RATIONAL BEHAVIOUR IN THE PRESENCE OF STOCHASTIC PERTURBATIONS

PANAYOTIS MERTIKOPOULOS AND ARIS L. MOUSTAKAS

Abstract. We study repeated games where players employ 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 [1], the notion of a Nash equilibrium and its refinements remain among the most prominent solution concepts of non-cooperative game theory. The reason for this is pretty simple and lies at the heart of any competitive scenario: the fear of losing is a strong deterrent for any rational player who might consider defecting unilaterally from such an equilibrium.

Still, this discouragement does little to settle the issue of why and how players might have arrived to equilibrial strategies in the first place. 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 [2], 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 [2] where subjective beliefs are also taken into account. Nevertheless, the seminal work of Maynard Smith on animal conflicts [3] 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 behaviour. As a consequence, instinctive “fight or flight” responses that are deeply ingrained in a species can be seen as a form of rational behaviour, 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 players 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 [4] for a most comprehensive account.

In both approaches, a fundamental selection mechanism is that of the replicator dynamics [5, 6] which reinforces a strategy proportionately 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 by Samuelson and Zhang in the multi-population setting of [7] (which is closer to learning than the self-interacting single-population scenario of [5, 6]), 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 [8] for a masterful survey.

We thus see that Nash equilibria arise over time as natural attractors for rational individuals. This fact further justifies their prominence among non-cooperative solution concepts but it is also conditional on the underlying game remaining stationary throughout the time horizon that it takes players to adapt to it. Unfortunately, in many practical applications this stationarity assumption cannot be met: in biological models, the reproductive fitness of an individual may be affected by the ever-changing weather conditions; in networks, communication 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 behaviour 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 of Fudenberg and Harris [9] has spurred quite a bit of interest and there is a number of features that differentiate it from the deterministic one. In [10] Cabrales showed 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 [11,12] revealed that even equilibrial play arises given enough 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.

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 behaviour 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 [9–12] and, as a result, it exhibits markedly different rationality properties as well. In stark contrast to the results of [10, 11], we show in section 3 that dominated strategies become extinct irrespective of the noise level. 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.

We then begin addressing 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), we can drop this assumption 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. This begs to be compared to the results of [11, 12] 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.

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 $i \in N = \{1, \ldots, N\}$. The players’ possible actions are drawn from their strategy sets $S_i = \{0, 1 \ldots S_i - 1\}$ and they can combine these (pure) strategies by choosing $\alpha_i \in S_i$ with probability $p_{i\alpha_i}$. In that case, the players’ mixed strategies will be described by the points $p$.
\( p_i = (p_{i,0}, p_{i,1}, \ldots) \in \Delta_i := \Delta(S_i) \) or, more succinctly, by the strategy profile \( p = (p_1, \ldots, p_N) \in \Delta := \prod_i \Delta_i \). Alternatively, if we wish to focus on the strategies of a particular player \( i \in N \) against the ones of his opponents \( N_{-i} := N \setminus \{i\} \), we will employ the shorthand notation \( (p_{-i}; q) = (p_1 \ldots q \ldots p_N) \) to denote the profile where \( i \) plays \( q \in \Delta_i \) against his opponents’ strategy \( p_{-i} \in \Delta_{-i} := \prod_{j \neq i} \Delta_j \).

Now, once players have made their strategic choices, they will be rewarded according to the (multilinear) payoff functions \( u_i : \Delta \to \mathbb{R} \):

\[
u_i(p) = \sum_{\alpha_1 \in S_1} \cdots \sum_{\alpha_N \in S_N} u_{i, \alpha_1 \ldots \alpha_N} p_{1, \alpha_1} \cdots p_{N, \alpha_N}
\]

where \( u_{i, \alpha_1 \ldots \alpha_N} \) is the reward of player \( i \) in the profile \( (\alpha_1 \ldots \alpha_N) \in S = \prod_i S_i \), i.e. the payoff that strategy \( \alpha_i \in S_i \) yields to player \( i \) against the strategy \( \alpha_{-i} \in S_{-i} := \prod_{j \neq i} S_j \) of \( i \)'s opponents. Under this light, the payoff that a player receives when playing a pure strategy \( \alpha \in S_i \) deserves special mention and will be given by:

\[
u_{i, \alpha}(p) := u_i(p_{-i}; \alpha) \equiv u_i(p_1, \ldots, p_N).
\]

This collection of players \( i \in N \), their strategies \( \alpha_i \in S_i \) and their payoffs \( u_i \) will be our working definition for a game in normal form, usually denoted by \( \mathcal{G} \) - or \( \mathcal{G}(N, S, u) \) 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

\[
u_i(p_{-i}; q_i) < \nu_i(p_{-i}; q_i')
\]

for all strategies \( p_{-i} \in \Delta_{-i} \) of \( i \)'s opponents \( N_{-i} \).

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 \ref{sec:iterated_dominated} 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 \( \mathcal{G} \) when

\[
u_i(p) \geq \nu_i(p_{-i}; q) \quad \text{for all } q \in \Delta_i, i \in N.
\]

If the equilibrium profile \( p \) only contains pure strategies \( \alpha_i \in S_i \), we will refer to it as a pure equilibrium; and if the inequality \ref{eq:strict} is strict for all \( q \neq p_i \in \Delta_i, i \in N \), the equilibrium \( p \) will carry instead the characterization strict. Clearly, only pure profiles can satisfy the strict version of \ref{eq:strict} and therefore 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

\[2\]The adjective “strict” characterizes the inequality \ref{eq:strict}; if the inequality is not strict, \( q_i \) will be called weakly dominated by \( q_i' \) and we will write \( q_i \preceq q_i' \). Because our primary interest lies in strictly dominated strategies, “dominated” should always be taken to mean “strictly dominated.”
same rewards. Since this occurrence has measure zero, we will relax our terminology somewhat and use the two terms interchangeably.

To recover now 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. Moreover, as we shall see a bit later, they are the only asymptotically stable states of the multi-population replicator dynamics.

Unfortunately, not all game possess strict equilibria (Rock-Paper-Scissors is 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 $G \equiv G(N, S, u)$ will be called a congestion game when:

1. all players $i \in N$ share a common set of facilities $F$ as their strategy set: $S_i = F$ for all $i \in N$;
2. the payoffs are functions of the number of players sharing a particular facility: $u_i(a_1, a_2, ..., a_N) \equiv u_a(N_a)$ where $N_a = \sum_j \delta_{a_j a}$ is the number of players choosing the same facility as $i$.

