HIGHER ORDER GAME DYNAMICS

RIDA LARAKI AND PANAYOTIS MERTIKOPOULOS*

ABSTRACT. Continuous-time dynamics for games are typically first order systems where payoffs determine the growth rate of the players' strategy shares. In this paper, we investigate what happens beyond first order by viewing payoffs as higher order forces of change, specifying e.g. the acceleration of the players' evolution instead of its velocity (a viewpoint which emerges naturally when it comes to aggregating empirical data of past instances of play). To that end, we derive a wide class of higher order game dynamics, generalizing first order imitative dynamics, and, in particular, the replicator dynamics. We show that strictly dominated strategies become extinct in n-th order payoff-monotonic dynamics n orders as fast as in the corresponding first order dynamics; furthermore, in stark contrast to first order, weakly dominated strategies also become extinct for $n \geq 2$. All in all, higher order payoff-monotonic dynamics lead to the elimination of weakly dominated strategies, followed by the iterated deletion of strictly dominated strategies, thus providing a dynamic justification of the well-known epistemic rationalizability process of Dekel and Fudenberg (1990). Finally, we also establish a higher order analogue of the folk theorem of evolutionary game theory, and we show that convergence to strict equilibria in n-th order dynamics is n orders as fast as in first order.

1. INTRODUCTION.

Owing to the considerable complexity of computing Nash equilibria and other rationalizable outcomes in non-cooperative games, a fundamental question that arises is whether these outcomes may be regarded as the result of a dynamic learning process where the participants "accumulate empirical information on the relative advantages of the various pure strategies at their disposal" (Nash, 1950, p. 21). To that end, numerous classes of game dynamics have been proposed (from both a learning and an evolutionary "mass-action" perspective), each with its own distinct set of traits and characteristics – see e.g. the comprehensive survey by Sandholm (2010) for a most recent account.

Be that as it may, there are few rationality properties that are shared by a decisive majority of game dynamics. For instance, if we focus on the continuous-time, deterministic regime, a simple comparison between the well-known replicator

dynamics (Taylor and Jonker, 1978) and the Smith dynamics (Smith, 1984) reveals that game dynamics can be imitative (replicator) or innovative (Smith), rest points might properly contain the game's Nash set or coincide with it (Hofbauer and Sandholm, 2009), and strictly dominated strategies might become extinct (Samuelson and Zhang, 1992) or instead survive (Hofbauer and Sandholm, 2011). In fact, negative results seem to be much more ubiquitous: there is no class of uncoupled game dynamics that always converges to equilibrium (Hart and Mas-Colell, 2003) and weakly dominated strategies may survive in the long run, even in simple 2×2 games (Samuelson, 1993; Weibull, 1995).

From a mathematical standpoint, the single unifying feature of the vast majority of game dynamics is that they are first order dynamical systems. Interestingly however, this restriction to first order is not present in the closely related field of optimization (corresponding to games against nature): as it happens, the second order "heavy ball with friction" method studied by Alvarez (2000) and Attouch, Goudou, and Redont (2000) has some remarkable optimization properties that first order schemes do not possess. In particular, by interpreting the gradient of the function to be maximized as a physical, Newtonian force (and not as a first order vector field to be tracked by the system's trajectories), one can give the system enough energy to escape the basins of attraction of local maxima and converge instead to the *global* maximum of the objective function (something which is not possible in ordinary first order dynamics).

This, therefore, begs the question: *can second (or higher) order dynamics be introduced and justified in a game theoretic setting?* And if yes, *do they allow us to obtain better convergence results and/or escape any of the first order impossibility results?*

The first challenge to overcome here is that second order methods in optimization apply to *unconstrained* problems, whereas game dynamics must respect the (constrained) structure of the game's strategy space. To circumvent this constraint, Flåm and Morgan (2004) proposed a heavy-ball method as in Attouch et al. (2000) above, and they enforced consistency by projecting the orbits' velocity to a subspace of admissible directions when the updating would lead to inadmissible strategy profiles (say, assigning negative probability to an action). Unfortunately, as is often the case with projection-based schemes (see e.g. Sandholm, Dokumacı, and Lahkar, 2008), the resulting dynamics are not continuous, so even basic existence and uniqueness results are hard to obtain.

On the other hand, if players try to improve their performance by aggregating information on the relative payoff differences of their pure strategies, then this cumulative empirical data is not constrained (as mixed strategies are). Thus, a promising way to obtain a well-behaved second order dynamical system for learning in games is to use the player's accumulated data to define an unconstrained performance measure for each strategy (this is where the dynamics of the process come in), and then map these "scores" to mixed strategies by means e.g. of a logit choice model (Hofbauer, Sorin, and Viossat, 2009; Mertikopoulos and Moustakas, 2010; Rustichini, 1999; Sorin, 2009). In other words, the dynamics can first be specified on an unconstrained space, and then mapped to the game's strategy space via the players' choice model.

This use of aggregate performance estimates also has important implications from the point of view of evolutionary game theory and population dynamics.

Indeed, it is well-known that the replicator dynamics arise naturally in populations of myopic agents that evolve based on "imitation of success" (Hofbauer, 1995; Sandholm, 2010) or on "imitation driven by dissatisfaction" (Björnerstedt and Weibull, 1996). Revision protocols of this kind are invariably steered by the players' *instantaneous* payoffs; remarkably however, if players are more sophisticated and keep an aggregate (or average) of their payoffs over time, then the same revision rules driven by *long-term* success or dissatisfaction give rise to the same higher order dynamics discussed above.

Paper outline. After a few preliminaries in Section 2, we make this approach precise in Section 3, where we derive a higher order variant of the well-known replicator dynamics of Taylor and Jonker (1978). Regarding the rationality properties of the derived dynamics, we show in Section 4 that the higher order replicator dynamics eliminate strictly dominated strategies, including iteratively dominated ones: in the long run, only iteratively undominated strategies survive. Qualitatively, this result is the same as its first order counterpart; quantitatively however, the rate of extinction increases dramatically with the order of the dynamics: *dominated strategies become extinct in the n-th order replicator dynamics n orders as fast as in first order* (Theorem 4.1).

The reason for this enhanced rate of elimination is that empirical data accrues much faster if a higher order scheme is used rather than a lower order one: players who use a higher order learning rule end up looking deeper into the past, so they identify consistent payoff differences and annihilate dominated strategies much faster. As a consequence of the above, in the higher order ($n \geq 2$) replicator dynamics, *even weakly dominated strategies become extinct* (Theorem 4.3, a result which comes in stark contrast to the first order setting. The higher order replicator dynamics thus perform one round of elimination of weakly dominated strategies followed by the iterated elimination of strictly dominated strategies; from an epistemic point of view, Dekel and Fudenberg (1990) showed that the outcome of this deletion process is all that can be expected from rational players who are not certain of their opponents' payoffs, so our result may be regarded as a dynamic justification of this form of rational behavior.

Extending our analysis to equilibrium play, we show in Section 5 that modulo certain technical modifications, the folk theorem of evolutionary game theory (Hofbauer and Sigmund, 1988; Weibull, 1995) continues to hold in our higher order setting. More specifically, we show that: *a)* if an interior solution orbit converges, then its limit is Nash; *b)* if a point is Lyapunov stable, then it is also Nash; and *c)* if players start close enough to a strict equilibrium and with a small learning bias, then they converge to it; conversely, only strict equilibria have this property (Theorem 5.1). In fact, echoing our results on the rate of extinction of dominated strategies, we show that the *n*-th order replicator dynamics converge to strict equilibria *n* orders as fast as in first order.

Finally, in Section 6, we consider a much wider class of higher order dynamics that extends the familiar *imitative dynamics* of Björnerstedt and Weibull (1996) – including all payoff-monotonic dynamics (Samuelson and Zhang, 1992) and, in particular, the replicator dynamics. The results that we described above go through essentially unchanged for all higher order payoff-monotonic dynamics, with one notable trait standing out: the property that only pure strategy profiles can be attracting holds in *all* higher order imitative dynamics for $n \geq 2$, and not

only for the *n*-th order replicator dynamics. As with the elimination of weakly dominated strategies, this is not the case in first order: for instance, the payoff-adjusted replicator dynamics of Maynard Smith exhibit interior attractors even in simple 2×2 games (see e.g. Ex. 5.3 in Weibull, 1995).

2. Notation and preliminaries.

2.1. **Notational conventions.** If $\mathcal{S} = \{s_\alpha\}_{\alpha=0}^n$ is a finite set, the vector space spanned by $\mathcal{S}$ over $\mathbb{R}$ will be the set $\mathbb{R}^{\mathcal{S}}$ of all maps $x \colon \mathcal{S} \to \mathbb{R}$, $s \in \mathcal{S} \mapsto x_s \in \mathbb{R}$. The canonical basis $\{e_s\}_{s \in \mathcal{S}}$ of this space consists of the indicator functions $e_s \colon \mathcal{S} \to \mathbb{R}$ which take the value $e_s(s) = 1$ on s and vanish otherwise, so, thanks to the identification $s \mapsto e_s$, we will not distinguish between $s \in \mathcal{S}$ and the corresponding basis vector e_s of $\mathbb{R}^{\mathcal{S}}$. In the same spirit, we will use the index α to refer interchangeably to either s_α or e_α (writing e.g. x_α instead of x_{s_α}); likewise, if $\{\mathcal{S}_k\}_{k \in \mathcal{K}}$ is a finite family of finite sets indexed by $k \in \mathcal{K}$, we will write $(\alpha; \alpha_{-k})$ for the tuple $(\alpha_0, \ldots, \alpha_{k-1}, \alpha, \alpha_{k+1}, \ldots) \in \prod_k \mathcal{S}_k$ and $\sum_\alpha^k$ in place of $\sum_{\alpha \in \mathcal{S}_k}$.

We will also identify the set $\Delta(\mathcal{S})$ of probability measures on $\mathcal{S}$ with the *n*-dimensional simplex of $\mathbb{R}^{\mathcal{S}}$: $\Delta(\mathcal{S}) \equiv \{x \in \mathbb{R}^{\mathcal{S}} : \sum_\alpha x_\alpha = 1 \text{ and } x_\alpha \geq 0\}$. Finally, regarding players and their actions, we will follow the original convention of Nash and employ Latin indices $(j, k, \ldots)$ for players, while keeping Greek ones $(\alpha, \beta, \ldots)$ for their actions (pure strategies); also, unless otherwise mentioned, we will use $\alpha, \beta, \ldots$, for indices that start at 0, and $\mu, \nu, \ldots$, for those which start at 1.

2.2. **Finite games.** A *finite game in normal form* will comprise a finite set of *players* $\mathcal{N} = \{1, \ldots, N\}$, each with a finite set of *actions* (or *pure strategies*) $\mathcal{A}_k = \{\alpha_{k,0}, \alpha_{k,1}, \ldots\}$ that can be mixed by means of a probability distribution (*mixed strategy* $x_k \in \Delta(\mathcal{A}_k)$. The set $\Delta(\mathcal{A}_k)$ of a player's mixed strategies will be denoted by X_k, and aggregating over all players, the space of *strategy profiles* $x = (x_1, \ldots, x_N) \in \prod_k \mathbb{R}^{\mathcal{A}_k}$ will be the product $X \equiv \prod_k X_k$; in this way, if $\mathcal{A} = \bigsqcup_k \mathcal{A}_k$ denotes the (disjoint) union of the players' action sets, X may be seen as a product of simplices embedded in $\mathbb{R}^{\mathcal{A}} \cong \prod_k \mathbb{R}^{\mathcal{A}_k}$.

As is customary, when we wish to focus on the strategy of a specific (focal) player $k \in \mathcal{N}$ versus that of his *opponents* $\mathcal{N}_{-k} \equiv \mathcal{N} \setminus \{k\}$, we will use the shorthand $(x_k; x_{-k}) \equiv (x_1, \ldots, x_k, \ldots, x_N) \in X$ to denote the strategy profile where player k plays $x_k \in X_k$ against the strategy $x_{-k} \in X_{-k} \equiv \prod_{\ell \neq k} X_\ell$ of his opponents. The players' (expected) rewards are then prescribed by the game's *payoff* (or *utility* functions $u_k \colon X \to \mathbb{R}$:

$$u_k(x) = \sum_{\alpha_1}^1 \cdots \sum_{\alpha_N}^N u_k(\alpha_1, \ldots, \alpha_N) \, x_{1,\alpha_1} \cdots x_{N,\alpha_N}, \tag{2.1}$$

where $u_k(\alpha_1, \ldots, \alpha_N)$ denotes the reward of player k in the profile $(\alpha_1, \ldots, \alpha_N) \in \prod_k \mathcal{A}_k$; specifically, if player k plays $\alpha \in \mathcal{A}_k$, we will use the notation:

$$u_{k\alpha}(x) \equiv u_k(\alpha; x_{-k}) = u_k(x_1, \ldots, \alpha, \ldots, x_N). \tag{2.2}$$

In light of the above, a game in normal form with players $k \in \mathcal{N}$, action sets $\mathcal{A}_k$ and payoff functions $u_k \colon X \to \mathbb{R}$ will be denoted by $\mathfrak{G} \equiv \mathfrak{G}(\mathcal{N}, \mathcal{A}, u)$. A *restriction* $\mathfrak{G}'$ of $\mathfrak{G}$ (denoted $\mathfrak{G}' \leq \mathfrak{G}$) will then be a game $\mathfrak{G}' \equiv \mathfrak{G}'(\mathcal{N}, \mathcal{A}', u')$ played by the players of $\mathfrak{G}$, each with a subset $\mathcal{A}'_k \subseteq \mathcal{A}_k$ of their original actions, and with payoff functions $u'_k \equiv u_k|_{X'}$ suitably restricted to the reduced strategy space $X' = \prod_k \Delta(\mathcal{A}'_k)$ of $\mathfrak{G}'$.

