.5) takes the form

(5.10)
$$L = \sum_i x_i (1 - x_i) \left[\Delta u_i(x) + \frac{1}{2}(1 - 2x_i)\eta_i^2(x) \right] \frac{\partial}{\partial x_i}$$
$$+ \frac{1}{2} \sum_i x_i^2 (1 - x_i)^2 \eta_i^2(x) \frac{\partial^2}{\partial x_i^2},$$

where now $\Delta u_i \equiv u_{i,0} - u_{i,1}$ and $\eta_i^2 = \eta_{i,0}^2 + \eta_{i,1}^2$.

It thus appears particularly appealing to introduce a new set of variables y_i such that $\frac{\partial}{\partial y_i} = x_i(1 - x_i) \frac{\partial}{\partial x_i}$; this is just the "logit" transformation: $y_i = \text{logit } x_i \equiv \log \frac{x_i}{1-x_i}$. In these new variables, (5.10) assumes the astoundingly suggestive guise

(5.11)
$$L = \sum_i \left(\Delta u_i \frac{\partial}{\partial y_i} + \frac{1}{2}\eta_i^2 \frac{\partial^2}{\partial y_i^2} \right),$$

which reveals that the noise coefficients can be effectively decoupled from the payoffs. We can then take advantage of this by letting L act on the function $f(y) = \sum_i e^{-a_i y_i}$ ($a_i > 0$):

(5.12)
$$Lf(y) = -\sum_i a_i \left(\Delta u_i - \frac{1}{2}a_i\eta_i^2 \right) e^{-a_i y_i}.$$

Hence, if a_i is chosen small enough so that $\Delta u_i - \frac{1}{2}a_i\eta_i^2 \geq m_i > 0$ for all sufficiently large y_i [recall that $\Delta u_i(q_0) > 0$ since q_0 is a strict equilibrium], we get

(5.13)
$$Lf(y) \leq -\sum_i a_i m_i e^{-a_i y_i} \leq -kf(y),$$

where $k = \min_i \{a_i m_i\} > 0$. And since f is strictly positive for $y_{i,0} > 0$ and only vanishes as $y \to \infty$ (i.e., at the equilbrium q_0), a trivial modification of the stochastic Lyapunov method (see, e.g., pages 314 and 315 of [7]) yields the following.

PROPOSITION 5.2. *The strict equilibria of minority games are stochastically asymptotically stable in the uniform replicator equation* (3.5).

REMARK 5.3. It is trivial to see that strict equilibria of minority games will also be stable in the rate-adjusted dynamics (3.5′): in that case we simply need to choose a_i such that $\Delta u_i - \frac{1}{2}a_i\lambda_i\eta_i^2 \geq m_i > 0$.

REMARK 5.4. A closer inspection of the calculations leading to Proposition 5.2 reveals that nothing hinges on the minority mechanism per se: it is (5.11) that is crucial to our analysis and L takes this form whenever the underlying game is a *dyadic* one (i.e., $|S_i| = 2$ for all $i \in \mathcal{N}$). In other words, Proposition 5.2 also holds for all games with 2 strategies and should thus be seen as a significant extension of Proposition 5.1.

PROPOSITION 5.5.   *The strict equilibria of dyadic games are stochastically asym ptotically stable in the replicator dynamics* (3.5), (3.5′) *of exponential learning.*

6. Stability of equilibrial play.   In deterministic environments, the "folk theorem" of evolutionary game theory provides some pretty strong ties between equilibrial play and stability: strict equilibria are asymptotically stable in the multipopulation replicator dynamics (2.9) [11]. In our stochastic setting, we have already seen that this is always true in two important classes of games: those that can be solved by iterated elimination of dominated strategies (Corollary 4.4) and dyadic ones (Proposition 5.5).

Although interesting in themselves, these results clearly fall short of adding up to a decent analogue of the folk theorem for stochastically perturbed games. Nevertheless, they are quite strong omens in that direction and such expectations are vindicated in the following.