Amazingly enough, it turns out that these games are equivalent to the class of potential games, first studied by Monderer and Shapley in [13]:

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

$$u_i(p_{-i}; q_i) - u_i(p_{-i}; q'_i) = -(V(p_{-i}; q_i) - V(p_{-i}; q'_i))$$

for all players $i \in N$ 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 $\Delta$ where the (necessarily multilinear) potential function $V$ is minimised.

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 minimising 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 behaviour where players play the game repeatedly and keep records of their strategies’ performance. In more detail, at each instance of the game all players $i \in N$ update the cumulative scores $U_{ia}$ of their strategies $a \in S_i$ as specified by the recursive formula:

$$U_{ia}(t + 1) = U_{ia}(t) + u_{ia}(p(t))$$

where $p(t) \in \Delta$ is the players’ strategy profile at the $t$-th iteration of the game. These scores reinforce the perceived success of each strategy as measured by the

---

3This is not true in the single-population case: for example, there exists a fully mixed evolutionarily stable strategy in the Hawk-Dove game.

4In the absence of initial bias, we assume that $U_{ia}(0) = 0$ for all $i \in N, a \in S_i$. 


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:

\[
p_{i\alpha}(t + 1) = \frac{e^{U_{i\alpha}(t+1)}}{\sum_{\beta \in S_i} e^{U_{i\beta}(t+1)}}.
\]

For simplicity, we will only consider the case where players update their scores in continuous time, i.e. according to the coupled equations:

\[
\begin{align*}
\frac{dU_{i\alpha}(t)}{dt} &= u_{i\alpha}(x(t))dt \\
x_{i\alpha}(t) &= \frac{e^{U_{i\alpha}(t)}}{\sum_{\beta} e^{U_{i\beta}(t)}}.
\end{align*}
\]

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

\[
\frac{dx_{i\alpha}}{dt} = x_{i\alpha} \left( u_{i\alpha}(x) - \sum_{\beta} x_{i\beta} u_{i\beta}(x) \right) = x_{i\alpha} \left( u_{i\alpha}(x) - u_{i}(x) \right).
\]

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

\[
x_{i\alpha}(t) = \frac{e^{\lambda_i U_{i\alpha}(t)}}{\sum_{\beta} e^{\lambda_i U_{i\beta}(t)}}
\]

where \(\lambda_i\) represents the learning rate of player \(i\), i.e. 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:

\[
\frac{dx_{i\alpha}}{dt} = \lambda_i x_{i\alpha} \left( u_{i\alpha}(x) - u_{i}(x) \right).
\]

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 [4] 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 \(x_i\):

\[
H(x_i) = -\sum_{\beta: x_{i\beta}>0} x_{i\beta} \log x_{i\beta}.
\]

That being so, we deduce that the learning rates \(\lambda_i\) act the part of (player-specific) inverse temperatures: in high temperatures (small \(\lambda_i\)), 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.

Now, the replicator dynamics were first derived in [5] 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; [8] and [14] provide excellent surveys). In both these cases, one begins with large populations of individuals that are programmed to a particular behaviour (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) $i \in N$ in some multi-population model where individuals are matched to play a game $G$ with payoff functions $u_i$. Then, the relative frequency (share) of $\alpha$ 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 $u_i(x)$, and if these payoffs represent a proportionate increase in the phenotype’s fitness (measured as the number of offspring in the unit of time), we will have:

\begin{equation}
(2.13) \quad dz_{i\alpha}(t) = z_{i\alpha}(t)u_i(x(t))dt.
\end{equation}

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

\begin{equation}
(2.14) \quad \frac{dx_{i\alpha}}{dt} = \frac{1 - x_{i\alpha}}{\sum_\beta z_{i\beta}} \frac{dz_{i\alpha}}{dt} - \sum_{\mu \neq \alpha} \frac{x_{i\mu}}{\sum_\beta z_{i\beta}} \frac{dz_{i\mu}}{dt} = x_{i\alpha}(u_{i\alpha}(x) - u_i(x)),
\end{equation}

which is exactly (2.4) 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 be specific, assume that individuals are randomly drawn from a large population and are matched against one another in a (symmetric) 2-player game $G$ with strategy space $S = \{1, \ldots S\}$ and payoff matrix $u = \{u_{\alpha\beta}\}$. Then, if $x_{\alpha}$ 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 $u(x, x) = \sum_\alpha u_\alpha(x)$. Hence, following the same procedure as above, we get the single-population replicator dynamics:

\begin{equation}
(2.15) \quad \frac{dx_{\alpha}}{dt} = x_{\alpha}(u_{\alpha}(x) - u(x, x))
\end{equation}

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

As far as rational behaviour is concerned, the replicator dynamics have some far-reaching ramifications. If we focus on multi-population models, Samuelson and Zhang showed in [7] 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 [8].

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.

\footnote{An important observation to keep in mind is that this is not the case in the dynamics (2.15).}
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):

$$dX(t) = b(X(t)) \, dt + \sum_{\beta} \sigma_{\alpha\beta}(X(t)) \, dW_{\beta}(t).$$

(2.16)

Following [15, 16], the notion of asymptotic stability in this SDE is expressed by:

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

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

for all initial conditions $X(0) = x \in V_\varepsilon$ 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 the equation (2.16), namely the **generator** $L$ of $X(t)$:

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

(2.18)

The importance of this operator can easily be 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 $Y(t) = f(X(t))$:

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

(2.19)

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

**Definition 2.4.** Let $q \in \mathbb{R}^n$ and let $U$ be an open neighbourhood 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 $x \in U$.

Accordingly, 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 314–315 of [15].

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.\(^6\) 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).\(^6\)

\(^6\)We should note here that this information has to be accurate but not necessarily *global* (as in the case of fictitious play). Just as in regret-matching schemes [17], players only need to “know the game”, i.e. the payoff that they would have received if they had chosen a strategy other than the one that they actually played. This information is usually easier to acquire than the empirical distribution of play: e.g. the received payoffs actually suffice in minority games [18].
To account for these random perturbations, we will assume that the players’ scores are now governed instead by the stochastic differential equation:

\[ dU_{ia}(t) = u_{ia}(X(t)) \, dt + \eta_{ia}(X(t)) \, dW_{ia}(t) \]

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

\[ X_{ia}(t) = \frac{e^{U_{ia}(t)}}{\sum_{\beta} e^{U_{i\beta}(t)}} \]

and \( W(t) \) is a standard Wiener process living in \( \prod_{i} \mathbb{R}^{S_i} \). The difference from the deterministic case obviously lies in the additive noise term \( \eta_{ia}(X(t)) \, dW_{ia}(t) \) where the coefficients \( \eta_{ia} \) measure the impact of the noise on the payoffs. Of course, since this impact might depend on the state of the game, these coefficients may well depend on the strategy profile \( X(t) \) itself; the only assumption that we will make is that they be continuous on \( \Delta \). In particular, if \( \eta_{ia}(x_{-i}; \alpha) = 0 \) for all \( i \in \mathbb{N}, \alpha \in S_i, x_{-i} \in \Delta_{-i} \), 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.