Given a game $\mathfrak{G} \equiv \mathfrak{G}(\mathcal{N}, \mathcal{A}, u)$, we will say that the pure strategy $\alpha \in \mathcal{A}_k$ is (*strictly*) *dominated* by $\beta \in \mathcal{A}_k$ (and we will write $\alpha \prec \beta$) when

$$u_{k\alpha}(x) < u_{k\beta}(x) \quad \text{for all strategy profiles } x \in X. \tag{2.3}$$

More generally, we will say that $q_k \in X_k$ *is dominated by* $q_k' \in X_k$ if

$$u_k(q_k; x_{-k}) < u_k(q_k'; x_{-k}) \quad \text{for all strategies } x_{-k} \in X_{-k} \text{ of } k\text{'s opponents.} \tag{2.4}$$

Finally, if the above inequalities are only strict for some (but *not all*) $x \in X$, then we will employ the term *weakly dominated* and write $q_k \precsim q_k'$ instead.

Of course, by removing dominated (and, thus, rationally unjustifiable) strategies from a game $\mathfrak{G}$, other strategies might become dominated in the resulting restriction of $\mathfrak{G}$, leading inductively to the notion of *iteratively dominated strategies*: specifically, if a strategy survives all rounds of elimination, then it will be called *iteratively undominated*, and if the space X^∞ of iteratively undominated strategies is a singleton, the game $\mathfrak{G}$ will be called *dominance-solvable*.

On the other hand, when a game cannot be solved by removing dominated strategies, we will turn to the equilibrium concept of Nash which characterizes profiles that are resilient against unilateral deviations; formally, we will say that $q \in X$ is a *Nash equilibrium* of $\mathfrak{G}$ if

$$u_k(x_k; q_{-k}) \leq u_k(q) \quad \text{for all } x_k \in X_k \text{ and for all } k \in \mathcal{N}. \tag{2.5}$$

If (2.5) is strict for all $x_k \in X_k \setminus \{q_k\}$, $k \in \mathcal{N}$, q itself will be called *strict*; finally, equilibria of restrictions $\mathfrak{G}'$ of $\mathfrak{G}$ will be called *restricted equilibria* of $\mathfrak{G}$.

2.3. **Dynamical systems.** Following Lee (2003), a *flow* on X will be a smooth map $\Theta \colon X \times \mathbb{R}_+ \to X$ such that *a)* $\Theta(x, 0) = x$ for all $x \in X$; and *b)* $\Theta(\Theta(x, t), s) = \Theta(x, t + s)$ for all $x \in X$ and for all $s, t \geq 0$. The curve $\Theta^x \colon \mathbb{R}_+ \to X$, $t \mapsto \Theta(x, t)$, will be called the *orbit* (or *trajectory*) of x under Θ, and when there is no danger of confusion, $\Theta^x(t)$ will be denoted more simply by $x(t)$. In this way, Θ induces a vector field V on X via the mapping $x \mapsto V(x) \equiv \dot{x}(0) \in T_x^c X$ where $\dot{x}(0)$ is the initial velocity of $x(t)$ and $T_x^c X$ denotes the tangent cone to X at x, viz.:

$$T_x^c X \equiv \{z \in \mathbb{R}^\mathcal{A} : \sum_\alpha^k z_{k\alpha} = 0 \text{ for all } k \in \mathcal{N} \text{ and } z_{k\alpha} \geq 0 \text{ if } x_{k\alpha} = 0\}. \tag{2.6}$$

By the fundamental theorem on flows, $x(t)$ will be the unique solution to the (first order) *dynamical system* $\dot{x}(t) = V(x(t))$, $t \geq 0$. Accordingly, we will say that $q \in X$ is:

- *stationary* if $V(q) = 0$ (i.e. if $q(t) \equiv \Theta(q, t) = q$ for all $t \geq 0$).
- *Lyapunov stable* if, for every neighborhood U of q, there exists a neighborhood V of q such that $x(t) \in U$ for all $x \in V$, $t \geq 0$.
- *attracting* if $x(t) \to q$ for all x in a neighborhood U of q in X.
- *asymptotically stable* if it is Lyapunov stable and attracting.

Higher order dynamics of the form "$x^{(n)} = V$" are defined via the recursive formulation:

$$\begin{aligned}
\dot{x}(t) &= x^1(t) \\
\dot{x}^1(t) &= x^2(t) \\
&\cdots \\
\dot{x}^{n-1}(t) &= V(x(t), x^1(t), \ldots, x^{n-1}(t)).
\end{aligned} \tag{2.7}$$

An *n*-th order dynamical system on X will thus correspond to a flow on the *phase space* $\Omega = \coprod_x (T_x^c X)^{n-1}$ whose points (*n*-tuples of the form $(x, x^1, \ldots, x^{n-1})$ as above) represent all possible *states* of the system;[1] by contrast, we will keep the designation "points" for base points $x \in X$, and X itself will be called the *configuration space* of the system. Obviously, the evolution of an *n*-th order dynamical system depends on the entire *initial state* $\omega = (x(0), \dot{x}(0), \ldots, x^{(n-1)}(0)) \in \Omega$ and not only on $x(0)$, so stationarity and stability definitions will be phrased in terms of states $\omega \in \Omega$. On the other hand, if we wish to characterize the evolution of an initial *position* $x(0) \in X$ over time, we will do so by means of the corresponding *rest state* $(x(0), 0, \ldots, 0)$ which signifies that the system starts *at rest*: using the natural embedding $x \mapsto (x, 0 \ldots, 0) \in \Omega$, we may thus view X as a subset of Ω, and when there is no danger of confusion, we will identify $x \in X$ with the associated rest state $(x, 0, \ldots, 0) \in \Omega$.

3. Derivation of higher order dynamics.

A fundamental requirement for any class of game dynamics is that solution trajectories must remain in the game's strategy space X for all time. For a first order system of the form

$$\dot{x}_{k\alpha} = F_{k\alpha}(x), \tag{3.1}$$

with Lipschitz $F_{k\alpha} \colon \mathbb{R}^{\mathcal{A}} \to \mathbb{R}$, this is guaranteed by the tangency requirements *a)* $\sum_\alpha^k F_{k\alpha} = 0$ for all $k \in \mathcal{N}$, and *b)* $F_{k\alpha} \geq 0$ whenever $x_{k\alpha} = 0$. In second order however, this does not suffice: if we simply replace $\dot{x}_{k\alpha}$ with $\ddot{x}_{k\alpha}$ in (3.1) and players start with sufficiently high velocity $\dot{x}(0)$ pointing towards the exterior of X, then they will escape X in finite time.

Flåm and Morgan (2004) forced solutions to remain in X by exogenously projecting the velocity $v(t) \equiv \dot{x}(t)$ of an orbit to the tangent cone $T_x^c X$ of "admissible" velocity vectors. This approach however has the problem that projections do not vary continuously with x, so existence and (especially) uniqueness of solutions might fail; moreover, players need to know exactly when they hit a boundary face of X in order to change their projection operator, so machine precision errors are bound to arise (Cantrell, 2000). To circumvent these problems, we will take an approach rooted in reinforcement learning, allowing us to respect the restrictions imposed by the simplicial structure of X in a natural way.

3.1. **The second order replicator dynamics in dyadic games.** We will first describe our higher order reinforcement learning approach in the simpler context of two-strategy games, the main idea being that players keep and update an *unconstrained* measure of their strategies' payoff differences instead of updating their (constrained) strategies directly. In this way, second (or higher) order effects arise naturally when players look two (or more) steps into the past, and it is the dynamics of these "scores" that induce a well-behaved dynamical system on the game's strategy space.

More precisely, consider an *n*-person game where every player $k \in \mathcal{N}$ has two possible actions, "0" and "1", that are appraised based on the associated payoff differences $\Delta u_k \equiv u_{k,1} - u_{k,0}$, $k \in \mathcal{N}$. With this regret-like information at hand,

---

[1]By convention, we let $(T_x^c X)^0 = \{0\}$, so $\Omega = \coprod_x \{0\} \cong X$ for $n = 1$.

players can measure the performance of their strategies over time by updating the auxiliary *score variables* (or *propensities*):

$$U_k(t+h) = U_k(t) + h\Delta u_k(x(t)), \tag{3.2}$$

where h is the time interval between updates and $\Delta u_k(x(t))$ represents the payoff difference between actions "1" and "0" at time t (assumed discrete for the moment). The players' strategies $x_k \in X_k$ are then updated following the *logit* (or *exponential weight*) choice model whereby actions that score higher are played exponentially more often (Hofbauer et al., 2009; Mertikopoulos and Moustakas, 2010; Rustichini, 1999; Sorin, 2009):

$$x_k(t) = \frac{\exp(U_k(t))}{1 + \exp(U_k(t))}. \tag{3.3}$$

This process is repeated indefinitely, so, for simplicity, we will descend to continuous time by letting $h \to 0$ in (3.6).[2] In this way, by collecting terms in the LHS of (3.2) and replacing the discrete difference ratio $(U_k(t+h) - U_k(t))/h$ with $\dot{U}_k$, the system of (3.2) and (3.3) becomes:

$$\dot{U}_k = \Delta u_k(x) \tag{3.4a}$$

$$x_k = (1 + \exp(-U_k))^{-1}. \tag{3.4b}$$

Hence, by differentiating (3.4b) to decouple it from (3.4a), we readily obtain the 2-strategy replicator dynamics of Taylor and Jonker (1978):

$$\dot{x}_k = \frac{dx_k}{dU_k}\dot{U}_k = x_k(1 - x_k)\,\Delta u_k(x). \tag{3.5}$$

In this well-known derivation of the replicator dynamics from the exponential reinforcement rule (3.4) (see also Hofbauer et al., 2009 and Mertikopoulos and Moustakas, 2010), the constraints $x_k \in (0,1)$, $k \in \mathcal{N}$, are automatically satisfied thanks to (3.3). On the downside however, (3.4a) itself "forgets" a lot of past (and potentially useful) information because the "discrete-time" recursion (3.2) only looks one iteration in the past. To remedy this, players could take (3.2) one step further by aggregating the scores U_k themselves so as to gather more momentum towards the strategies that tend to perform better.

This reasoning yields the double-aggregation reinforcement scheme:

$$U_k(t+h) = U_k(t) + h\Delta u_k(x(t)) \tag{3.6a}$$

$$Z_k(t+h) = Z_k(t) + hU_k(t), \tag{3.6b}$$

where, as before, the profile $x(t)$ is updated following the logistic distribution (3.3) applied to the double aggregate Z, viz. $x_k(t) = 1/(1 + \exp(-Z_k(t)))$.[3] Thus, by eliminating the intermediate (first order) aggregation variables U_k from (3.6), we obtain the *second order* recursion:

$$\frac{Z_k(t+2h) - Z_k(t+h)}{h} = \frac{Z_k(t+h) - Z_k(t)}{h} + h\Delta u_k(x(t)), \tag{3.7}$$

---

[2]We should stress that the passage to continuous time is done here at a heuristic level – see Rustichini (1999) and Sorin (2009) for some of the discretization issues that arise. This discretization is a very important topic in itself, but since our focus is the properties of the underlying continuous-time dynamics, we will not address it here.

[3]Of course, players could look even deeper into the past by taking further aggregates in (3.6), but we will not deal with this issue here in order to keep our example as simple as possible.

which in turn leads to the continuous-time variant:

$$\ddot{Z}_k = \Delta u_k(x) \tag{3.8a}$$

$$x_k = (1 + \exp(-Z_k))^{-1}. \tag{3.8b}$$

The second order system (3.8) automatically respects the simplicial structure of X by virtue of the logistic updating rule (3.3), so this overcomes the hurdle of staying in X; still, it is quite instructive to also derive the dynamics of the strategy profile $x(t)$ itself. To that end, (3.8b) gives $Z_k = \log(x_k) - \log(1 - x_k)$, so a simple differentiation yields:

$$\dot{Z}_k = \frac{\dot{x}_k}{x_k} + \frac{\dot{x}_k}{1 - x_k} = \frac{\dot{x}_k}{x_k(1 - x_k)}. \tag{3.9}$$

Differentiating yet again, we thus obtain

$$\ddot{Z}_k = \frac{\ddot{x}_k x_k(1 - x_k) - \dot{x}_k^2(1 - x_k) + \dot{x}_k^2 x_k}{x_k^2(1 - x_k)^2}, \tag{3.10}$$

and some algebra readily yields the *second order replicator dynamics for dyadic games*:

$$\ddot{x}_k = x_k(1 - x_k)\Delta u_k + \frac{1 - 2x_k}{x_k(1 - x_k)}\dot{x}_k^2. \tag{3.11}$$

This derivation of a second order dynamical system on X will be the archetype for the significantly more general class of higher order dynamics of the next section, so we will pause here for some remarks:

Remark 1 (Initial Conditions)*.* In the first order exponential learning scheme (3.4), the players' initial scores $U_k(0)$ determine their initial strategies via the logit rule (3.4b), namely $x_k(0) = (1 + \exp(-U_k(0)))^{-1}$. The situation however is less straightforward in (3.8) where we have two different types of initial conditions: the players' initial scores $Z_k(0)$ and the associated initial velocities $\dot{Z}_k(0)$ (themselves corresponding to the initial values of the intermediate aggregation variables U_k in (3.6)).