THEOREM 6.1.   *The strict equilibria of a game $\mathfrak{G}$ are stochastically asymptotically stable in the replicator dynamics* (3.5), (3.5′) *of exponential learning.*

Before proving Theorem 6.1, we should first take a slight detour in order to properly highlight some of the issues at hand. On that account, assume again that the profile $q_0 = (e_{1,0}, \ldots, e_{N,0})$ is a strict equilibrium of $\mathfrak{G}$. Then, if q_0 is to be stochastically stable, say in the uniform dynamics (3.5), one would expect the strategy scores $U_{i,0}$ of player i to grow much faster than the scores $U_{i\mu}, \mu \in \mathcal{S}_i \setminus \{0\}$ of his other strategies. This is captured remarkably well by the "adjusted" scores

$$(6.1a) \qquad Z_{i,0} = \lambda_i U_{i,0} - \log\left(\sum_{\mu} e^{\lambda_i U_{i\mu}}\right),$$

$$(6.1b) \qquad Z_{i\mu} = \lambda_i (U_{i\mu} - U_{i,0}),$$

where $\lambda_i > 0$ is a sensitivity parameter akin (but not identical) to the learning rates of (3.5′) (the choice of common notation is fairly premeditated though).

Clearly, whenever $Z_{i,0}$ is large, $U_{i,0}$ will be much greater than any other score $U_{i\mu}$ and hence, the strategy $0 \in \mathcal{S}_i$ will be employed by player i far more often. To see this in more detail, it is convenient to introduce the variables

$$(6.2a) \qquad Y_{i,0} := e^{Z_{i,0}} = \frac{e^{\lambda_i U_{i,0}}}{\sum_{\nu} e^{\lambda_i U_{i\nu}}},$$

$$(6.2b) \qquad Y_{i\mu} := \frac{e^{Z_{i\mu}}}{\sum_{\nu} e^{Z_{i\nu}}} = \frac{e^{\lambda_i U_{i\mu}}}{\sum_{\nu} e^{\lambda_i U_{i\nu}}},$$

where $Y_{i,0}$ is a measure of how close X_i is to $e_{i,0} \in \Delta_i$ and $(Y_{i,1}, Y_{i,2}, \ldots) \in \Delta^{S_i-1}$ is a direction indicator; the two sets of coordinates are then related by the transformation $Y_{i\alpha} = X_{i\alpha}^{\lambda_i} / \sum_{\mu} X_{i\mu}^{\lambda_i}$, $\alpha \in \mathcal{S}_i$, $\mu \in \mathcal{S}_i \setminus \{0\}$. Consequently, to show that

the strict equilibrium $q_0 = (e_{1,0}, \ldots, e_{N,0})$ is stochastically asymptotically stable in the replicator equation (3.5), it will suffice to show that $Y_{i,0}$ diverges to infinity as $t \to \infty$ with arbitrarily high probability.

Our first step in this direction will be to derive an SDE for the evolution of the $Y_{i\alpha}$ processes. To that end, Itô's lemma gives

$$
dY_{i\alpha} = \sum_{j,\beta} \frac{\partial Y_{i\alpha}}{\partial U_{j\beta}} dU_{j\beta} + \frac{1}{2} \sum_{j,k} \sum_{\beta,\gamma} \frac{\partial^2 Y_{i\alpha}}{\partial U_{j\beta} \partial U_{k\gamma}} dU_{j\beta} \cdot dU_{k\gamma}
$$

(6.3)

$$
= \sum_{\beta} \left(u_{i\beta} \frac{\partial Y_{i\alpha}}{\partial U_{i\beta}} + \frac{1}{2} \eta_{i\beta}^2 \frac{\partial^2 Y_{i\alpha}}{\partial U_{i\beta}^2} \right) dt + \sum_{\beta} \eta_{i\beta} \frac{\partial Y_{i\alpha}}{\partial U_{i\beta}} dW_{i\beta},
$$

where, after a simple differentiation of (6.2a), we have

(6.4a)
$$
\frac{\partial Y_{i,0}}{\partial U_{i,0}} = \lambda_i Y_{i,0}, \qquad \frac{\partial^2 Y_{i,0}}{\partial U_{i,0}^2} = \lambda_i^2 Y_{i,0},
$$