At any rate, to decouple equations (3.1) and (3.2), we can simply apply Itô’s lemma to the process \( X(t) \). Indeed, since \( dW_{j\beta} \cdot dW_{k\gamma} = \delta_{jk} \delta_{\beta\gamma} \, dt \) (recall that \( W(t) \) has independent components across both players and strategies), we get:

\[
dX_{ia} = \sum_{j} \sum_{\beta} \frac{\partial X_{ia}}{\partial U_{j\beta}} dU_{j\beta} + \frac{1}{2} \sum_{j,k} \sum_{\beta,\gamma} \frac{\partial^2 X_{ia}}{\partial U_{j\beta} \partial U_{k\gamma}} dU_{j\beta} \cdot dU_{k\gamma} = \sum_{\beta} \left( u_{i\beta}(X) \frac{\partial X_{ia}}{\partial U_{i\beta}} + \frac{1}{2} \eta_{i\beta}(X) \sum_{\gamma} \eta_{i\gamma}(X) \frac{\partial^2 X_{ia}}{\partial U_{i\beta} \partial U_{i\gamma}} \right) dt + \sum_{\beta} \eta_{i\beta}(X) \frac{\partial X_{ia}}{\partial U_{i\beta}} \, dW_{i\beta}
\]

and, after a few more calculations:

\[ dX_{ia} = X_{ia} \left[ u_{ia}(X) - u_i(X) \right] dt + X_{ia} \left[ \frac{1}{2} \eta_{i\beta}(X)(1 - 2X_{ia}) - \frac{1}{2} \sum_{\beta} \eta^2_{i\beta}(X)X_{i\beta}(1 - 2X_{i\beta}) \right] dt + X_{ia} \left[ \eta_{i\alpha}(X) \, dW_{ia} - \sum_{\beta} \eta_{i\beta}(X)X_{i\beta} \, dW_{i\beta} \right]. \]

On the other hand, if players update their strategies with different learning rates \( \lambda_i \), we should instead apply Itô’s formula to equation (2.10). In so doing, we obtain:

\[ dX_{ia} = \lambda_i X_{ia} \left[ u_{ia}(X) - u_i(X) \right] dt + \frac{\lambda_i^2}{2} X_{ia} \left[ \eta^2_{i\alpha}(X)(1 - 2X_{ia}) - \sum_{\beta} \eta^2_{i\beta}(X)X_{i\beta}(1 - 2X_{i\beta}) \right] dt + \lambda_i X_{ia} \left[ \eta_{i\alpha}(X) \, dW_{ia} - \sum_{\beta} \eta_{i\beta}(X)X_{i\beta} \, dW_{i\beta} \right] = b_{ia}(X) dt + \sum_{\beta} \sigma_{i,\alpha\beta}(X) \, dW_{i\beta}. \]

\(^7\)In fact, even this assumption can be relaxed: it suffices for \( \eta_{ia} \) to be measurable and bounded on \( \Delta \) (perhaps after redefinition on a null set). As it turns out, the particular form of the coefficients is not important and all that matters is their worst value: as long as they are bounded (which is more than reasonable from a practical point of view), our results will not be affected.
where, in obvious notation, \( b_{i\alpha}(x) \) and \( \sigma_{i,\alpha\beta}(x) = \lambda_i x_{i\alpha} \eta_{i\beta}(x)(\delta_{\alpha\beta} - x_{\beta}), x \in \Delta \)
are respectively the drift and diffusion coefficients of the diffusion \( X(t) \). Obviously, when \( \lambda_i = 1 \), we recover the uniform dynamics (3.4); equivalently (and this is an interpretation that is well worth keeping in mind), the rates \( \lambda_i \) can simply be regarded as a commensurate inflation of the payoffs and noise coefficients of player \( i \in \mathbb{N} \) in the uniform logistic model (3.2).

Equation (3.4) and its rate-adjusted sibling (3.4') 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 SDE’s (e.g. theorem 5.2.1 in [19]). Instead, an addition over \( \alpha \in S_i \) reveals that every simplex \( \Delta_i \subseteq \Delta \) remains invariant under (3.4): 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 \( \Delta_i \) for all \( t \geq 0 \).

So, if \( \phi \) is a smooth bump function that is equal to 1 on some open neighbourhood of \( U \supset \Delta \) and which vanishes outside some compact set \( K \supset U \), the SDE
\[
(3.5) \quad dX_{i\alpha} = \phi(X) \left( b_{i\alpha}(X) \, dt + \sum_{\beta} \sigma_{i,\alpha\beta}(X) \, dW_{i\beta} \right)
\]
will have bounded diffusion and drift coefficients and will thus admit a unique strong solution. But since this last equation agrees with (3.4) on \( \Delta \) and any solution of (3.4) always stays in \( \Delta \), 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 (3.4) (3.4’) with the “aggregate shocks” approach of Fudenberg and Harris [9] 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 \( \alpha \in S_i \) of the species \( i \in \mathbb{N} \) in some multi-population evolutionary game \( \mathcal{G} \), its growth will be determined by:
\[
(3.6) \quad dZ_{i\alpha}(t) = Z_{i\alpha}(t)(u_{i\alpha}(X(t)) \, dt + \eta_{i\alpha} \, dW_{i\alpha}(t))
\]
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:
\[
(3.7) \quad dX_{i\alpha} = X_{i\alpha} \left[ (u_{i\alpha}(X) - u_i(X)) - \left( \eta_{i\alpha}^2 X_{i\alpha} - \sum_{\beta} \eta_{i\beta}^2 X_{i\beta} \right) \right] \, dt
\]
\[+ X_{i\alpha} \left[ \eta_{i\alpha} \, dW_{i\alpha} - \sum_{\beta} \eta_{i\beta} X_{i\beta} \, dW_{i\beta} \right].
\]

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 (3.4) and (3.7) 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
\[
(3.8) \quad X_{i\alpha} \left[ \frac{1}{2} \eta_{i\alpha}^2 - \frac{1}{2} \sum_{\beta} \eta_{i\beta}^2 X_{i\beta} \right] \, dt.
\]

\footnote{Actually, it is not harder to see that every face of \( \Delta \) is a trap for \( X(t) \).}
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.