In second order, the initial scores $Z_k(0)$ determine the players' initial strategies via (3.8b). On the other hand, the scores' initial velocities $\dot{Z}_k(0) = U_k(0)$ play a somewhat more convoluted role: indeed, differentiating (3.8b) yields

$$\dot{x}_k(0) = \dot{Z}_k(0)\frac{\exp(Z_k(0))}{1 + \exp(Z_k(0))}\frac{1 - \exp(Z_k(0))}{1 + \exp(Z_k(0))} = x_k(0)(1 - x_k(0))\dot{Z}_k(0), \tag{3.12}$$

so the initial score velocities $\dot{Z}_k(0)$ control the initial growth rate $\dot{x}_k(0)$ of the players' strategies. As we shall see in the following, nonzero $\dot{Z}_k(0)$ introduce an inherent bias in a players' learning scheme; hence, given that $\dot{x}_k(0) = 0$ when $\dot{Z}_k(0) = 0$ (independently of the players' initial strategy), starting "at rest" ($\dot{x}_k(0) = 0$) is basically equivalent to learning with no initial bias ($\dot{Z}_k(0) = 0$).

Remark 2 (Past information)*.* The precise sense in which the double aggregation scheme (3.8) is "looking deeper into the past" can be understood more clearly by

writing out explicitly the first and second order scores U_k and Z_k as follows:

$$U_k(t) = U_k(0) + \int_0^t \Delta u_k(x(s)) \, ds, \tag{3.13a}$$

$$Z_k(t) = Z_k(0) + \int_0^t U_k(s) \, ds = Z_k(0) + \dot{Z}_k(0)t + \int_0^t (t-s) \, \Delta u_k(x(s)) \, ds. \tag{3.13b}$$

We thus see that the first order aggregate scores U_k assign uniform weight to all past instances of play, while the second order aggregates Z_k put (linearly) more weight on instances that are farther removed into the past. This mode of weighing can be interpreted as players being reluctant to forget what has occurred, and this is precisely the reason that we describe the second order scheme (3.8) as "looking deeper into the past". From the point of view of learning, this may appear counter-intuitive because past information is usually discounted (e.g. by an exponential factor) and ultimately discarded in favor of more recent observations (Fudenberg and Levine, 1998). As we shall see, "refreshing" observations in this way results in the players' scores U_k growing at most linearly in time (see e.g. Hofbauer et al., 2009, Rustichini, 1999, and Sorin, 2009); on the flip side, if players *reinforce* past observations by using (3.13b) in place of (3.13a), then their scores may grow quadratically instead of linearly.

From (3.13) we also see that nonzero initial score velocities $\dot{Z}_k(0) \neq 0$ introduce a skew in the players' learning scheme: for instance, in a constant game ($\Delta u_k \equiv 0$), the integral expression (3.13b) gives $\lim_{t\to\infty} Z_k(t) = \text{sgn}(Z_k(0)) \cdot \infty$, i.e. $x_k(t)$ will converge to 0 or 1, depending only on the sign of $\dot{Z}_k(0)$. Put differently, the bias introduced by the initial velocities $\dot{Z}_k(0)$ is not static (as the players' choice of initial strategy/score), but instead drives the player to a particular direction, even in the absence of external stimuli.

3.2. **Reinforcement learning and higher order dynamics.** In the general case, we will consider the following reinforcement learning setup:

(1) For every action $\alpha \in \mathcal{A}_k$, player k keeps and updates a *score* (or *propensity*) variable $y_{k\alpha} \in \mathbb{R}$ which measures the performance of α over time.

(2) Players transform the scores $y_k \in \mathbb{R}^{\mathcal{A}_k}$ into mixed strategies $x_k \in X_k$ by means of the Gibbs map $G_k \colon \mathbb{R}^{\mathcal{A}_k} \to X_k$, $y_k \mapsto G_k(y_k)$:

$$x_{k\alpha} = G_{k\alpha}(y_k) \equiv \frac{\exp\left(\lambda_k y_{k\alpha}(t)\right)}{\sum_\beta^k \exp\left(\lambda_k y_{k\beta}(t)\right)}, \tag{GM}$$

where the "inverse temperature" $\lambda_k > 0$ controls the model's sensitivity to external stimuli (Landau and Lifshitz, 1976).

(3) The game is played and players record the payoffs $u_{k\alpha}(x)$ for each $\alpha \in \mathcal{A}_k$.

(4) Players update their scores and the process is repeated ad infinitum.

Needless to say, the focal point of this learning process is the exact way in which players update the performance scores $y_{k\alpha} \in \mathbb{R}$ at each iteration of the game. In the previous section, these scores were essentially defined as double aggregates of the received payoffs via the two-step process (3.6). Here, we will further extend this framework by considering an n-fold aggregation scheme in

which the scores $y_{k\alpha}$ are formed via the following n-step process:

$$
\begin{aligned}
Y_{k\alpha}^{(n-1)}(t+h) &= Y_{k\alpha}^{(n-1)}(t) + h u_{k\alpha}(x(t)), \\
Y_{k\alpha}^{(n-2)}(t+h) &= Y_{k\alpha}^{(n-2)}(t) + h Y_{k\alpha}^{(n-1)}(t) \\
&\quad \cdots \\
Y_{k\alpha}^{(1)}(t+h) &= Y_{k\alpha}^{(1)}(t) + h Y_{k\alpha}^{(2)}(t), \\
y_{k\alpha}(t+h) &= y_{k\alpha}(t) + h Y_{k\alpha}^{(1)}(t).
\end{aligned}
\tag{3.14}
$$

In other words, at each update period, players first aggregate their payoffs by updating the first order aggregation variables $Y_{k\alpha}^{(n-1)}$; they then re-aggregate these intermediate variables by updating the second order aggregation scores $Y_{k\alpha}^{(n-2)}$ above, and repeat this step up to n levels, leading to the n-fold aggregation score $y_{k\alpha}$.

Similarly to the analysis of the previous section, if we eliminate the intermediate aggregation variables $Y^{(1)}$, $Y^{(2)}$ and so forth, we readily obtain the straightforward n-th order recursion:

$$
\frac{\Delta_h^{(n)} y_{k\alpha}(t)}{h^n} = u_{k\alpha}(x(t)),
\tag{3.15}
$$

where $\Delta_h^{(n)} y_{k\alpha}$ denotes the n-th order finite difference of $y_{k\alpha}$, defined inductively as $\Delta_h^{(n)} y(t) = \Delta_h^{(n-1)} y(t+h) - \Delta_h^{(n-1)} y(t)$, with $\Delta_h^{(1)} y(t) = y(t+h) - y(t)$.[4] Thus, if we descend to continuous time by letting $h \to 0$, we obtain the *n-th order learning dynamics*:

$$
y_{k\alpha}^{(n)}(t) = u_{k\alpha}(x(t)),
\tag{LD$_n$}
$$

with $x(t)$ given by the Gibbs map (GM) applied to $y(t)$.

The learning dynamics (LD$_n$) together with the logit choice model (GM) completely specify the evolution of the players' mixed strategy profile $x(t)$ and will thus constitute the core of our considerations. However, it will also be important to derive the associated higher order dynamics induced by (LD$_n$) on the players' strategy space X; to that end, we begin with the identity

$$
\log(x_{k\alpha}) - \log(x_{k\beta}) = \lambda_k (y_{k\alpha} - y_{k\beta}),
\tag{3.16}
$$

itself an easy consequence of (GM). By Faà di Bruno's higher order chain rule (Fraenkel, 1978), we then obtain

$$
\frac{d^n}{dt^n} \log(x_{k\alpha}(t)) = \sum \frac{n!}{m_1! \cdots m_n!} \frac{(-1)^{m-1}(m-1)!}{x_{k\alpha}^m} \prod_{r=1}^n \left(x_{k\alpha}^{(r)}(t)/r! \right)^{m_r},
\tag{3.17}
$$

where $m = m_1 + \cdots + m_n$, and the sum is taken over all non-negative integers $m_1, \ldots, m_n$ such that $\sum_{r=1}^n r m_r = n$. In particular, since the only term that contains $x_{k\alpha}^{(n)}$ has $m_1 = m_2 = \ldots = m_{n-1} = 0$ and $m_n = 1$, we may rewrite (3.17) as

$$
\frac{d^n}{dt^n} \log(x_{k\alpha}(t)) = \frac{x_{k\alpha}^{(n)}(t)}{x_{k\alpha}(t)} + R_{k\alpha}^{n-1}\big(x(t), \dot{x}(t), \ldots, x^{(n-1)}(t)\big),
\tag{3.18}
$$

---

[4]In fact, we have $Y_{k\alpha}^{(r)}(t) = \Delta_h^{(r)} y_{k\alpha}(t)/h^r$ for all $r = 1, \ldots, n-1$, explaining our choice of notation.

where $R_{k\alpha}^{n-1}$ denotes the $(n-1)$-th order remainder of the RHS of (3.17):

$$R_{k\alpha}^{n-1}\left(x, \dot{x}, \ldots, x^{(n-1)}\right) = \sum \frac{(-1)^{m-1} n!}{m_1! \cdots m_{n-1}!} \frac{(m-1)!}{x_{k\alpha}^m} \prod_{r=1}^{n-1}\left(x_{k\alpha}^{(r)}(t)/r!\right)^{m_r}. \quad (3.19)$$

By taking the n-th derivative of (3.16) and substituting, we thus get

$$\lambda_k\left(u_{k\alpha} - u_{k\beta}\right) = \frac{x_{k\alpha}^{(n)}}{x_{k\alpha}} - \frac{x_{k\beta}^{(n)}}{x_{k\beta}} + R_{k\alpha}^{n-1} - R_{k\beta}^{n-1}, \quad (3.20)$$

so, by multiplying both sides with $x_{k\beta}$ and summing over $\beta \in \mathcal{A}_k$ (recall that $\sum_\beta^k x_{k\beta}^{(n)} = 0$), we finally obtain the *n-th order (asymmetric) replicator dynamics*:

$$x_{k\alpha}^{(n)} = \lambda_k x_{k\alpha}\left(u_{k\alpha} - u_k\right) - x_{k\alpha}\left(R_{k\alpha}^{n-1} - \sum_\beta^k x_{k\beta} R_{k\beta}^{n-1}\right). \quad (\text{RD}_n)$$

The higher order replicator equation (RD_n) above will be the chief focus of our paper; as such, a few remarks are in order:

Remark 1. As one would expect, for $n = 1$, we trivially obtain $R_{k\alpha}^0 = 0$ for all $\alpha \in \mathcal{A}_k$, $k \in \mathcal{N}$, so (RD_n) reduces to the standard (asymmetric) replicator dynamics of Taylor and Jonker (1978):

$$\dot{x}_{k\alpha} = \lambda_k x_{k\alpha}\left(u_{k\alpha}(x) - u_k(x)\right). \quad (\text{RD}_1)$$

On the other hand, for $n = 2$, the only lower order term that survives in (3.19) is for $m_1 = 2$; a bit of algebra then yields the second order replicator equation:

$$\ddot{x}_{k\alpha} = \lambda_k x_{k\alpha}\left(u_{k\alpha}(x) - u_k(x)\right) + x_{k\alpha}\left(\dot{x}_{k\alpha}^2/x_{k\alpha}^2 - \sum_\beta^k \dot{x}_{k\beta}^2/x_{k\beta}\right). \quad (\text{RD}_2)$$

At first glance, the above equation seems different from the dynamics (3.11) that we derived in Section 3.1, but this is just a matter of reordering: if we restrict (RD_2) to two strategies, "0" and "1", and set $x_k \equiv x_{k,1} = 1 - x_{k,0}$, we will have $\dot{x}_k = \dot{x}_{k,1} = -\dot{x}_{k,0}$, and (3.11) follows immediately.