(6.4a′)
$$
\frac{\partial Y_{i,0}}{\partial U_{i\nu}} = -\lambda_i Y_{i,0} Y_{i\nu}, \qquad \frac{\partial^2 Y_{i,0}}{\partial U_{i\nu}^2} = -\lambda_i^2 Y_{i,0} Y_{i\nu}(1 - 2Y_{i\nu})
$$

and, similarly, from (6.2b)

(6.4b)
$$
\frac{\partial Y_{i\mu}}{\partial U_{i,0}} = 0, \qquad \frac{\partial^2 Y_{i\mu}}{\partial U_{i,0}^2} = 0,
$$

$$
\frac{\partial Y_{i\mu}}{\partial U_{i\nu}} = \lambda_i Y_{i\mu}(\delta_{\mu\nu} - Y_{i\nu}),
$$

(6.4b′)
$$
\frac{\partial^2 Y_{i\mu}}{\partial U_{i\nu}^2} = \lambda_i^2 Y_{i\mu}(\delta_{\mu\nu} - Y_{i\nu})(1 - 2Y_{i\nu}).
$$

In this way, by plugging everything back into (6.3) we finally obtain

$$
dY_{i,0} = \lambda_i Y_{i,0} \left[u_{i,0} - \sum_{\mu} Y_{i\mu} u_{i\mu} + \frac{\lambda_i}{2} \eta_{i,0}^2 - \frac{\lambda_i}{2} \sum_{\mu} Y_{i\mu}(1 - 2Y_{i\mu})\eta_{i\mu}^2 \right] dt
$$

(6.5a)

$$
+ \lambda_i Y_{i,0} \left[\eta_{i,0} dW_{i,0} - \sum_{\mu} \eta_{i\mu} Y_{i\mu} dW_{i\mu} \right],
$$

$$
dY_{i\mu} = \lambda_i Y_{i\mu} [u_{i\mu} - \sum_{\nu} u_{i\nu} Y_{i\nu}] dt
$$

(6.5b)

$$
+ \frac{\lambda_i^2}{2} Y_{i\mu} \left[\eta_{i\mu}^2(1 - 2Y_{i\mu}) - \sum_{\nu} \eta_{i\nu}^2 Y_{i\nu}(1 - 2Y_{i\nu}) \right] dt
$$

$$
+ \lambda_i Y_{i\mu} \left[\eta_{i\mu} dW_{i\mu} - \sum_{\nu} \eta_{i\nu} Y_{i\nu} dW_{i\nu} \right],
$$

where we have suppressed the arguments of u_i and η_i in order to reduce notational clutter.

This last SDE is particularly revealing: roughly speaking, we see that if λ_i is chosen small enough, the deterministic term $u_{i,0} - \sum_\mu Y_{i\mu} u_{i\mu}$ will dominate the rest (cf. with the "soft" learning rates of Proposition 5.1). And, since we know that strict equilibria are asymptotically stable in the deterministic case, it is plausible to expect the SDE (6.5) to behave in a similar fashion.

PROOF OF THEOREM 6.1.   Tying in with our previous discussion, we will establish stochastic asymptotic stability of strict equilibria in the dynamics (3.5) by looking at the processes $Y_i = (Y_{i,0}, Y_{i,1}, \ldots) \in \mathbb{R} \times \Delta^{S_i-1}$ of (6.2). In these coordinates, we just need to show that for every $M_i > 0, i \in \mathcal{N}$ and any $\varepsilon > 0$, there exist $Q_i > M_i$ such that if $Y_{i,0}(0) > Q_i$, then, with probability greater than $1 - \varepsilon$, $\lim_{t\to\infty} Y_{i,0}(t) = \infty$ and $Y_{i,0}(t) > M_i$ for all $t \geq 0$. In the spirit of the previous section, we will accomplish this with the help of the stochastic Lyapunov method.