Moreover, this difference suggests that we can shift from (3.4) to (3.7) simply by modifying the game’s payoffs to
\[ \tilde{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2} \eta_{i\alpha}^2 \]
This modified game was precisely the one that came up in the analysis of [11,12] and it plays a pivotal role in setting apart learning and evolution in a stochastic setting. In effect, whereas this modification seems deeply ingrained in the process of natural selection, exponential learning gives players a clearer picture of the actual underlying game.

4. Extinction of Dominated Strategies

Thereby armed with the stochastic replicator equations (3.4), (3.4′) 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; only the rationally admissible ones survive.

As in [10] (implicitly) and [11] (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 \( i \in N \):
\[
H(q_i, x_i) := -\sum_{\alpha: q_{i\alpha} > 0} q_{i\alpha} \log(x_{i\alpha}) \equiv H(q_i) + d_{KL}(q_i, x_i)
\]
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):
\[
d_{KL}(q_i, x_i) := H(q_i, x_i) - H(q_i) = \sum_{\alpha: q_{i\alpha} > 0} q_{i\alpha} \log \frac{x_{i\alpha}}{q_{i\alpha}}.
\]
This divergence function is central in the stability analysis of the (deterministic) replicator dynamics because it serves as a distance measure in probability space [8]. 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 \( d_{KL}(q_i, x_i) = \infty \) iff \( x_i \) does not employ a pure strategy \( \alpha \in S_i \) that is present in \( q_i \). Therefore, if \( d_{KL}(q_i, x_i) = \infty \) for all dominated strategies \( q_i \) of player \( i \), it immediately follows that \( x_i \) cannot be dominated itself. In this vein, we have:

**Proposition 4.1.** Let \( X(t) \) be a solution of the stochastic replicator dynamics (3.4) for some interior initial condition \( X(0) = x \in \text{Int}(\Delta) \). Then, if \( q_i \in \Delta_i \) is (strictly) dominated:
\[
\lim_{t \to \infty} d_{KL}(q_i, X_i(t)) = \infty \quad \text{almost surely.}
\]
In particular, if \( q_i = \alpha \in S_i \) 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 \( X(0) = x \in \text{Int}(\Delta) \) and hence, \( X_i(t) \) will almost surely stay in \( \text{Int}(\Delta_i) \) for all \( t \geq 0 \); this is a simple consequence of the uniqueness of strong solutions and the invariance of the faces of \( \Delta_i \) under the dynamics (3.4).

---

9Strictly speaking, this presumes that the noise coefficients \( \eta_{i\alpha} \) be constant; the general case requires us to allow for games whose payoffs may not be multilinear. This would not really change our results but, for the time being, we prefer to avoid this complication.
Let us now consider the cross entropy $G_{q_i}(t)$ between $q_i$ and $X_i(t)$:

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

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:

$$
dG_{q_i} = \sum_\beta \frac{\partial G_{q_i}}{\partial X_{i\beta}} dX_{i\beta} + \frac{1}{2} \sum_\beta \sum_\gamma \frac{\partial^2 G_{q_i}}{\partial X_{i\beta} \partial X_{i\gamma}} dX_{i\beta} \cdot dX_{i\gamma}
$$

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

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

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

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

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

and, after integrating:

$$
G_{q_i' - q_i}(t) = \int_0^t u_i(X - i; q_i') ds + \sum_\beta (q_{i'\beta} - q_{i\beta}) \int_0^t \eta_{i\beta}(X(s)) dW_{i\beta}(s)
$$

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

On the other hand, 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:

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

where $S_i = |S_i|$ is the number of pure strategies available to player $i$ and $\eta_i = \max\{|\eta_{i\beta}(x)| : x \in \Delta_i, \beta \in \Delta_i\}$ recall also that $q_i, q_i' \in \Delta_i$ for the last step. Therefore, if $\psi_i(t) = \int_0^t \xi_i(s) ds$ denotes the martingale part of (4.10) and $\rho_i(t)$ is its quadratic variation, the previous inequality becomes:

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

However, by the time-change theorem for martingales (e.g. theorem 3.4.6 in [20]), there exists a Wiener process $\tilde{W}_i$ such that $\psi_i(t) = \tilde{W}_i(\rho_i(t))$. Hence, by the law of

---

10If the coefficients $\eta_{i\beta}$ are not continuous but only bounded (perhaps after redefinition on a null set), one should take the essential supremum instead and set $\eta_i = \text{ess sup}\{\eta_{i\beta}\}$. 

the iterated logarithm we get:
\[
\liminf_{t \to \infty} G_{q_i - q_i'}(t) \geq \liminf_{t \to \infty} \left( H(q_i - q_i', x) + v_i t + \tilde{W}_i(\rho_i(t)) \right)
\]
\[
\geq \liminf_{t \to \infty} \left( v_i t - \sqrt{2 \rho_i(t) \log \log \rho_i(t)} \right)
\]
\[
(4.11)
\]
\[
\geq \liminf_{t \to \infty} \left( v_i t - 2 \eta \sqrt{S_i t \log \log(2S_i \eta t^2)} \right) = \infty \text{ (a.s.)}
\]
Since \( G_{q_i}(t) \geq G_{q_i}(t) - G_{q_j}(t) \to \infty \), it follows that \( \lim_{t \to \infty} d_{KL}(q_i, X_i(t)) = \infty \) (almost surely) and, with \( G_{\alpha}(t) = -\log X_{\alpha}(t) \) for all pure strategies \( \alpha \in S_i \), our proof is complete.