Remark 2. In terms of structure, (RD_n) consists of a replicator-like term (driven by the game's payoffs) and a game-independent adjustment $R_{k\alpha}^{n-1}$ which reflects the higher order character of (RD_n). As noted by the associate editor (whom we thank for this remark), if we put the order of the dynamics aside, there is a structural similarity between the higher order replicator dynamics (RD_n), the replicator dynamics with aggregate shocks of Fudenberg and Harris (1992), and the stochastic replicator dynamics of exponential learning (Mertikopoulos and Moustakas, 2010). The reason for this similarity is that all these models are first defined in terms of an auxiliary set of variables: absolute population sizes in Fudenberg and Harris (1992) and payoff scores here and in Mertikopoulos and Moustakas (2010). Differentiation of these variables with respect to time then always yields a replicator-like term carrying the game's payoffs, plus a correction term which is independent of the game being played (because it is coming from Itô calculus or higher order considerations).

Remark 3. Technically, given that the higher order adjustment terms $R_{k\alpha}^{n-1}$ blow up for $x_{k\alpha} = 0$ and $n > 1$, the dynamics (RD_n) are only defined for strategies that lie in the (relative) interior $\text{rel int}(X)$ of X. If the players' initial strategy profile is itself interior, then this poses no problem to (RD_n) because the Gibbs map

(GM) ensures that every strategy share $x_{k\alpha}(t)$ will remain positive for all time. For the most part, we will not need to consider non-interior orbits; nonetheless, if required, we can consider initial conditions on any subface X' of X simply by restricting (GM) to the corresponding restriction $\mathfrak{G}'$ of $\mathfrak{G}$, i.e. by effectively setting the score of an action that is not present in the initial strategy distribution to $-\infty$. In this manner, we may extend (RD$_n$) to any subface of X, and it is in this sense that we will interpret (RD$_n$) for non-interior initial conditions.

Remark 4. In the same way that we derived the integral expressions (3.13) for the payoff scores, we obtain the following integral representation for the higher order learning dynamics (LD$_n$):

$$y_{k\alpha}(t) = \frac{1}{(n-1)!} \int_0^t (t-s)^{n-1} u_{k\alpha}(x(s)) \, ds + \sum_{r=0}^{n-1} y_{k\alpha}^{(r)}(0) \frac{t^r}{r!}. \qquad (3.21)$$

As with our previous discussion in Section 3.1, the players' initial scores $y_{k\alpha}(0)$ determine the players' initial strategies. Similarly, the higher order initial conditions $y_{k\alpha}^{(r)} \neq 0$, $r \geq 1$, control the initial derivates $\dot{x}_{k\alpha}(0),\dots$ of (RD$_n$), and it is easy to see that starting "at rest" is equivalent to having no initial learning bias that could lead players to a particular strategy in the absence of external stimuli (e.g. in a constant game; see also the concluding remarks of Section 3.1).

3.3. **Evolutionary interpretations of the higher order replicator dynamics.** In the mass-action interpretation of evolutionary game theory, it is assumed that there is a nonatomic population linked to each player role $k \in \mathcal{N}$, and that the governing dynamics arise from individual interactions within these populations. In the context of (RD$_n$), such an evolutionary interpretation may be obtained as follows: focusing on the case $n = 2$ for simplicity, assume that each (nonatomic) player receives an opportunity to switch strategies at every ring of a Poisson alarm clock as described in detail in Chapter 3 of Sandholm (2010). In this context, if $\rho_{\alpha\beta}^k$ denotes the *conditional switch rate* from strategy $\alpha \in \mathcal{A}_k$ to strategy $\beta \in \mathcal{A}_k$ in population $k \in \mathcal{N}$ (i.e. the probability of an α-strategist becoming a β-strategist up to a normalization factor), then the strategy shares $x_{k\alpha}$ will follow the *mean dynamics associated to ρ*:

$$\dot{x}_{k\alpha} = \sum_{\beta}^k x_{k\beta} \rho_{\beta\alpha}^k - x_{k\alpha} \sum_{\beta}^k \rho_{\alpha\beta}^k. \qquad (\text{MD}_\rho)$$

Conditional switch rates are usually functions of the current population state $x \in X$ and the corresponding payoffs $u_{k\alpha}$: for instance, the well-known "imitation of success" revision protocol is described by the rule

$$\rho_{\alpha\beta}^k = x_{k\beta} u_{k\beta}(x), \qquad (3.22)$$

and the resulting mean field (MD$_\rho$) is simply the standard replicator equation (Hofbauer, 1995; Sandholm, 2010).[5] On the other hand, if players are more sophisticated and keep track of the *long-term* performance $U_{k\alpha}(t) = \int_0^t u_{k\alpha}(x(s)) \, ds$

---

[5]Other revision protocols that lead to the replicator dynamics are Schlag's (1998)"pairwise proportional imitation" and the protocol of "pure imitation driven by dissatisfaction" of Björnerstedt and Weibull (1996). We will only focus here on "imitation of success" for simplicity; that said, the discussion that follows can easily be adapted to these revision protocols as well.

of a strategy over time (instead of only considering the instantaneous payoffs $u_{k\alpha}$), then the "long-term" analogue of the revision rule (3.22) will be

$$\tilde{\rho}^k_{\alpha\beta} = x_{k\beta} U_{k\beta} = x_{k\beta} \int_0^t u_{k\beta}(x(s)) \, ds, \tag{3.23}$$

leading in turn to the mean dynamics:

$$\dot{x}_{k\alpha} = x_{k\alpha} \left(U_{k\alpha} - \sum_\beta^k x_{k\beta} U_{k\beta} \right). \tag{3.24}$$

Of course, (3.24) is not a dynamical system per se, but a system of integro-differential equations (recall that $U_{k\alpha}$ has an integral dependence on x). However, by differentiating (3.24) with respect to time and recalling that $\dot{U}_{k\alpha} = u_{k\alpha}$, we readily obtain:

$$\ddot{x}_{k\alpha} = \dot{x}_{k\alpha} \left(U_{k\alpha} - \sum_\beta^k x_{k\beta} U_{k\beta} \right) + x_{k\alpha} \left(u_{k\alpha} - \sum_\beta^k x_{k\beta} u_{k\beta} \right) - x_{k\alpha} \sum_\beta^k \dot{x}_{k\beta} U_{k\beta}. \tag{3.25}$$

By (3.24), the first term in the RHS of (3.25) above will be equal to $\dot{x}_{k\alpha}^2 / x_{k\alpha}$; moreover, some easy algebra also yields

$$\begin{aligned}
\sum_\beta^k \dot{x}_{k\beta} U_{k\beta} &= \sum_\beta^k x_{k\beta} \left(U_{k\beta} - \sum_\gamma^k x_{k\gamma} U_{k\gamma} \right) U_{k\beta} \\
&= \sum_\beta^k x_{k\beta} U_{k\beta}^2 - \left(\sum_\beta^k x_{k\beta} U_{k\beta} \right)^2 \\
&= \sum_\beta^k x_{k\beta} \left(U_{k\beta} - \sum_\gamma^k x_{k\gamma} U_{k\gamma} \right)^2 = \sum_\beta \dot{x}_{k\beta}^2 / x_{k\beta}.
\end{aligned} \tag{3.26}$$

Thus, after some rearranging, (3.25) becomes

$$\ddot{x}_{k\alpha} = x_{k\alpha} \left(u_{k\alpha}(x) - u_k(x) \right) + x_{k\alpha} \left(\dot{x}_{k\alpha}^2 / x_{k\alpha}^2 - \sum_\beta^k \dot{x}_{k\beta}^2 / x_{k\beta} \right), \tag{3.27}$$

i.e. the mean dynamics associated to the "imitation of long-term success" revision protocol (3.23) is just the second order replicator equation (RD$_2$) with $\lambda_k = 1$.

The higher order dynamics (RD$_n$) may be derived from similar considerations, simply by taking a revision protocol of the form (3.23) with U replaced by a different (higher order) payoff aggregation scheme. Accordingly, the evolutionary significance of higher order is similar to its learning interpretation: higher order dynamics arise when players revise their strategies based on long-term performance estimates instead of instantaneous payoff information. Obviously, this opens the door to higher order variants of other population dynamics that arise from revision protocols (such as the Smith dynamics and other pairwise comparison dynamics), but since this discussion would take us too far afield, we will delegate it to a future paper.

Remark. The definition of the payoff aggregates $U_{k\beta}(t) = \int_0^t u_{k\beta}(x(s)) \, ds$ gives $\dot{x}_{k\beta}(0) = 0$ in (3.24), so players will be starting "at rest" in (3.25). On the other hand, just as in (3.13a), if players are inherently predisposed towards one strategy or another, these aggregates could be offset by some nonzero initial bias $U_{k\beta}(0)$, which would then translate into a nonzero initial velocity $\dot{x}(0)$. In view of the above, when it is safe to assume that players are not ex-ante biased, starting at rest will be our baseline assumption.

4. Elimination of dominated strategies.

A fundamental rationality requirement for any class of game dynamics is that dominated strategies die out in the long run. Formally, if play evolves over time, say along the path $x(t)$, $t \geq 0$, we will say that the strategy $\alpha \in \mathcal{A}_k$ *becomes extinct along $x(t)$* if $x_{k\alpha}(t) \to 0$ as $t \to \infty$; more generally, for mixed strategies $q_k \in X_k$, we will follow Samuelson and Zhang (1992) and say that q_k becomes extinct along $x(t)$ if $\min\{x_{k\alpha}(t) : \alpha \in \mathrm{supp}(q_k)\} \to 0$, with the minimum taken over the support $\mathrm{supp}(q_k) \equiv \{\beta \in \mathcal{A}_k : q_{k\beta} > 0\}$ of q_k. Equivalently, if we let

$$D_{\mathrm{KL}}(q_k \,\|\, x_k) = \sum_{\alpha} q_{k\alpha} \log\left(q_{k\alpha}/x_{k\alpha}\right) \tag{4.1}$$

denote the *Kullback-Leibler divergence* of x_k with respect to q_k (with the usual convention $0 \cdot \log 0 = 0$ when $q_{k\alpha} = 0$), then $D_{\mathrm{KL}}(q_k \,\|\, x_k)$ blows up to $+\infty$ whenever $\min\{x_{k\alpha} : \alpha \in \mathrm{supp}(q_k)\} \to 0$, so $q_k \in X_k$ becomes extinct along $x(t)$ if and only if $D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) \to \infty$ as $t \to \infty$ (see e.g. Weibull 1995).

In light of the above, our first result is to show that in the n-th order replicator dynamics, dominated strategies die out at a rate which is exponential in t^n:

Theorem 4.1. *Let $x(t)$ be an interior solution orbit of the n-th order replicator dynamics* (RD$_n$). *If $q_k \in X_k$ is iteratively dominated, we will have*

$$D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) \geq \lambda_k c t^n/n! + \mathcal{O}(t^{n-1}), \tag{4.2}$$

for some constant $c > 0$. In particular, for pure strategies $\alpha \prec \beta$, we will have

$$x_{k\alpha}(t)/x_{k\beta}(t) \leq \exp\left(-\lambda_k \Delta u_{\beta\alpha} t^n/n! + \mathcal{O}(t^{n-1})\right), \tag{4.3}$$

where $\Delta u_{\beta\alpha} = \min_{x \in X}\{u_{k\beta}(x) - u_{k\alpha}(x)\} > 0$.

As an immediate corollary, we then obtain:

Corollary 4.2. *In dominance-solvable games, the n-th order replicator dynamics* (RD$_n$) *converge to the game's rational solution.*

Remark. Before proving Theorem 4.1, it is worth nothing that even though (4.2) and (4.3) have been stated as inequalities, one can use any upper bound for the game's payoffs to show that the rate of extinction of dominated strategies in terms of the K-L divergence is indeed $\mathcal{O}(t^n)$.[6] As a result, the asymptotic rate of extinction of dominated strategies in the n-th order replicator dynamics (RD$_n$) is n orders as fast as in the standard first order dynamics (RD$_1$), so irrational play becomes extinct much faster in higher orders.