Our first task will be to calculate the generator of the diffusion $Y = (Y_1, \ldots, Y_N)$, that is, the second order differential operator

$$(6.6) \qquad L = \sum_{\substack{i\in\mathcal{N} \\ \alpha\in\mathcal{S}_i}} b_{i\alpha}(y) \frac{\partial}{\partial y_{i\alpha}} + \frac{1}{2} \sum_{\substack{i\in\mathcal{N} \\ \alpha,\beta\in\mathcal{S}_i}} (\sigma_i(y)\sigma_i^T(y))_{\alpha\beta} \frac{\partial^2}{\partial y_{i\alpha}\,\partial y_{i\beta}},$$

where b_i and σ_i are the drift and diffusion coefficients of the SDE (6.5), respectively. In particular, if we restrict our attention to sufficiently smooth functions of the form $f(y) = \sum_{i\in\mathcal{N}} f_i(y_{i,0})$, the application of L yields

$$
\begin{aligned}
Lf(y) = \sum_{i\in\mathcal{N}} \lambda_i y_{i,0} \bigg[& u_{i,0} + \frac{\lambda_i}{2}\eta_{i,0}^2 \\
& - \sum_\mu y_{i\mu}\left(u_{i\mu} - \frac{\lambda_i}{2}(1-2y_{i\mu})\eta_{i\mu}^2\right)\bigg] \frac{\partial f_i}{\partial y_{i,0}} \\
& + \frac{1}{2} \sum_{i\in\mathcal{N}} \lambda_i^2 y_{i,0}^2 \bigg[\eta_{i,0}^2 + \sum_\mu \eta_{i\mu}^2 y_{i\mu}^2\bigg] \frac{\partial^2 f_i}{\partial^2 y_{i,0}}.
\end{aligned}
$$

(6.7)

Therefore, let us consider the function $f(y) = \sum_i 1/y_{i,0}$ for $y_{i,0} > 0$. With $\frac{\partial f}{\partial y_{i,0}} = -1/y_{i,0}^2$ and $\frac{\partial^2 f}{\partial y_{i,0}^2} = 2/y_{i,0}^3$, (6.7) becomes

$$
\begin{aligned}
Lf(y) = -\sum_{i\in\mathcal{N}} \frac{\lambda_i}{y_{i,0}} \bigg[& u_{i,0} - \sum_\mu u_{i\mu}y_{i\mu} - \frac{\lambda_i}{2}\eta_{i,0}^2 \\
& - \frac{\lambda_i}{2}\sum_\mu y_{i\mu}(1-y_{i\mu})\eta_{i\mu}^2\bigg].
\end{aligned}
$$

(6.8)

However, since $q_0 = (e_{1,0}, \dots, e_{N,0})$ has been assumed to be a strict Nash equilibrium of $\mathfrak{G}$, we will have $u_{i,0}(q_0) > u_{i\mu}(q_0)$ for all $\mu \in \mathcal{S}_i \setminus \{0\}$. Then, by continuity, there exists some positive constant $v_i > 0$ with $u_{i,0} - \sum_\mu u_{i\mu} y_{i\mu} \geq v_i > 0$ whenever $y_{i,0}$ is large enough (recall that $\sum_\mu y_{i\mu} = 1$). So, if we set $\eta_i = \max\{|\eta_{i\beta}(x)| : x \in \Delta, \beta \in \mathcal{S}_i\}$ and pick positive λ_i with $\lambda_i < v_i/\eta_i^2$, we get

$$(6.9) \qquad Lf(y) \leq -\sum_{i \in \mathcal{N}} \frac{\lambda_i v_i}{2} \frac{1}{y_{i,0}} \leq -\frac{1}{2} \min_i \{\lambda_i v_i\} f(y)$$

for all sufficiently large $y_{i,0}$. Moreover, f is strictly positive for $y_{i,0} > 0$ and vanishes only as $y_{i,0} \to \infty$. Hence, as in the proof of Proposition 5.2, our claim follows on account of f being a (local) stochastic Lyapunov function.