As in [11], we can now obtain the following estimate for the lifespan of pure dominated strategies:

**Proposition 4.2.** Let \( X(t) \) be an interior solution path of \([3,4]\) with initial condition \( X(0) = x \in \text{Int}(\Delta) \) and let \( P_x \) denote its law. Assume further that the strategy \( \alpha \in S_i \) is dominated; then, for any \( M > 0 \) and for \( t \) large enough, we have:
\[
P_x \{ X_{\alpha}(t) < e^{-M} \} \geq \frac{1}{2} \text{erfc} \left( \frac{M - h_i(x_i) - v_i t}{2 \eta \sqrt{S_i t}} \right)
\]
where \( S_i = |S_i| \) is the number of strategies available to player \( i \), \( \eta_i = \max\{|\eta_{i\beta}(y)| : y \in \Delta, \beta \in S_i\} \) 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 [11]. Surely enough, if \( \alpha < p_i \in \Delta_i \) and we use the same notation as in the proof of proposition 4.1 we have:
\[
-\log X_{\alpha}(t) = G_{\alpha}(t) \geq G_{\alpha}(t) - G_{p_i}(t) \geq H(\alpha, x) - H(p_i, x) + v_i t + \tilde{W}_i(\rho_i(t))
\]
\[
= h_i(x_i) + v_i t + \tilde{W}_i(\rho_i(t))
\]
where \( v_i := \min_x \{ u_i(x_{-i} ; p_i) - u_i(x_{-i} ; \alpha) \} \) and \( h_i(x_i) := \log x_{\alpha} - \sum \beta p_{i\beta} \log x_{i\beta} \) are both positive. Then:
\[
P_x(X_{\alpha}(t) < e^{-M}) \geq P_x \left\{ \tilde{W}_i(\rho_i(t)) > M - h_i(x_i) - v_i t \right\}
\]
\[
= \frac{1}{2} \text{erfc} \left( \frac{M - h_i(x_i) - v_i t}{\sqrt{2} \rho_i(t)} \right)
\]
and, since the quadratic variation \( \rho_i(t) \) is bounded above by \( 2S_i \eta^2 t \) (eq. (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 [10] and Imhof [11] 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 \([3,4]\) and, in particular, the extra terms that are propagated there by exponential learning and which are absent from the aggregate shocks dynamics \([3,7]\). 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 \( \bar{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2} \eta_{i\alpha}^2 \) that come up in [11, 12] (and, indirectly, in [10] as well).

\[\text{It should be noted here again that the single-population dynamics studied in [11] are even further differentiated from [3,4] by the fact that they are cubic and not quadratic.}\]
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 an interior solution path of (3.4) starting at $X(0) = x \in \text{Int}(\Delta)$. Then, if $q_i \in \Delta$, is iteratively dominated:

$$\lim_{t \to \infty} d_{KL}(q_i, X_i(t)) = \infty \quad \text{almost surely},$$

i.e. only rationally admissible strategies survive in the long run.

**Proof.** As in the deterministic case [7], the main idea is that the solution path $X(t)$ gets progressively closer to the faces of $\Delta$ that are spanned by the pure strategies which have not yet been eliminated. Following [10], 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 $\text{Adm}(A_i, A_{-i})$ 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 $A_i^0 = \Delta_i$, and $A_{-i}^0 = \prod_{j \neq i} A_j^0$, we may define inductively the set of strategies that remain admissible after $n$ elimination rounds by $A_i^n := \text{Adm}(A_i^{n-1}, A_{-i}^{n-1})$ where $A_{-i}^{n-1} := \prod_{j \neq i} A_j^{n-1}$; similarly, the pure strategies that have survived after $n$ such rounds will be denoted by $S_i^n := S_i \cap A_i^n$. Clearly, this sequence forms a descending chain $A_i^0 \supseteq A_i^1 \supseteq \ldots$ and the set $A_i^\infty := \bigcap_{n=0}^\infty A_i^n$ will consist precisely of the strategies of player $i$ that are rationally admissible.

Assume then that the cross entropy $G_{q_i}(t) = H(q_i, X_i(t)) = -\sum q_{i\alpha} \log X_{i\alpha}(t)$ diverges as $t \to \infty$ for all strategies $q_i \notin A_i^k$ that die out within the first $k$ rounds; in particular, if $\alpha \notin S_i^k$ 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 $q_i \in A_i^k$ but $q_i \notin A_i^{k+1}$, there will exist some $q_i' \in A_i^{k+1}$ such that:

$$u_i(x_{-i}; q_i') > u_i(x_{-i}; q_i) \quad \text{for all } x_{-i} \in A_i^{k+1}.$$

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

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

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

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

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

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

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

As a result, if there exists only one rationally admissible strategy, we get:

**Corollary 4.4.** Let $X(t)$ be an interior solution path of the replicator equation [3.4] for some dominance-solvable game $\mathcal{G}$ and let $x_0 \in \delta$ be the (unique) strict equilibrium of $\mathcal{G}$. Then:

$$\lim_{t \to \infty} X(t) = x_0$$

i.e. 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 [3.4] as well: this is just a matter of rescaling. The only difference from using different learning rates $\lambda_i$ comes about in proposition [4.2] where the estimate (4.12) becomes

$$\mathbb{P}_x \left\{ \frac{X_{i\alpha}(t)}{x_{i\alpha}(t)} < e^{-M} \right\} \geq \frac{1}{2} \text{erfc} \left( \frac{M - h_i(x_i) - \lambda_i v_i t}{2 \lambda_i \eta_i \sqrt{S_i t}} \right).$$

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 $\lambda_i$ also carries a commensurate increase in the noise coefficients $\eta_i$, 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 analysed 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. To wit, assume that player $i \in \mathbb{N}$ 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:

$$\frac{dV}{dt} = \sum_{i,\alpha} \frac{\partial V}{\partial x_{i\alpha}} \frac{dx_{i\alpha}}{dt} = - \sum_{i,\alpha} u_{i\alpha}(x) \lambda_i x_{i\alpha} (u_{i\alpha}(x) - u_i(x))$$

$$= - \sum_i \lambda_i \left( \sum_{\alpha} x_{i\alpha} u^2_{i\alpha}(x) - u^2_i(x) \right) \leq 0$$

where the last step stems from Jensen’s inequality - recall that $\frac{\partial V}{\partial x_{i\alpha}} = -u_{i\alpha}(x)$ on account of equation [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 \( \Delta \).