Proof of Theorem 4.1. We will begin by showing that if $q_k \in X_k$ is dominated by $q_k' \in X_k$, then $D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) \geq c t^n/n!$ for some positive constant $c > 0$. Indeed, let $V_k(x) = D_{\mathrm{KL}}(q_k \,\|\, x_k) - D_{\mathrm{KL}}(q_k' \,\|\, x_k)$, and rewrite (GM) as $\log x_{k\alpha} = \lambda_k y_{k\alpha} - \log(\mathcal{Z}_k(y))$ where $\mathcal{Z}_k(y) = \sum_{\beta}^k \exp(\lambda_k y_{k\beta})$ denotes the partition function of player

---

[6]In fact, the coefficients that make (4.2) and (4.3) into asymptotic equalities can also be determined, but we will not bother with this calculation here.
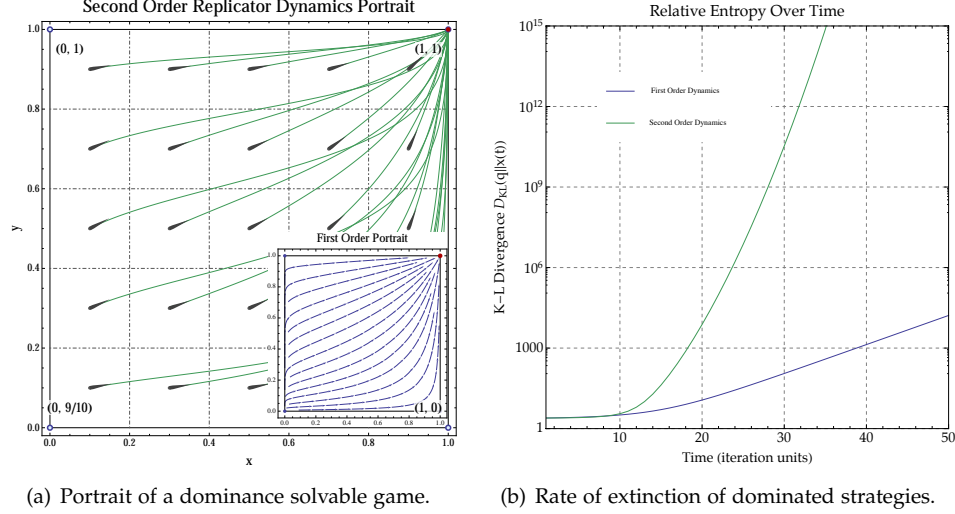
(a) Portrait of a dominance solvable game.   (b) Rate of extinction of dominated strategies.

FIGURE 1.   Extinction of dominated strategies in the first and second order replicator dynamics. In Fig. 1(a) we plot the second order solution orbits of a dominance solvable game with payoff matrices $U_1 = ((1,1),(0,0))$ for the "x" player and $U_2 = ((1,0),(1,9/10))$ for the "y" player (see also the figure's labels). In Fig. 1(b), we illustrate the rate of extinction of the dominated strategy of player 1 by plotting the corresponding K-L divergence of a typical trajectory: the K-L distance grows exp-quadratically in second order dynamics compared to exp-linearly in first order.

k. Then, some algebra yields:

$$V_k(x) = \sum_{\alpha \in \text{supp}(q)} q_{k\alpha} \log (q_{k\alpha}/x_{k\alpha}) \ - \sum_{\alpha \in \text{supp}(q')} q'_{k\alpha} \log (q'_{k\alpha}/x_{k\alpha})$$
$$= \sum_\alpha^k (q'_{k\alpha} - q_{k\alpha}) \log x_{k\alpha} + h_k(q_k, q'_k)$$
$$= \sum_\alpha^k (q'_{k\alpha} - q_{k\alpha}) \lambda_k y_{k\alpha} + h_k(q_k, q'_k), \tag{4.4}$$

where $h_k(q_k, q'_k)$ is a constant depending only on q_k and q'_k, and the last equality follows from the fact that $\sum_\alpha^k (q'_{k\alpha} - q_{k\alpha}) \log \mathcal{Z}_k = 0$ (recall that $\sum_\alpha^k q_{k\alpha} = \sum_\alpha^k q'_{k\alpha} = 1$). In this way, we obtain:

$$\frac{d^n}{dt^n} V_k(x(t)) = \lambda_k \sum_\alpha^k (q'_{k\alpha} - q_{k\alpha}) y_{k\alpha}^{(n)} = \lambda_k \sum_\alpha^k (q'_{k\alpha} - q_{k\alpha}) u_{k\alpha}(x(t))$$
$$= \lambda_k \left[u_k(q'_k; x_{-k}(t)) - u_k(q_k; x_{-k}(t)) \right] \geq \lambda_k \Delta u_k > 0, \tag{4.5}$$

where the constant Δu_k is defined as $\Delta u_k = \min_{X_{-k}} \{ u_k(q'_k; x_{-k}) - u_k(q_k; x_{-k}) \}$ and its positivity follows from the fact that X is compact and u_k is continuous. Hence, if we set $c_r = (r!)^{-1} \frac{d^r V_k}{dt^r}\Big|_{t=0}$, $r = 0 \ldots n-1$, Taylor's theorem with Lagrange remainder readily gives:

$$V_k(x(t)) \geq \lambda_k \Delta u_k t^n / n! + \sum_{r=0}^{n-1} c_r t^r, \tag{4.6}$$

and our assertion follows by noting that $D_{\text{KL}}(q_k \| x_k(t)) \geq V_k(x(t))$. In particular, for pure strategies $\alpha \prec \beta$, we will have $V_k(x(t)) = \log x_{k\beta}(t) - \log x_{k\alpha}(t)$, so (4.6)

gives:

$$\log\left(x_{k\beta}(t)/x_{k\alpha}(t)\right) \geq \lambda_k \, \Delta u_{\beta\alpha} t^n/n! + \mathcal{O}(t^{n-1}), \tag{4.7}$$

and (4.3) follows by exponentiating.

Now, to establish the theorem for *iteratively* dominated strategies, we will resort to induction on the rounds of elimination. To that end, let X_k^r denote the space of strategies that survive r elimination rounds, and assume that $D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) = \mathcal{O}(t^n)$ for all strategies $q_k \notin X_k^r$; in particular, if $\alpha \notin \mathcal{A}_k^r \equiv \mathcal{A}_k \cap X_k^r$, this implies that $x_{k\alpha}(t) \to 0$ as $t \to \infty$. We will show that this also holds if $q_k \in X_k^r$ survives for r deletion rounds but dies in the subsequent one. Indeed, if $q_k \in X_k^r \setminus X_k^{r+1}$, there will be some $q_k' \in X_k^r$ with $u_k(q_k'; x_{-k}) > u_k(q_k; x_{-k})$ for all $x_{-k} \in X_{-k}^r$. With this in mind, decompose $x \in X$ as $x = x^r + z^r$ where x^r denotes the "r-rationalizable" part of x, i.e. the orthogonal projection of x on the subspace of X spanned by the surviving pure strategies $\mathcal{A}_\ell^r$, $\ell \in \mathcal{N}$. Then, if we set $\Delta u_k^r = \min\{u_k(q_k'; \alpha_{-k}) - u_k(q_k; \alpha_{-k}) : \alpha_{-k} \in \mathcal{A}_{-k}^r\}$, we will also have:

$$u_k(q_k'; x_{-k}^r) - u_k(q_k; x_{-k}^r) \geq \Delta u_k^r > 0 \quad \text{for all } x_{-k} \in X_{-k}. \tag{4.8}$$

Moreover, it is easy to see that our induction hypothesis implies $z^r(t) \to 0$ as $t \to \infty$ (recall that $x_{k\alpha}(t) \to 0$ for all $\alpha \notin \mathcal{A}_k^r$), so, for large enough t, we also get:

$$|u_k(q_k'; z_{-k}^r(t)) - u_k(q_k; z_{-k}^r(t))| < \Delta u_k^r/2. \tag{4.9}$$

Hence, by combining (4.8) and (4.9), we obtain $u_k(q_k'; x_{-k}(t)) - u_k(q_k; x_{-k}(t)) > \Delta u_k^r/2$ for large t, and the induction is complete by plugging this last estimate into (4.5) and proceeding as in the base case $r = 0$ (our earlier assertion). $\quad\square$

On the other hand, if a strategy is only *weakly* dominated, the payoff differences $\Delta u_{\beta\alpha}$ in (4.3) and related estimates vanish, so Theorem 4.1 cannot guarantee that it will be annihilated. In fact, it is well-known that weakly dominated strategies may survive in the standard first order replicator dynamics: if the pure strategy $\alpha \in \mathcal{A}_k$ of player k is weakly dominated by $\beta \in \mathcal{A}_k$, and if all adversarial strategies $\alpha_{-k} \in \mathcal{A}_{-k}$ against which β performs better than α die out, then α may survive for an open set of initial conditions (for instance, see Example 5.4 and Proposition 5.8 in Weibull, 1995).

Quite remarkably, this can never be the case in a higher order setting if players start unbiased:

Theorem 4.3. *Let $x(t)$ be an interior solution orbit of the n-th order ($n \geq 2$) replicator dynamics (RD_n) that starts at rest: $\dot{x}(0) = \ldots = x^{(n-1)}(0) = 0$. If $q_k \in X_k$ is weakly dominated, then it becomes extinct along $x(t)$ with rate*

$$D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) \geq \lambda_k c t^{n-1}/(n-1)!, \tag{4.10}$$

where λ_k is the learning rate of player k and $c > 0$ is a positive constant.

The intuition behind this surprising result can be gleaned by looking at the reinforcement learning scheme (LD_n). If we take the case $n = 2$ for simplicity, we see that the "payoff forces" $F_{k\alpha} \equiv u_{k\alpha}$ never point towards a weakly dominated strategy. As a result, solution trajectories are always accelerated away from weakly dominated strategies, and even if this acceleration vanishes in the long run, the trajectory still retains a growth rate that drives it away from the dominated strategy. By comparison, this is not the case in first order dynamics: there, we only know that *growth rates* point away from weakly dominated strategies,

and if these rates vanish in the long run, solution trajectories might ultimately converge to a point where weakly dominated strategies are still present (see for instance Fig. 2. The proof follows by formalizing these ideas:

Proof of Theorem 4.3. Let $q_k \preccurlyeq q'_k$ and let $\mathcal{A}'_{-k} \equiv \{\alpha_{-k} \in \mathcal{A}_{-k} : u_k(q'_k; \alpha_{-k}) > u_k(q_k; \alpha_{-k})\}$ be the set of pure strategy profiles of k's opponents against which q'_k yields a strictly greater payoff than q_k. Then, with notation as in the Proof of Theorem 4.1, we will have:

$$\frac{d^n}{dt^n} V_k(x(t)) = \lambda_k \sum_{\alpha_{-k} \in \mathcal{A}'_{-k}} \left[u_k(q'_k; \alpha_{-k}) - u_k(q_k; \alpha_{-k})\right] x_{\alpha_{-k}}(t), \qquad (4.11)$$

where $x_{\alpha_{-k}} \equiv \prod_{\ell \neq k} x_{\alpha_\ell}$ denotes the α_{-k}-th component of x. Thus, with $x(t)$ starting at rest, Faà di Bruno's formula gives $\frac{d^r V_k}{dt^r}\Big|_{t=0} = 0$ for all $r = 1, \dots, n-1$, and a simple integration then yields:

$$\frac{d^{n-1}}{dt^{n-1}} V_k(x(t)) = \lambda_k \sum_{\alpha_{-k} \in \mathcal{A}'_{-k}} \left[u_k(q'_k; \alpha_{-k}) - u_k(q_k; \alpha_{-k})\right] \int_0^t x_{\alpha_{-k}}(s)\, ds, \quad (4.12)$$

However, with $x(t)$ interior, the integrals in the above equation will be positive and increasing, so for some suitably chosen $c > 0$ and t large enough, we obtain

$$\frac{d^{n-1}}{dt^{n-1}} V_k(x(t)) \geq \lambda_k c > 0, \qquad (4.13)$$

and our claim follows from a $(n-1)$-fold integration. $\qquad\square$

In view of this qualitative difference between first and higher order dynamics, some further remarks are in order:

Remark 1. In the first order replicator dynamics, the elimination of weakly dominated strategies when evidence of their domination survives requires that all players adhere to the same dynamics (see e.g. the proof of Proposition 3.2 in Weibull, 1995). To wit, consider a simple Entry Deterrence game where a competitor (Player 1) "enters" or "stays out" of a market controlled by a monopolist (Player 2) who can either "fight" the entrant or "share" the market, and where "fighting" is a weakly dominated strategy that yields a strictly worse payoff if the competitor "enters" (Weibull, 1995, Ex. 5.4). Under the replicator dynamics, "fight" becomes extinct if "enter" survives (cf. Figure 2); however, if Player 1 were to follow a different process under which "enter" survives but the integral of its population share over time is bounded, then "fight" does not become extinct (cf. the proof of Proposition 3.2 in Weibull, 1995). In higher orders though, the proof of Theorem 4.3 goes through for any continuous play $x_{-k}(t) \in X_{-k}$, $t \geq 0$, of k's opponents, so weakly dominated strategies become extinct independently of how one's opponents evolve over time.