Finally, in the case of the rate-adjusted replicator dynamics (3.5′), the proof is similar and only entails a rescaling of the parameters λ_i. $\quad\square$

REMARK 6.2. If we trace our steps back to the coordinates $X_{i\alpha}$, our Lyapunov candidate takes the form $f(x) = \sum_i (x_{i,0}^{-\lambda_i} \sum_\mu x_{i\mu}^{\lambda_i})$. It thus begs to be compared to the Lyapunov function $\sum_\mu x_\mu^\lambda$ employed by Imhof and Hofbauer in [8] to derive a conditional version of Theorem 6.1 in the evolutionary setting. As it turns out, the obvious extension $f(x) = \sum_i \sum_\mu x_{i\mu}^{\lambda_i}$ works in our case as well, but the calculations are much more cumbersome and they are also shorn of their ties to the adjusted scores (6.1).

REMARK 6.3. We should not neglect to highlight the dual role that the learning rates λ_i play in our analysis. In the logistic learning model (2.10), they measure the players' convictions and how strongly they react to a given stimulus (the scores $U_{i\alpha}$); in this role, they are fixed at the outset of the game and form an intrinsic part of the replicator dynamics (3.5′). On the other hand, they also make a virtual appearance as free temperature parameters in the adjusted scores (6.1), to be softened until we get the desired result. For this reason, even though Theorem 6.1 remains true for any choice of learning rates, the function $f(x) = \sum_i x_{i,0}^{-\lambda_i} \sum_\mu x_{i\mu}^{\lambda_i}$ is Lyapunov only if the sensitivity parameters λ_i are small enough. It might thus seem unfortunate that we chose the same notation in both cases, but we feel that our decision is justified by the intimate relation of the two parameters.

7. Discussion. Our aim in this last section will be to discuss a number of important issues that we have not been able to address thoroughly in the rest of the paper; truth be told, a good part of this discussion can be seen as a roadmap for future research.

Ties with evolutionary game theory. In single-population evolutionary models, an evolutionarily stable strategy (ESS) is a strategy which is robust against invasion by mutant phenotypes [15]. Strategies of this kind can be considered as a stepping

stone between mixed and strict equilibria and they are of such significance that it makes one wonder why they have not been included in our analysis.

The reason for this omission is pretty simple: even the weakest evolutionary criteria in multi-population models tend to reject all strategies which are not strict Nash equilibria [11]. Therefore, since our learning model (2.9) corresponds exactly to the multi-population environment (2.14), we lose nothing by concentrating our analysis only on the strict equilibria of the game. If anything, this equivalence between ESS and strict equilibria in multi-population settings further highlights the importance of the latter.

However, this also brings out the gulf between the single-population setting and our own, even when we restrict ourselves to 2-player games (which are the norm in single-population models). Indeed, the single-population version of the dynamics (3.8) is:

$$dX_\alpha = X_\alpha \Bigg[\big(u_\alpha(X) - u(X, X)\big) - \Big(\eta_\alpha^2 X_\alpha - \sum_\beta \eta_\beta^2 X_\beta^2\Big) \Bigg] dt$$

(7.1)

$$+ X_\alpha \Big[\eta_\alpha \, dW_\alpha - \sum \eta_\beta X_\beta \, dW_\beta \Big].$$

As it turns out, if a game possesses an interior ESS and the shocks are mild enough, the solution paths $X(t)$ of the (single-population) replicator dynamics will be recurrent (Theorem 2.1 in [10]). Theorem 6.1 rules out such behavior in the case of strict equilibria (the multi-population analogue of ESS), but does not answer the following question: if the underlying game only has mixed equilibria, will the solution $X(t)$ of the dynamics (3.5) be recurrent?