It is therefore reasonable to ask whether similar conclusions can be drawn in the noisy setting of (5.3). 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 (2.3)\footnote{As mentioned before, to avoid unnecessary complications, we plead guilty to a slight abuse of terminology in assuming that all equilibria in pure strategies are also strict.}. 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 neighbourhood of \( q_0 \). We are thus left to examine the negativity condition of definition (2.4)\footnote{Strictly speaking, since our analysis is constrained on \( \Delta \), the “neighbourhoods” of definitions (2.3) and (2.4) should be taken to mean “neighbourhoods in \( \Delta \), i.e. neighbourhoods in the subspace topology of \( \Delta \hookrightarrow \prod_i R^S_i \). This minor point should always be clear from the context.} i.e. whether there exists some \( k > 0 \) such that \( LV(x) \leq -kV(x) \) 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 (3.1) applied to \( V \) produces:

\[
LV(x) = -\sum_{i,\alpha} \lambda_i x_{i,\alpha} u_{i,\alpha}(x) (u_{i,\alpha}(x) - u_i(x))
- \sum_{i,\alpha} \frac{\lambda^2}{2} x_{i,\alpha} u_{i,\alpha}(x) \left( \eta_{i,\alpha}^2 (1 - 2x_{i,\alpha}) - \sum_{\beta} \eta_{i,\beta}^2 x_{i,\beta} (1 - 2x_{i,\beta}) \right)
\]

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

So, let \( \varepsilon > 0 \) and consider the perturbed strategies \( x_i = (1 - \varepsilon_i) e_{i,0} + \varepsilon_i y_i \) with \( y_i \) belonging to the face of \( \Delta \) that lies opposite to \( e_{i,0} \) (i.e. \( y_{i,\mu} \geq 0, \mu = 1,2,\ldots \) and \( \sum_\mu y_{i,\mu} = 1 \)). After a series of calculations, we obtain:

\[
\sum_\alpha x_{i,\alpha} u_{i,\alpha}(x) (u_{i,\alpha}(x) - u_i(x))
= \varepsilon_i^2 u_{i,0}^2 (q_0) + \varepsilon_i \sum_\mu y_{i,\mu} (u_{i,\mu}^2(q_0) - 2u_{i,\mu}(q_0)u_{i,\alpha}(q_0)) + O(\varepsilon_i^2)
\]

(5.3a)

and that

\[
\sum_\alpha x_{i,\alpha} u_{i,\alpha}(x) (u_i - u_{i,\alpha} x_{i,\alpha})
= -\varepsilon_i \left( \eta_{i,0}^2 + \sum_\mu y_{i,\mu} \eta_{i,\mu}^2 \right) + \varepsilon_i \sum_\mu y_{i,\mu} u_{i,\mu}(q) \left( \eta_{i,\mu}^2 + \eta_{i,0}^2 \right) + O(\varepsilon_i^2)
\]

(5.3b)

and, finally:

\[
V(x) = \sum_\alpha \varepsilon_i \Delta u_{i,\mu} + O(\varepsilon^2)
\]

where \( \varepsilon^2 = \sum_i \varepsilon_i^2 \). Therefore, if we combine equations (5.3a)–(5.3c), the condition \( LV(x) \leq -kV(x) \) becomes:

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

(5.4)
and if \( \Delta u_{i\mu} > \frac{\lambda_i}{2}(\eta^2_{i\mu} + \eta^2_{i,0}) \) for all \( \mu \in S_i \setminus \{0\} \), this last inequality will be satisfied for some \( k > 0 \) whenever \( \varepsilon \) is small enough. Essentially, we have proven:

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

\[
V(q-i, \mu) > \frac{\lambda_i}{2}(\eta^2_{i\mu} + \eta^2_{i,0}).
\]

Then \( q \) is stochastically asymptotically stable in the rate-adjusted dynamics (5.7).

We thus see that no matter how loud the noise \( \eta_i \) might be, stochastic stability is always guaranteed if the players choose a learning rate that is slow enough and if \( \Delta q \) is small enough. Essentially, we have proven:

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

\[
V(q-i, \mu) > \frac{\lambda_i}{2}(\eta^2_{i\mu} + \eta^2_{i,0}).
\]

Then \( q \) is stochastically asymptotically stable in the rate-adjusted dynamics (5.7).

We thus see that no matter how loud the noise \( \eta_i \) 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. \( \lambda_i < \Delta V_i/\eta^2_i \)). 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 \( \lambda_i \). On top of that, the very form of the condition (5.6) is strongly reminiscent of the “modified” game of [11, 12], a similarity which seems to contradict our statement that exponential learning favours rational reactions in the original game. The catch here is that condition (5.5) 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 (5.5) altogether.

### 5.2. The Dyadic Case.

To gain some further intuition into why the condition (5.5) is redundant, it will be particularly helpful to examine the case where players compete for the resources of only two facilities (i.e. \( S_i = \{0, 1\} \), \( i \in \mathbb{N} \)) and try to learn the game with the help of the uniform replicator equation (3.4). This is the natural setting for the El Farol bar problem [21] and the ensuing minority game [18] 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 [22], these games always possess strict equilibria, even when players have distinct payoff functions. So, by relabeling indices if necessary, 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 (3.4) takes the form:

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

where now \( \Delta u_i \equiv u_{i,0} - u_{i,1} \) and \( \eta^2_i = \eta^2_{i,0} + \eta^2_{i,1} \).

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, (5.6) assumes the astoundingly suggestive guise:

\[
L = \sum_i \left( \Delta u_i \frac{\partial}{\partial y_i} + \frac{1}{2} \eta^2_i \frac{\partial^2}{\partial y_i^2} \right)
\]

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...
\[ f(y) = \sum_i e^{-a_i y_i} (a_i > 0): \]

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

Indeed, 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:

\[ Lf(y) \leq -\sum_i a_i m_i e^{-a_i y_i} \leq -kf(y) \]

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 equilibrium \( q_0 \)), a trivial modification of the stochastic Lyapunov method (see e.g. pp. 314–315 of [15]) yields:

**Proposition 5.2.** The strict equilibria of minority games are stochastically asymptotically stable in the uniform replicator equation (3.4).

**Remark 5.2.1.** It is trivial to see that strict equilibria of minority games will also be stable in the rate-adjusted dynamics (3.4′): 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 5.2.2.** A closer inspection of the calculations leading to proposition 5.2 reveals that nothing hinges on the minority mechanism per se: it is (5.7) 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 \( i \in N \)). 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 5.3.** The strict equilibria of dyadic games are stochastically asymptotically stable in the replicator dynamics (3.4), (3.4′) 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 multi-population replicator dynamics (2.9) [8]. 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 5.3).

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 \( \mathcal{G} \) are stochastically asymptotically stable in the replicator dynamics (3.4), (3.4′) of exponential learning.