Remark 2. As noted in Section 3, starting "at rest" is a natural assumption to make from both learning and evolutionary considerations. First, as far as learning is concerned, this assumption means that players may start with any mixed strategy they wish, but that the learning process (LD$_n$) is not otherwise skewed towards one strategy or another; similarly, with regards to evolution, starting with $\dot{x}(0) = \dots = 0$ is just the baseline of the "imitation of long-term success" revision protocol (3.23).
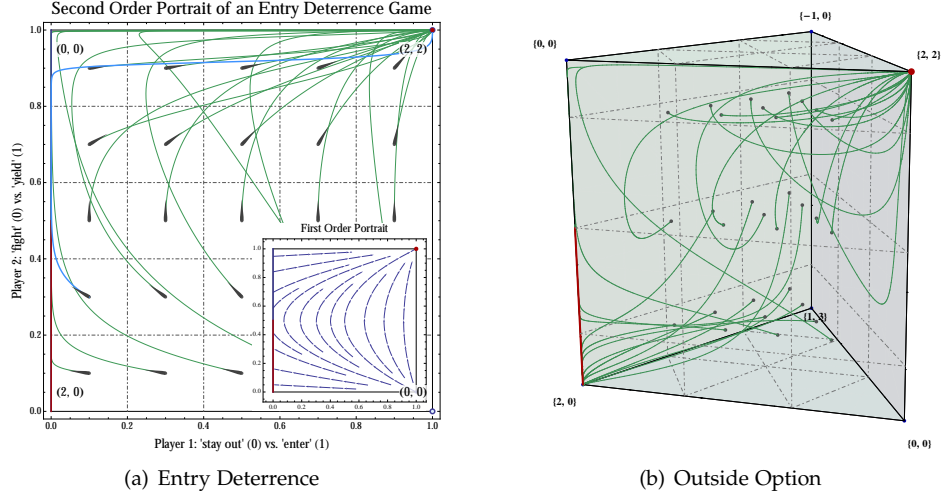
(a) Entry Deterrence   (b) Outside Option

FIGURE 2.   Extinction of weakly dominated strategies and survival of itera-
tively weakly dominated ones in the second order replicator dynamics. Fig. 2(a)
shows solution orbits starting at rest in an Entry Deterrence game: the weakly
dominated strategy "fight" of Player 2 becomes extinct, in stark contrast to the
first order case (compare the highlighted trajectory with the first order portrait
in the inlay). Fig. 2(b) shows an Outside Option supergame where the strat-
egy "fight" in Fig. 2(a) is only *iteratively* weakly dominated; this strategy pays
very well against certain initial conditions, so it ends up surviving when all evi-
dence that it is (iteratively weakly) dominated vanishes. (The payoff matrices for
the Outside Option supergame are $U_1 = ((2,0),(0,2),(-1,1))$ for Player 1 and
$U_2 = ((2,0),(0,0),(0,3))$ for Player 2; see also the corresponding figure labels.)

That said, Theorem 4.3 still holds if the players' initial velocities (or higher
order derivates) are nonzero but small; if they are too large, weakly dominated
strategies may indeed survive.[7] This observation is important for strategies which
are only *iteratively* weakly dominated because, if a strategy becomes weakly dom-
inated after removing a strictly dominated strategy, then the system's solutions
could approach the face of X associated with the resulting restriction of the game
with a high velocity towards the newly weakly dominated strategy (e.g. if the
iteratively weakly dominated strategy pays very well against the disappearing
strictly dominated one; cf. Fig. 2). Thus, although Theorem 4.3 guarantees the
elimination of weakly dominated strategies, its conclusions do not extend to iter-
atively weakly dominated ones.

Remark 3. A joint application of Theorems 4.3 and 4.1 reveals that the higher
order replicator dynamics (RD_n) perform one round of elimination of weakly
dominated strategies followed by the elimination of all strictly dominated strate-
gies. This result may thus be seen as a dynamic justification of the claim of Dekel
and Fudenberg (1990) who argue that asking for the iterated deletion of weakly
dominated strategies is too strong a requirement for "rational" play.

---

[7]More precisely, it suffices for the RHS of (4.12) to exceed $V_k^{(r)}(0)$ for some $t > 0$.

In particular, Dekel and Fudenberg posit that if players are not certain about their opponents' payoffs, then they will not choose a weakly dominated strategy; however, to proceed with a second round of elimination, players must know that other players will not choose certain strategies, and since weak dominance is destroyed by arbitrarily small amounts of payoff uncertainty, only strictly dominated strategies may henceforth be deleted. In the same spirit, weakly dominated strategies are eliminated in the higher order replicator dynamics when players begin unbiased; however, because of the inertial character of (LD_n), players may develop such a bias over time, so only (iteratively) strictly dominated strategies are sure to become extinct after that phase.

Remark 4. In a very recent paper, Balkenborg, Hofbauer, and Kuzmics (2013) observe a similar behavior in a refined variant of the best reply dynamics of Gilboa and Matsui (1991) where the players' best reply correspondences are modified to include only strategies that are best replies to an open set of nearby states of play. This refined update process also eliminates weakly dominated strategies, but it requires players to be significantly more informed than in the myopic context of continuous-time deterministic dynamics: it applies to highly informed, highly rational players who know not only their payoffs at the current state of play, but also their payoffs in all nearby states as well. Instead, Theorem 4.3 shows that weakly dominated strategies (and weakly dominated equilibria) become extinct under much milder information assumptions, namely the players' payoffs at the current state.

Remark 5. Tying in with Remark 2 above, we get the following result for weakly dominated Nash equilibria (or for Nash equilibria whose support contains a weakly dominated strategy): if q is such an equilibrium and players start with sufficiently small learning bias $\dot{y}(0), \ddot{y}(0)$, etc., then $D_{\mathrm{KL}}(q \,\|\, x(t)) \to +\infty$. In particular, there exists a neighborhood V of q in X such that every solution orbit $x(t)$ of (RD_n) which starts at rest in V will escape V in finite time, never to return. In this sense, weakly dominated equilibria are *repelling*, so they may not be selected in the higher order replicator dynamics (RD_n) if players start at rest (see also Theorem 5.1 in the following section).

Remark 6. Finally, it is important to note that our estimate of the rate of extinction of weakly dominated strategies is one order lower than that of strictly dominated ones; as a result, Theorem 4.3 does not imply the annihilation of weakly dominated strategies in first order dynamics (as well it shouldn't). Instead, in first order, if there is some adversarial strategy against which the weakly dominant strategy gives a strictly greater payoff than the weakly dominated one, and if the share of this strategy always remains above a certain level, then the weakly dominated strategy becomes extinct (see e.g. Proposition 3.2 in Weibull, 1995). In our higher order setting, this assumption instead implies that weakly dominated strategies become extinct as fast as *strictly* dominated ones:

Proposition 4.4. *Let $x(t)$ be an interior solution of the n-th order replicator dynamics* (RD_n)*, and let $q_k \preccurlyeq q_k'$. If there exists $\alpha_{-k} \in \mathcal{A}_{-k}$ with $u_k(q_k; \alpha_{-k}) < u_k(q_k'; \alpha_{-k})$ and $x_{\alpha_{-k}}(t) \geq \varepsilon > 0$ for all $t \geq 0$, then:*

$$D_{\mathrm{KL}}(q_k \,\|\, x_k(t)) \geq \varepsilon \lambda_k \big[u_k(q_k'; \alpha_{-k}) - u_k(q_k; \alpha_{-k}) \big] t^n / n! + \mathcal{O}(t^{n-1}). \qquad (4.14)$$

Proof. Simply note that the estimate (4.11) is bounded from below by $\varepsilon \lambda_k u_k(q_k' - q_k; \alpha_{-k})$ and follow the same reasoning as in the proof of Theorem 4.1. $\square$

5. Stability of Nash play and the folk theorem.

In games that cannot be solved by the successive elimination of dominated strategies, one usually tries to identify the game's Nash equilibria instead. Thus, given the prohibitive complexity of these solutions (Daskalakis, Goldberg, and Papadimitriou, 2006), one of the driving questions of evolutionary game theory has been to explain how Nash play might emerge over time as the byproduct of a simpler, adaptive dynamic process.

5.1. **The higher order folk theorem.** A key result along these lines is the *folk theorem of evolutionary game theory* (Hofbauer and Sigmund, 1988; Sandholm, 2010; Weibull, 1995); for the multi-population replicator dynamics (RD$_1$), this theorem can be summarized as follows:

   I. Nash equilibria are stationary.
  II. If an interior solution orbit converges, its limit is Nash.
 III. If a point is Lyapunov stable, then it is also Nash.
 IV. A point is asymptotically stable if and only if it is a strict equilibrium.

Accordingly, our aim in this section will be to extend the above in the context of the higher order dynamics (RD$_n$). To that end however, it is important to recall that the higher order playing field is fundamentally different because the choice of an initial strategy profile $x(0) \in X$ does not suffice to determine the evolution of (RD$_n$); instead, one must prescribe the full initial state $\omega(0) = (x(0), \dot{x}(0), \dots, x^{(n-1)}(0))$ in the system's phase space Ω. Regardless, a natural way to discuss the stability of initial *points* $q \in X$ is via the corresponding *rest states* $(q, 0, \dots, 0) \in \Omega$ (recall also the relevant discussion in Section 2.3, Section 3, and the remarks following Theorem 4.3). With this in mind, we will say that $q \in X$ is *stationary* (resp. *Lyapunov stable*, resp. *attracting*) when the associated rest state $(q, 0, \dots, 0) \in \Omega$ is itself stationary (resp. Lyapunov stable, resp. attracting).

In spite of these differences, we have:

Theorem 5.1. *Let $x(t)$ be a solution orbit of the n-th order replicator dynamics (RD$_n$), $n \geq 1$, for a normal form game $\mathfrak{G} \equiv \mathfrak{G}(\mathcal{N}, \mathcal{A}, u)$, and let $q \in X$. Then:*

   I. *$x(t) = q$ for all $t \geq 0$ iff q is a restricted equilibrium of $\mathfrak{G}$ (i.e. $u_{k\alpha}(q) = \max\{u_{k\beta}(q) : q_{k\beta} > 0\}$ whenever $q_{k\alpha} > 0$).*

  II. *If $x(0) \in \text{int}(X)$ and $\lim_{t \to \infty} x(t) = q$, then q is a Nash equilibrium of $\mathfrak{G}$.*

 III. *If every neighborhood U of q in X admits an interior orbit $x_U(t)$ such that $x_U(t) \in U$ for all $t \geq 0$, then q is a Nash equilibrium of $\mathfrak{G}$.*

 IV. *Let q be a strict equilibrium of $\mathfrak{G}$. For every neighborhood U of q in X, there exists a neighborhood V of q in X and an open set $W \subseteq \Omega$ containing $V \setminus \{(q, 0, \dots, 0)\}$ such that $x(t) \in U$ and $x(t) \to q$ for all initial states $(x(0), \dot{x}(0), \dots) \in W$; put differently, for every $x \in X$ sufficiently close to q, we can find a neighborhood V_x of $(x, 0, \dots, 0)$ in Ω such that $x(t)$ converges to q for all $(x(0), \dot{x}(0), \dots) \in V_x$, but the bound on $\dot{x}(0)$, $\ddot{x}(0)$, etc. will depend on $x(0)$. Conversely, only strict equilibria have this property.*

As an immediate corollary of (IV), we also have:

IV$'$. *If q is a strict equilibrium of $\mathfrak{G}$, then it is attracting: there exists a neighborhood U of q in X such that $x(t) \to q$ whenever $x(t)$ starts at rest in U (that is, $x(0) \in U$ and $\dot{x}(0) = \ldots = 0$); conversely, only strict equilibria have this property.*

The basic intuition behind Theorem 5.1 is as follows: First, stationarity is a trivial consequence of the replicator-like term of the dynamics (RD$_n$). Parts II and III follow by noting that if a trajectory eventually spends all time near a stationary point q, then this point must be Nash – otherwise, the forces of (LD$_n$) would drive the orbit away. Finally, convergence to strict equilibria is a consequence of the fact that they are locally dominant, so the payoff-driven forces (LD$_n$) point in their direction. However, before making these ideas precise, it will be important to draw the following parallels between Theorem 5.1 and the standard first order folk theorem:

Parts I and II of Theorem 5.1 are direct analogues of the corresponding first order claims; note however that (II) can now be inferred from (III).

Part III is a slightly stronger assertion than the standard statement that Lyapunov stability implies Nash equilibrium in first order: indeed, Lyapunov stability posits that all trajectories which start close enough will remain nearby, whereas Theorem 5.1 only asks for one such trajectory. Actually, this last property is all that is required for the proof of the corresponding part of the first order folk theorem as well, but since there are cases which satisfy the latter property but not the former,[8] we will use this stronger formulation which seems closer to a "bare minimum" characterization of Nash equilibria (especially in higher orders).

Part IV shows that strict equilibria attract all nearby rest states $(x(0), 0, \ldots, 0) \in \Omega$, but it is not otherwise tantamount to higher order asymptotic stability – it would be if W were a neighborhood of V in Ω (or, equivalently, of $(q, 0, \ldots, 0)$ in Ω) instead of $V \setminus \{(q, 0, \ldots, 0)\}$.[9]

This difference between first and higher orders is intimately tied to the bias that higher order initial conditions (such as $\dot{y}(0)$, $\ddot{y}(0)$, etc.) introduce in the learning scheme (LD$_n$). More precisely, recall that a nonzero initial score velocity $\dot{y}(0)$ skews the learning scheme (LD$_n$) to such an extent that it might end up converging to an arbitrary pure strategy even in the absence of external stimuli (viz. in a constant game; cf. the relevant discussion at the end of Section 3.1). This behavior is highly unreasonable and erratic, so players are more likely to adhere to an unbiased version of (LD$_n$) with $\dot{y}(0) = \ddot{y}(0) = \ldots = 0$. In that case however, Faà di Bruno's chain rule shows that we will also have $\dot{x}(0) = \ddot{x}(0) = \ldots = 0$ in (RD$_n$), so Part IV$'$ of Theorem 5.1 allows us to recover the first order statement to the effect that all nearby initial strategy choices are attracted to q. Similarly, if we consider the mass-action interpretation of (RD$_n$) that we put forth in Section 3.3, players who are unbiased in the calculation of their aggregate payoffs will also have $\dot{x}(0) = \cdots = 0$, so Part IV$'$ of the theorem essentially boils down to the first order asymptotic stability result in that case.