This question is equivalent to showing that a profile x is stochastically asymptotically stable in the replicator equations (3.5), (3.5′) only if it is a strict equilibrium. Since Theorem 6.1 provides the converse "if" part, an answer in the positive would yield a strong equivalence between stochastically stable states and strict equilibria; we leave this direction to be explored in future papers.

Itô vs. Stratonovich. For comparison purposes (but also for simplicity), let us momentarily assume that the noise coefficients $\eta_{i\alpha}$ do not depend on the state $X(t)$ of the game. In that case, it is interesting (and very instructive) to note that the SDE (3.1) remains unchanged if we use Stratonovich integrals instead of Itô ones:

$$(7.2) \qquad dU_{i\alpha}(t) = u_{i\alpha}(X(t)) \, dt + \eta_{i\alpha} \, \partial W_{i\alpha}(t).$$

Then, after a few calculations, the corresponding replicator equation reads

$$(7.3) \quad \partial X_{i\alpha} = X_{i\alpha}\big(u_{i\alpha}(X) - u_i(X)\big) dt + X_{i\alpha}\Big(\eta_{i\alpha} \, \partial W_{i\alpha} - \sum \eta_{i\beta} X_{i\beta} \, \partial W_{i\beta}\Big).$$

The form of this last equation is remarkably suggestive. First, it highlights the role of the modified game $\widetilde{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2}\eta_{i\alpha}^2$ even more crisply than (3.5): the payoff terms are completely decoupled from the noise, in contrast to what one obtains

by introducing Stratonovich perturbations in the evolutionary setting [8, 13]. Secondly, one can seemingly use this simpler equation to get a much more transparent proof of Proposition 4.1: the estimates for the cross entropy terms $G_{q_i - q_i'}$ are recovered almost immediately from the Stratonovich dynamics. However, since (7.3) takes this form only for constant coefficients $\eta_{i\alpha}$ (the general case is quite a bit uglier), we chose the route of consistency and employed Itô integrals throughout our paper.

Applications in network design. Before closing, it is worth pointing out the applicability of the above approach to networks where the presence of noise or uncertainty has two general sources. The first of these has to do with the time variability of the connections which may be due to the fluctuations of the link quality because of mobility in the wireless case or because of external factors (e.g., load conditions) in wireline networks. This variability is usually dependent on the state of the network and was our original motivation in considering noise coefficients $\eta_{i\alpha}$ that are functions of the players' strategy profile; incidentally, it was also our original motivation for considering randomly fluctuating payoffs in the first place: travel times and delays in traffic models are not determined solely by the players' choices, but also by the fickle interference of nature.

The second source stems from errors in the measurement of the payoffs themselves (e.g., the throughput obtained in a particular link) and also from the lack of information on the payoff of strategies that were not employed. The variability of the noise coefficients $\eta_{i\alpha}$ again allows for a reasonable approximation to this problem. Indeed, if $\eta_{i\alpha} : \Delta \to \mathbb{R}$ is continuous and satisfies $\eta_{i\alpha}(x_{-i}; \alpha) = 0$ for all $i \in \mathcal{N}, \alpha \in \mathcal{S}_i$, this means that there are only errors in estimating the payoffs of strategies that were not employed (or small errors for pure strategies that are employed with high probability). Of course, this does not yet give the full picture [one should consider the discrete-time dynamical system (2.6) instead where the players' *actual* choices are considered], but we conjecture that our results will remain essentially unaltered.

Acknowledgments. We would like to extend our gratitude to the anonymous referee for his insightful comments and to David Leslie from the university of Bristol for the fruitful discussions on the discrete version of the exponential learning model.

Some of the results of Section 4 were presented in the conference "Game Theory for Networks" in Boğaziçi University, Istanbul, May 2009 [16].