---

14 Instead, if players have more than 2 strategies, the not-so-convenient analogue of (5.7) is:

\[ Lf(y) = \sum_i \left[ u_{i,0} - \sum_\mu u_{i_\mu} y_{i_\mu} - \frac{1}{2} \sum_\mu \eta_{i_\mu} y_{i_\mu} (1 - y_{i_\mu}) \right] \frac{\partial f}{\partial y_{i,0}} + \frac{1}{2} \left[ \eta_{i,0}^2 + \sum_\mu \eta_{i_\mu}^2 \eta_{i_\mu}^2 \right] \frac{\partial^2 f}{\partial y_{i,0}^2}, \]

where \( y_{i,0} = \text{logit} x_{i,0}, \ y_{i_\mu} = x_{i_\mu}/(1 - x_{i,0}) \) and we are assuming that \( f \) does not depend on the \( y_{i_\mu} \) variables.
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 \( q_0 = (\epsilon_1, 0, \ldots, \epsilon_N, 0) \) is a strict equilibrium of \( \Theta \). Then, if \( q_0 \) is to be stochastically stable, say in the uniform dynamics \( \{3.4\} \), one would expect the strategy scores \( U_{i,0} \) of player \( i \) to grow much faster than the scores \( U_{i,\mu}, \mu \in S_i \setminus \{0\} \) of his other strategies. This is captured remarkably well by the “adjusted” scores:

\[
\begin{align*}
Z_{i,0} &= \lambda_i U_{i,0} - \log \left( \sum_{\mu} e^{\lambda_i U_{i,\mu}} \right), \\
Z_{i,\mu} &= \lambda_i (U_{i,\mu} - U_{i,0})
\end{align*}
\]

where \( \lambda_i > 0 \) is a sensitivity parameter akin (but not identical) to the learning rates of proposition [5.1]. And, since it is plausible to expect the SDE \( \{6.3\} \) to behave in a similar fashion.

To that end, after some calculations in the \( \{Y_{i,0}\} \) coordinates, Itô’s lemma gives:

\[
\begin{align*}
\frac{dY_{i,0}}{dt} &= \lambda_i Y_{i,0} \left[ u_{i,0} - \sum_{\mu} Y_{i,\mu} u_{i,\mu} + \frac{\lambda_i}{2} \eta_{i,0}^2 - \frac{\lambda_i}{2} \sum_{\mu} Y_{i,\mu} (1 - 2Y_{i,\mu}) \eta_{i,\mu}^2 \right] dt \\
&+ \lambda_i Y_{i,0} \left[ \eta_{i,0} dW_{i,0} - \sum_{\mu} \eta_{i,\mu} Y_{i,\mu} dW_{i,\mu} \right] \\
\frac{dY_{i,\mu}}{dt} &= \lambda_i Y_{i,\mu} \left[ u_{i,\mu} - \sum_{\nu} u_{i,\nu} Y_{i,\nu} \right] dt \\
&+ \frac{\lambda_i^2}{2} Y_{i,\mu} \left[ \eta_{i,\mu}^2 (1 - 2Y_{i,\mu}) - \sum_{\nu} \eta_{i,\nu}^2 Y_{i,\nu} (1 - 2Y_{i,\nu}) \right] dt \\
&+ \lambda_i Y_{i,\mu} \left[ \eta_{i,\mu} dW_{i,\mu} - \sum_{\nu} \eta_{i,\nu} Y_{i,\nu} dW_{i,\nu} \right].
\end{align*}
\]

where we have suppressed the arguments of \( u_i \) and \( \eta_i \) in order to reduce notational clutter. In fact, this last SDE is particularly revealing: roughly speaking, we see that if \( \lambda_i \) 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 [6.2]). And, since we know that strict equilibria are asymptotically stable in the deterministic case, it is plausible to expect the SDE \( \{6.3\} \) 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 \( \{3.4\} \) by looking at the processes \( Y_i = (Y_{i,0}, Y_{i,1}, \ldots) \in \mathbb{R} \times \Delta^{S_i - 1} \) of equation [6.2]. In these coordinates, we just need to show that for every \( M_i > 0, i \in \mathbb{N} \) 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}(t) = \infty and Y_{i}(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 \( Y = (Y_1, \ldots, Y_N) \), i.e. the second order differential operator:

\[
L = \sum_{i \in N} b_{i\alpha}(y) \frac{\partial}{\partial y_{i\alpha}} + \frac{1}{2} \sum_{\alpha, \beta \in S_i} \left( \sigma_i(y) \sigma_i^T(y) \right)_{\alpha\beta} \frac{\partial^2}{\partial y_{i\alpha} \partial y_{i\beta}}
\]

where \( b_i \) and \( \sigma_i \) are the drift and diffusion coefficients of the SDE (5.3) respectively. In particular, if we restrict our attention to sufficiently smooth functions of the form \( f(y) = \sum_{i \in N} f_i(y_{i,0}) \), the application of \( L \) yields:

\[
L f(y) = \sum_{i \in N} \lambda_i y_{i,0} \left[ u_{i,0} + \frac{\lambda_i}{2} \eta_{i,0}^2 - \sum_{\mu} y_{i\mu} \left( u_{i\mu} - \frac{\lambda_i}{2} (1 - 2 y_{i\mu}) \eta_{i\mu}^2 \right) \right] \frac{\partial f_i}{\partial y_{i,0}} + \frac{1}{2} \sum_{i \in N} \lambda_i^2 y_{i,0}^2 \left( \eta_{i,0}^2 + \sum_{\mu} \eta_{i\mu}^2 y_{i\mu}^2 \right) \frac{\partial^2 f_i}{\partial y_{i,0}^2}.
\]

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 \( \frac{\partial^2 f}{\partial y_{i,0}^2} = 2/y_{i,0}^3 \), equation (6.5) becomes:

\[
L f(y) = -\sum_{i \in N} \frac{\lambda_i}{y_{i,0}} \left[ u_{i,0} - \sum_{\mu} u_{i\mu} y_{i\mu} - \frac{\lambda_i}{2} \eta_{i,0}^2 - \frac{\lambda_i}{2} \sum_{\mu} y_{i\mu} (1 - y_{i\mu}) \eta_{i\mu}^2 \right].
\]

However, since \( q_0 = (e_1, \ldots, e_{N,0}) \) has been assumed to be a strict Nash equilibrium of \( \mathcal{O} \), we will have \( u_{i,0}(q_0) > u_{i\mu}(q_0) \) for all \( \mu \in S_i \setminus \{0\} \). 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 \( \eta_i = \max\{|\eta_{i\beta}(x)| : x \in \Delta, \beta \in S_i\} \) and pick positive \( \lambda_i \) with \( \lambda_i < v_i/\eta_i^2 \), we get:

\[
L f(y) \leq -\sum_{i \in N} \frac{\lambda_i v_i}{2} \frac{1}{y_{i,0}} \leq -\frac{1}{2} \min_i \{\lambda_i v_i\} f(y)
\]

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 (3.4), the proof is similar and only entails a rescaling of the parameters \( \lambda_i \).