---

[8] For instance, the equilibrium profile $q_1 = (1, 0)$, $q_2 = (1/2, 1/2)$ of the simple 2×2 game with payoff matrices $U_1 = \mathbf{I}$ and $U_2 = \mathbf{0}$ is neither Lyapunov stable under the replicator dynamics, nor an ω-limit of an interior trajectory, but it still satisfies the property asserted in Part III of Theorem 5.1.

[9] We thank Josef Hofbauer for this remark.

That said, it is also important to note that this convergence statement remains true even if the players' higher order learning bias $\dot{y}(0), \ddot{y}(0), \dots$, is nonzero but (uniformly) not too large. To wit, assume without loss of generality that the strict equilibrium q under scrutiny corresponds to everyone playing their "0"-th strategy, and consider the associated relative score variables

$$z_{k\mu} = y_{k\mu} - y_{k,0}, \quad \mu \in \mathcal{A}_k^* \equiv \mathcal{A}_k \setminus \{0\}. \tag{5.1}$$

As can be easily seen, these score differences are mapped to strategies $x \in X$ via the *reduced Gibbs map* $G_k^* : \mathbb{R}^{\mathcal{A}_k^*} \to X_k$, with $z_k \in \mathbb{R}^{\mathcal{A}_k^*} \mapsto G_k^*(z_k) \in X_k$ as follows:

$$G_{k,0}^*(z) = \left(1 + \sum_\nu^k e^{z_{k\nu}}\right)^{-1}, \qquad G_{k\mu}^*(z) = e^{z_{k\mu}}\left(1 + \sum_\nu^k e^{z_{k\nu}}\right)^{-1}. \tag{GM*}$$

More specifically, if the relative scores $z_{k\mu}$ are given by (5.1), we will have $G_k^*(z_k) = G_k(y_k)$, so the learning scheme (LD$_n$) will be equivalent to the *relative score dynamics*

$$z_{k\mu}^{(n)} = u_{k\mu}(x) - u_{k,0}(x), \tag{ZD$_n$}$$

with (reduced) logit choice $x_k = G_k^*(z)$. In this formulation, the proof of Theorem 5.1 shows that if players start with sufficiently negative $z_{k\mu}(0)$ and their learning bias $\dot{z}_{k\mu}(0), \ddot{z}_{k\mu}(0), \dots$, does not exceed some *uniform* $M > 0$, then the relative scores $z_{k\mu}$ will escape to $-\infty$. In other words, we will have $x(t) \to q$ whenever $x(0)$ is sufficiently close to q and the players' initial learning bias (which is what players use to update (LD$_n$) anyway) is uniformly small.[10]

Remark. Using the extended real arithmetic operations for $-\infty$, the reduced Gibbs map (GM*) maps $(-\infty, \dots, -\infty)$ to q. Interestingly, by adjoining $(-\infty, \dots, -\infty)$ to $\prod_k \mathbb{R}^{\mathcal{A}_k^*}$ in a topology which preserves the continuity of G_k^*, the statement above implies that this "point at negative infinity" is asymptotically stable in (ZD$_n$) – for a detailed statement and proof, see A.

To prove Theorem 5.1, we begin with a quick technical lemma:

Lemma 5.2. *The reduced Gibbs map* $G_k^* : \mathbb{R}^{\mathcal{A}_k^*} \to X_k$ *of* (GM*) *is a diffeomorphism onto its image.*

Proof. It is easy to check that the expressions $z_{k\mu} = \log(x_{k\mu}/x_{k,0})$ provide an inverse to G_k^* for $x_k \in \mathrm{rel\,int}(X_k)$; the claim then follows by noting that all expressions involved are smooth. $\qquad\square$

Proof of Theorem 5.1. We will begin with stationarity of restricted equilibria. Since the payoff term of (RD$_n$) does not contain any higher order derivatives, it will

---

[10]The reason that this reasoning does not apply to nonzero initial strategy growth rates $\dot{x}(0)$ etc. may be seen in a simple 2-strategy context as follows: by (3.9) we will have $\dot{x} = x(1-x)\dot{z}$, so a uniform bound on $\dot{x}$ does not correspond to a uniform bound on z. In particular, if players start with a finite initial score velocity $\dot{z}(0)$ near the boundary of X, then this will correspond to a vanishingly small initial strategy growth rate $\dot{x}(0)$; conversely, nonzero $\dot{x}(0)$ with $x(0)$ arbitrarily near $\mathrm{bd}(X)$ implies an arbitrarily large initial learning bias $\dot{z}(0)$.

Similar conclusions apply to the revision formulation (3.25) of the higher order replicator dynamics with offset aggregate payoffs $U_{k\alpha}(t) = U_{k\alpha}(0) + \int_0^t u_{k\alpha}(x(s))\, ds$. Indeed, if the initial offsets $U_{k\alpha}(0)$ are uniformly small in (3.25), then the corresponding initial conditions $\dot{x}_{k\alpha}(0)$ will scale with $x_{k\alpha}(0)$; it thus makes little evolutionary sense to ask for convergence under a *uniform* bound on the players' initial velocities $\dot{x}_{k\alpha}(0)$ when $x_{k\alpha}(0)$ is itsellf small.
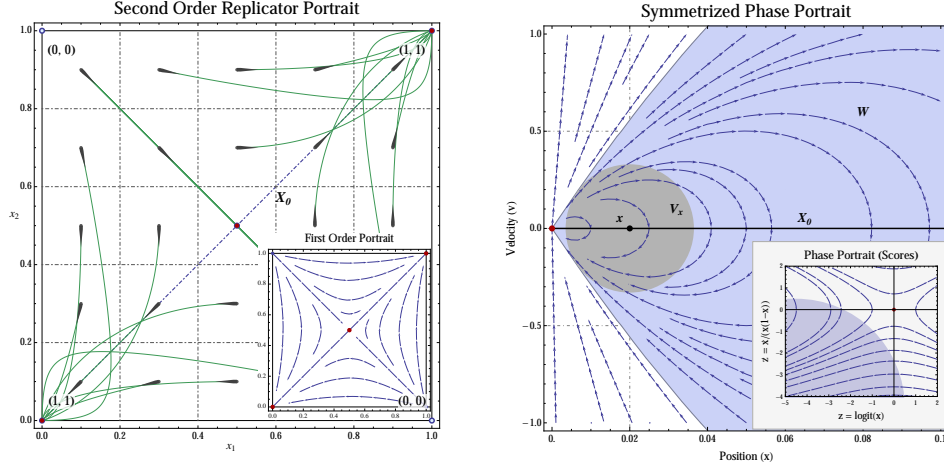
FIGURE 3.   Second order replicator trajectories in a 2×2 coordination game with payoff matrices $U_1 = U_2 = I$. Fig. 3(a) shows the phase portraits for the first and second order replicator dynamics, while Fig. 3(b) shows the restriction of the game's phase space to the symmetric invariant manifold X_0 which joins the game's equilibria. For every symmetric initial point (x, x) near $q = (0, 0)$, there exists a neighborhood of initial states V_x (gray) such that all orbits starting in V_x stay close and eventually converge to q. The union W of these neighborhoods (light blue) is not itself a neighborhood of q in $\Omega_0 \equiv \Omega(X_0)$, so q is not asymptotically stable in (RD$_n$); however, in terms of the score variables $z = \log \frac{x}{1-x}$, $\dot{z} = \dot{x}/x(1-x)$, the corresponding point at negative infinity $(-\infty, \ldots, -\infty)$ attracts all nearby initial states (inlay).

vanish at $q \in X$ if and only if $u_{k\alpha}(q) = u_k(q)$ for all $\alpha \in \text{supp}(q)$, implying that q is a restricted equilibrium. Conversely, let q be a Nash equilibrium in the restriction $\mathfrak{G}' \equiv \mathfrak{G}(\mathcal{N}, X', u_k|_{X'})$ of $\mathfrak{G}$ with $\mathcal{A}'_k = \text{supp}(q_k)$. Then, with $u_{k\alpha}(q) = u_{k\beta}(q)$ for all $\alpha, \beta \in \mathcal{A}'_k$, the updating scheme (LD$_n$) constrained to $\mathfrak{G}'$ and starting at q also gives $y^{(n)}_{k\alpha}(0) = y^{(n)}_{k\beta}(0)$ for all $\alpha, \beta \in \mathcal{A}'_k$. So, if (RD$_n$) starts at q with initial motion rates $\dot{x}(0) = \ddot{x}(0) = \cdots = 0$, we will have $y_{k\alpha}(t) - y_{k\alpha}(0) = y_{k\beta}(t) - y_{k\beta}(0)$ for all $\alpha, \beta \in \mathcal{A}'_k$, and, by the homogeneity of the Gibbs map ($G(y_1 + c, y_2 + c, \ldots) = G(y_1, y_2, \ldots)$ for all $c \in \mathbb{R}$), we readily obtain $x(t) = q$ for all t, i.e. q is stationary.[11]

We now turn to Part (III) of the theorem – which will also prove Part (II). To that end, suppose that every neighborhood U of q in X admits an interior orbit $x(t)$ that stays in U for all $t \geq 0$; we then claim that q is Nash. Indeed, assume instead that for some $k \in \mathcal{N}$, there exists $\beta \in \mathcal{A}_k$ and $\alpha \in \text{supp}(q_k)$ with $u_{k\alpha}(q) < u_{k\beta}(q)$. Then, pick $\varepsilon > 0$ and a neighborhood U of q such that $x_{k\alpha} > q_{k\alpha}/2 > 0$ and $u_{k\beta}(x) \geq u_{k\alpha}(x) + \varepsilon$ for all $x \in U$. By assumption, there exists an interior orbit $x(t)$ which stays in U for all time, so, for the associated score variables $y(t)$, we will have:

$$y^{(n)}_{k\beta}(t) - y^{(n)}_{k\alpha}(t) = u_{k\beta}(x(t)) - u_{k\alpha}(x(t)) \geq \varepsilon > 0.$$

[11]Importantly, Nash equilibria are not stationary in (LD$_n$): orbits that are parallel to the line $(t, \ldots, t)$ in $\mathbb{R}^{\mathcal{A}}$ are stationary in (RD$_n$).

This last inequality immediately implies that $\log \left(x_{k\beta}(t)/x_{k\alpha}(t) \right) \to +\infty$, contradicting the fact that $x_{k\alpha}(t) > q_{k\alpha}/2$ for all $t \geq 0$.

With regards to Part (IV), let $q = (e_{1,0}, \ldots, e_{N,0})$ be a strict equilibrium of $\mathfrak{G}$, and consider the relative scores $z_{k\mu} = y_{k\mu} - y_{k,0}$, $\mu \in \mathcal{A}_k^* \equiv \mathcal{A}_k \setminus \{0\}$ of (5.1). Since the reduced Gibbs map $G_k^* \colon \mathbb{R}^{\mathcal{A}_k^*} \to X_k$ of (GM*) is a diffeomorphism onto its image by Lemma 5.2, the same will hold for the direct sum $G^* \equiv \bigoplus_k G_k^* \colon \prod_k \mathbb{R}^{\mathcal{A}_k^*} \to X$ as well. Accordingly, if we take a neighborhood U_ε of q in X of the form $U_\varepsilon = \{x \in \text{int}(X) : x_{k,0} > 1 - \varepsilon, k \in \mathcal{N}\}$, its preimage under G^* will be the set $V_h = \{z : \mathcal{Z}_{k,0} < h, k \in \mathcal{N}\}$ where $\mathcal{Z}_{k,0} = \sum_\nu^k \exp(z_{k\nu})$ and $h = (1 - \varepsilon)^{-1} - 1$ ($\approx \varepsilon$ for small ε). We will show that if h is chosen small enough, then there exists $\delta > 0$ such that whenever a solution $z(t)$ of (ZD$_n$) starts at $z(0) \in V_h$ with $\|z^{(r)}(0)\| < \delta$ for $r = 1, \ldots, n - 1$, we will have $z(t) \in V_{2h}$ for all $t \geq 0$ and $z_{k\mu}(t) \to -\infty$ for all $\mu \in \mathcal{A}_k^*$, $k \in \mathcal{N}$. Since G^* is a diffeomorphism onto its image and $x \to q$ iff $z_{k\mu} \to -\infty$ for all $\mu \in \mathcal{A}_k^*$, $k \in \mathcal{N}$, this will establish the "if" direction of our claim.[12]

Indeed, let $z(t)$ be a solution of (ZD$_n$) starting in V_h and let $\tau_{2h} = \inf\{t : z(t) \notin V_{2h}\}$ be the time it takes $z(t)$ to escape from V_{2h} (with the usual convention $\inf(\varnothing) = \infty$). Then, if h is taken small enough, there will be some constant $M > 0$ such that $u_{k,0}(x) - u_{k,\mu}(x) \geq M > 0$ for all $x \in G^*(V_{2h})$ (recall that q is a strict equilibrium). In this way, for $t \leq \tau_{2h}$, Taylor's theorem with Lagrange remainder applied to (ZD$_n$) readily gives:

$$z_{k\mu}(t) \leq z_{k\mu}(0) + \sum_{r=1}^{n-1} z_{k\mu}^{(r)}(0) \, t^r/r! - Mt^n/n!. \tag{5.2}$$

Hence, pick $\delta > 0$ such that the maximum of the polynomial $\sum_{r=1}^{n-1} z_{k\mu}^{(r)}(0) \, t^r/r! - Mt^n/n!$ for $t \geq 0$ is strictly smaller than $\log 2$ whenever $|z_{k\mu}^{(r)}(0)| < \delta$ for $\mu \in \mathcal{A}_k^*$, $k \in \mathcal{N}$, and $r = 1, \ldots n - 1$. This readily yields $\mathcal{Z}_{k,0}(\tau_{2h}) < \sum_\mu^k \exp \left(z_{k\mu}(0) + \log 2 \right) = 2\mathcal{Z}_{k,0}(0) < 2h$, i.e. $z(\tau_{2h}) \in V_{2h}$, a conclusion which cannot hold unless $\tau_{2h} = \infty$. We thus obtain $z(t) \in V_{2h}$ for all $t \geq 0$, so the limit of (5.2) as $t \to \infty$ gives $z_{k\mu}(t) \to -\infty$.