REFERENCES

[1] ARNOLD, L. (1974). *Stochastic Differential Equations*: *Theory and Applications*. Wiley, New York. MR0443083
[2] ARTHUR, W. B. (1994). Inductive reasoning and bounded rationality (the El Farol problem). *American Economic Review* **84** 406–411.

[3] AUMANN, R. J. (1974). Subjectivity and correlation in randomized strategies. *J. Math. Econom.* **1** 67–96. MR0421694

[4] CABRALES, A. (2000). Stochastic Replicator Dynamics. *Internat. Econom. Rev.* **41** 451–481. MR1760461

[5] FUDENBERG, D. and HARRIS, C. (1992). Evolutionary dynamics with aggregate shocks. *J. Econom. Theory* **57** 420–441. MR1180005

[6] FUDENBERG, D. and LEVINE, D. K. (1998). *The Theory of Learning in Games. MIT Press Series on Economic Learning and Social Evolution* **2**. MIT Press, Cambridge, MA. MR1629477

[7] GIKHMAN, I. I. and SKOROKHOD, A. V. (1971). *Stochastische Differentialgleichungen*. Akademie-Verlag, Berlin.

[8] HOFBAUER, J. and IMHOF, L. A. (2009). Time averages, recurrence and transience in the stochastic replicator dynamics. *Ann. Appl. Probab.* **19** 1347–1368. MR2538073

[9] HOFBAUER, J. and SIGMUND, K. (1988). *The Theory of Evolution and Dynamical Systems*. Cambridge Univ. Press, Cambridge. MR1071180

[10] IMHOF, L. A. (2005). The long-run behavior of the stochastic replicator dynamics. *Ann. Appl. Probab.* **15** 1019–1045. MR2114999

[11] JÖRGEN, W. W. (1995). *Evolutionary Game Theory*. MIT Press, Cambridge, MA. MR1347921

[12] KARATZAS, I. and SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Springer, New York.

[13] KHAS'MINSKII, R. Z. and POTSEPUN, N. (2006). On the replicator dynamics behavior under Stratonovich type random perturbations. *Stoch. Dyn.* **6** 197–211. MR2239089

[14] MARSILI, M., CHALLET, D. and ZECCHINA, R. (2000). Exact solution of a modified El Farol's bar problem: Efficiency and the role of market impact. *Phys. A* **280** 522–553.

[15] MAYNARD SMITH, J. (1974). The theory of games and the evolution of animal conflicts. *J. Theoret. Biol.* **47** 209–221. MR0444115

[16] MERTIKOPOULOS, P. and MOUSTAKAS, A. L. (2009). Learning in the presence of noise. In *GameNets'09: Proceedings of the* 1*st International Conference on Game Theory for Networks* 308–313. IEEE Press, Piscataway, NJ.

[17] MILCHTAICH, I. (1996). Congestion games with player-specific payoff functions. *Games Econom. Behav.* **13** 111–124. MR1385132

[18] MONDERER, D. and SHAPLEY, L. S. (1996). Potential games. *Games Econom. Behav.* **14** 124–143. MR1393599

[19] NASH, J. F. (1951). Non-cooperative games. *Ann. Math.* **54** 286–295. MR0043432

[20] ØKSENDAL, B. (2006). *Stochastic Differential Equations*, 6th ed. Springer, Berlin.

[21] SAMUELSON, L. and ZHANG, J. (1992). Evolutionary stability in asymmetric games. *J. Econom. Theory* **57** 363–391. MR1180002

[22] SCHUSTER, P. and SIGMUND, K. (1983). Replicator dynamics. *J. Theoret. Biol.* **100** 533–538. MR0693413

[23] TAYLOR, P. D. and JONKER, L. B. (1978). Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40** 145–156. MR0489983

DEPARTMENT OF PHYSICS
UNIVERSITY OF ATHENS
PANEPISTIMIOUPOLIS, ZOGRAFOU
ATHENS, 15784
GREECE
E-MAIL: pmertik@phys.uoa.gr
            arislm@phys.uoa.gr