\( \square \)

Remark 6.1.1. If we trace our steps back to the coordinates \( X_{i\alpha} \), our Lyapunov candidate takes the form \( f(x) = \sum_i \left( x_{i,0}^{-\lambda_i} \sum_{\mu} x_{i\mu}^{\lambda_i} \right) \). It thus begs to be compared to the Lyapunov function \( \sum_{\mu} x_{i\mu}^{\lambda} \) employed by Imhof and Hofbauer in [12] to derive a conditional version of theorem 6.1 in the evolutionary setting. In fact, the obvious extension \( f(x) = \sum_i \sum_{\mu} x_{i\mu}^{\lambda} \) also works in our case, but the calculations are much more cumbersome and they are also shorn of their ties to the adjusted scores (6.1).

Remark 6.1.2. It is also important to highlight the dual role that the learning rates \( \lambda_i \) play in our analysis. In the logistic learning model [21], 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 (3.4). 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 \(f(x) = \sum \lambda_i x_i^{\lambda_i} \sum \mu_i x_i^\mu\) is Lyapunov only if the sensitivity parameters \(\lambda_i\) 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 their intimate relation.

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 [3]. 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 [8]. Therefore, since our learning model [2.14] 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 [3.1] is:

\[
\begin{aligned}
\text{(7.1)} & \quad dX_\alpha = X_\alpha \left[ \left( u_\alpha(X) - u(X, X) \right) - \left( \eta^2_\alpha X_\alpha - \sum \eta^2_\beta X_\beta \right) \right] dt \\
& \quad + X_\alpha \left[ \eta_\alpha dW_\alpha - \sum \eta_\beta X_\beta dW_\beta \right].
\end{aligned}
\]

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 [11]). Theorem \(6.1\) rules out such behaviour 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 \(3.1\) be recurrent?

This question is equivalent to showing that a profile \(x\) is stochastically asymptotically stable in the replicator equations \(6.1\), \(3.1\) 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 the future.

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:

\[
\text{(7.2)} \quad dU_{i\alpha}(t) = u_{i\alpha}(X(t)) dt + \eta_{i\alpha} dW_{i\alpha}(t).
\]
Then, after a few calculations, the corresponding replicator equation reads:

\[
\partial X_{i\alpha} = X_{i\alpha} \left( u_{i\alpha}(X) - u_i(X) \right) dt + X_{i\alpha} \left( \eta_{i\alpha} \partial W_{i\alpha} - \sum \eta_{i\beta} X_{i\beta} \partial W_{i\beta} \right).
\]

The form of this last equation is remarkably suggestive. First, it highlights the role of the modified game \( \tilde{u}_{i\alpha} = u_{i\alpha} + \frac{1}{2} \eta_{i\alpha}^2 \) even more crisply than equation (3.4): the payoff terms are completely decoupled from the noise, in contrast to what one obtains by introducing Stratonovich perturbations in the evolutionary setting [12,23]. 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.

Acknowledgements

We would like to extend our gratitude to David Leslie from the university of Bristol for the fruitful discussions and his helpful comments 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 [24].

References

[1] J. F. Nash, “Non-cooperative games,” The Annals of Mathematics, vol. 54, pp. 286–295, September 1951.
[2] R. J. Aumann, “Subjectivity and correlation in randomized strategies,” Journal of Mathematical Economics, vol. 1, pp. 67–96, March 1974.
[3] 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.
[4] D. Fudenberg and D. K. Levine, The Theory of Learning in Games, vol. 2 of MIT Press Series on Economic Learning and Social Evolution. The MIT Press, 1998.
[5] P. D. Taylor and L. B. Jonker, “Evolutionary stable strategies and game dynamics,” Mathematical Biosciences, vol. 40, no. 1-2, pp. 145–156, 1978.
[6] P. Schuster and K. Sigmund, “Replicator dynamics,” Journal of Theoretical Biology, vol. 100, no. 3, pp. 533–538, 1983.
[7] L. Samuelson and J. Zhang, “Evolutionary stability in asymmetric games,” Journal of Economic Theory, vol. 57, pp. 363–391, 1992.
[8] Jürgen W. Weibull, Evolutionary Game Theory. The MIT Press, 1995.
[9] D. Fudenberg and C. Harris, “Evolutionary dynamics with aggregate shocks,” Journal of Economic Theory, vol. 57, pp. 420–441, August 1992.
[10] A. Cabrales, “Stochastic replicator dynamics,” International Economic Review, vol. 41, 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] D. Monderer and L. S. Shapley, “Potential games,” Games and Economic Behavior, vol. 14, no. 1, pp. 124 – 143, 1996.
[14] J. Hofbauer and K. Sigmund, The Theory of Evolution and Dynamical Systems. Cambridge University Press, 1988.
[15] I. I. Gikhman and A. V. Skorokhod, Stochastische Differentialgleichungen. Akademie-Verlag, 1971.
[16] L. Arnold, Stochastic Differential Equations: Theory and Applications. Wiley, 1974.
[17] S. Hart and A. Mas-Colell, “A simple adaptive procedure leading to correlated equilibrium,” *Econometrica*, vol. 68, pp. 1127–1150, September 2000.

[18] M. Marsili, D. Challet, and R. Zecchina, “Exact solution of a modified El Farol’s bar problem: Efficiency and the role of market impact,” *Physica A*, vol. 280, pp. 522–553, 2000.

[19] B. Øksendal, *Stochastic Differential Equations*. Springer-Verlag, 6 ed., 2006.

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

[21] W. B. Arthur, “Inductive reasoning and bounded rationality (the El Farol problem),” *Am. Econ. Assoc. Papers Proc.*, vol. 84, pp. 406–411, 1994.

[22] I. Milchtaich, “Congestion games with player-specific payoff functions,” *Games and Economic Behavior*, vol. 13, pp. 111–124, 1996.

[23] R. Z. Has’minskii and N. Potsepun, “On the replicator dynamics behavior under Stratono-vich type random perturbations,” *Stochastic Dynamics*, vol. 6, pp. 197–211, 2006.

[24] P. Mertikopoulos and A. L. Moustakas, “Learning in the presence of noise,” in *GameNets ’09: Proceedings of the 1st International Conference on Game Theory for Networks*, May 2009.

Department of Physics, National and Kapodistrian University of Athens, 15784 Athens, Greece

E-mail address: {pmertik, arislm}@phys.uoa.gr