For the converse implication, it is easy to show that any vertex q of X which attracts an open neighborhood of initial rest states must also be a strict Nash equilibrium: extending the reasoning of Ritzberger and Weibull (1995, Thm. 1) to our higher order setting, it suffices to consider the evolution of the dynamics in the edge which joins $q = (\alpha_k; \alpha_{-k})$ to a vertex $q' = (\alpha_k'; \alpha_{-k})$ with $u_k(q') \geq u_k(q)$. However, Theorem 5.5 shows that only a vertex $q \in X$ can attract an open set of initial states $\omega \in \Omega$ containing a punctured neighborhood of q in X, so our assertion follows. □

Now, with regards to the equilibration speed of the higher order dynamics, it can be shown that the rate of convergence to a strict equilibrium in the n-th order dynamics (RD$_n$) is n orders as fast as in the first order regime. More specifically, we have:

---

[12]Non-interior trajectories can be handled similarly by looking at an appropriate restriction $\mathfrak{G}'$ of $\mathfrak{G}$.

Proposition 5.3. *Let $q = (e_{1,0}, \ldots, e_{N,0})$ be a strict Nash equilibrium of the finite game $\mathfrak{G}$, and let $x(t)$ be a solution orbit of the replicator dynamics* (RD$_n$) *which starts at rest and close enough to q. Then, there exists a positive constant $c > 0$ such that:*

$$x_{k,0}(t) \sim 1 - \exp\left(-ct^n/n! + \mathcal{O}(t^{n-1})\right). \tag{5.3}$$

Proof. If we choose a sufficiently small neighborhood of initial positions, Theorem 5.1 shows that the payoff differences $u_{k,0}(x(t)) - u_{k,\mu}(x(t))$ will be bounded away from 0 by some positive constant $c > 0$ for all $\mu \in \mathcal{A}_k^*$, $k \in \mathcal{N}$ and for all $t \geq 0$. Hence, with $z_{k\mu}^{(n)} \leq -c < 0$ by (ZD$_n$), our assertion follows from an $(n-1)$-fold application of the mean value theorem. $\square$

5.2. **Dynamic instability of mixed equilibria.** Theorem 5.1 and Proposition 5.3 above characterize the behavior of the n-th order replicator dynamics near strict equilibria from both a qualitative and a quantitative viewpoint; on the flip side, they do not address mixed equilibria. To study this issue, recall first that the standard asymmetric replicator dynamics preserve a certain volume form in the interior of X, so mixed equilibria cannot be attracting in first order. Ritzberger and Weibull (1995) establish this "incompressibility" property of the replicator dynamics by taking an ingenious extrinsic reparametrization which makes the replicator dynamics divergence-free in the interior of X (see also Ritzberger and Vogelsberger, 1990). On the other hand, by exploiting the interplay between the natural logarithm and the Gibbs map (GM), Hofbauer (1996) essentially showed that the replicator dynamics are incompressible in the space of the score variables $y_{k\alpha}$. In the same spirit, we have:

Proposition 5.4. *The flow of the higher order learning dynamics* (LD$_n$) *preserves volume in the usual Euclidean geometry of $\mathbb{R}^\mathcal{A}$ for all $n \geq 1$; the same holds for* (RD$_n$) *w.r.t. a non-Euclidean volume form on the system's phase space Ω.*

Proof. Rewriting (3.14) in continuous time, (LD$_n$) is equivalent to the first order system:

$$\dot{y}_{k\alpha}^{n-1}(t) = u_{k\alpha}(x(t)),$$
$$\cdots \tag{5.4}$$
$$\dot{y}_{k\alpha}^0(t) = y_{k\alpha}^1(t).$$

Thus, given that $y_{k\alpha}^r$ does not appear in the equation for $\dot{y}_{k\alpha}^r$ for $r = 0, \ldots, n-1$, it follows that the flow of (LD$_n$) will be incompressible in the standard Euclidean metric of $\mathbb{R}^\mathcal{A}$. Using the relative scores z of (5.1), the same argument applies to the dynamics (ZD$_n$), and since G^* is a diffeomorphism onto its image by Lemma 5.2, the result carries over to (RD$_n$) as well. $\square$

Thanks to this incompressibility property of (LD$_n$) and (RD$_n$), we have:

Theorem 5.5. *In the higher order replicator dynamics* (RD$_n$)*, interior points cannot attract open sets of initial states; only vertices of X can be attracting. More generally, a non-pure point $q \in X$ can only attract relatively open sets of initial states whose support in X properly contains that of q.*

Proof. We will prove that if $q \in \text{int}(X)$, then there is no open set of initial conditions in Ω that converges to q. The result for general non-pure $q \in X$ will then

follow by focusing on the face X' of X which is spanned by the support of q, i.e. $X' = \prod_k \Delta(\mathcal{A}'_k)$ with $\mathcal{A}'_k = \mathrm{supp}(q_k)$; since the dynamics (RD$_n$) preserve the faces of X, the assertion follows by noting that the intersection of X' with an open set in X is open in X' by definition.

Working with the variables z of (5.1) and recalling that the map $G^*\colon z \mapsto x$ is a diffeomorphism onto its image by Lemma 5.2, Proposition 5.4 shows that open sets of initial states in Ω cannot converge to the interior state $((G^*)^{-1}(q), 0, \ldots, 0)$. Thus, to establish the theorem's claim that $z(t)$ cannot converge to the interior point $z^* \equiv (G^*)^{-1}(q)$, it suffices to show that if $z(t) \to z^*$, then we would also have $\lim_{t\to\infty} \dot{z}(t) = \lim_{t\to\infty} \ddot{z}(t) = \cdots = 0$.

For notational simplicity, we will only prove the case $n = 2$. To that end, assume for the purposes of establishing a contradiction that $z_{k\mu}(t) \to z^*_{k\mu}$ for some $\mu \in \mathcal{A}^*_k$, $k \in \mathcal{N}$, but that $\dot{z}_{k\mu}(t) \not\to 0$. Then, without loss of generality, there exists $\varepsilon > 0$ and an increasing sequence of times $t_n \to \infty$ such that $\dot{z}_{k\mu}(t_n) \geq \varepsilon$ for all n. Hence, let J_n be the largest open interval which contains t_n and which is such that $\dot{z}_{k\mu} > \varepsilon/2$ in J_n; we then claim that the length $\delta_n = m(J_n)$ of J_n vanishes as $n \to \infty$. Indeed, by passing to a subsequence of t_n if necessary, assume that δ_n always exceeded some positive $\delta > 0$; then, with $\dot{z}_{k\mu} > \varepsilon/2$ in J_n, it follows that $z_{k\mu}(t)$ would grow by at least $\varepsilon\delta/2$ over J_n for all n, but since $z_{k\mu}(t)$ converges, every subsequence of $z_{k\mu}(t)$ must also be Cauchy, a contradiction. Then, by the definition of J_n, we will have $\dot{z}_{k\mu} > \varepsilon$ at some interior point of J_n and $\dot{z}_{k\mu} = \varepsilon/2$ at its endpoints; thus, by the mean value theorem, there exists some $\xi_n \in J_n$ with $\ddot{z}_{k\mu}(\xi_n) \geq \varepsilon/2\delta_n$, and hence, $\ddot{z}_{k\mu}(\xi_n) \to +\infty$. However, since $z^*_{k\mu}$ must also be a rest point of (ZD$_n$), the dynamics (ZD$_n$) give $\ddot{z}_{k\mu}(t) \to 0$ as $t \to \infty$, a contradiction. $\qquad\square$

The property that only vertices of X can be attracting in the higher order replicator dynamics (RD$_n$) directly mirrors the first order case. In the following section however, we will show that this is a property of a much more general class of higher order dynamics, so higher order considerations actually sharpen the instability of non-pure equilibria.

6. Extensions: imitative and payoff-monotonic dynamics.

In this section, our aim is to provide several extensions of the higher order dynamics (LD$_n$) and (RD$_n$) and to show how the rationality analysis of the previous sections applies to this more general setting. To that end, if players do not base the updating (LD$_n$) of their performance scores on the payoffs $u_{k\alpha}(x)$ of the game but on a different set of "payoff observables" $w_{k\alpha}\colon X \to \mathbb{R}$ (assumed continuous), then we obtain the generalized reinforcement scheme

$$y^{(n)}_{k\alpha} = w_{k\alpha}(x), \qquad\qquad\qquad (\mathrm{GLD}_n)$$

which, coupled with the logit choice model (GM), yields the generalized n-th order dynamics:

$$x^{(n)}_{k\alpha} = \lambda_k x_{k\alpha}\left(w_{k\alpha}(x) - \sum^k_\beta x_{k\beta} w_k(x)\right) - x_{k\alpha}\left(R^{n-1}_{k\alpha} - \sum^k_\beta x_{k\beta} R^{n-1}_{k\beta}\right). \qquad (\mathrm{GD}_n)$$

The dynamics (GD$_n$) are characterized by the property that if $x(0)$ lies in a subface X' of X, then $x(t)$ will remain in X' for all time: in other words, if the

strategy share $x_{k\alpha}$ of a pure strategy $\alpha \in \mathcal{A}_k$ is initially zero, then it remains zero for all time (see also Remark 3 at the end of Section 3.2). This invariance property is known as "imitation" (Weibull, 1995), so the dynamics (GD$_n$) may be seen as a higher order extension of the class of *imitative game dynamics* introduced by Björnerstedt and Weibull (1996): in particular, (GD$_n$) is the higher order extension of the general imitative equation $\dot{x}_{k\alpha} = x_{k\alpha}(w_{k\alpha} - \sum_{\beta}^{k} x_{k\beta} w_{k\beta})$ in the same way that (RD$_n$) extends the standard replicator dynamics (RD$_1$) to an n-th order setting.

Of course, if the payoff observables $w_{k\alpha}$ are not correlated with the game's payoffs $u_{k\alpha}$, the dynamics (GD$_n$) will not lead to any sort of rational play over time. To account for this, Samuelson and Zhang (1992) considered the aggregate-monotonicity criterion

$$w_k(q'_k; x_{-k}) > w_k(q_k; x_{-k}) \text{ if and only if } u_k(q'_k; x_{-k}) > u_k(q_k; x_{-k}), \quad \text{(AM)}$$

with $x_{-k} \in X_{-k} \equiv \prod_{\ell \neq k} X_\ell$ and $q_k, q'_k \in X_k$. Accordingly, following Hofbauer and Weibull (1996), we will say that the higher order dynamics (GD$_n$) are:

- *aggregate-monotonic* if (AM) holds for all $q_k, q'_k \in X_k$.
- *convex-monotonic* when the "if" direction of (AM) holds for all pure q'_k.
- *concave-monotonic* when the "if" direction of (AM) holds for all pure q_k.
- *payoff-monotonic* if (AM) holds for q_k and q'_k that are *both* pure.

In the first order regime, Samuelson and Zhang (1992) showed that payoff-monotonic (resp. aggregate-monotonic) dynamics eliminate all pure (resp. mixed) dominated strategies. This result was extended by Hofbauer and Weibull (1996) to pure strategies which are dominated by mixed ones in convex-monotonic dynamics, while Viossat (2011) recently established the dual result for concave dynamics. In the same spirit, the rationality analysis of Sections 4 and 5 yields:

Proposition 6.1. *For any interior initial condition, we have:*

- *Aggregate-monotonic n-th order dynamics eliminate all dominated strategies.*
- *Convex (resp. concave) monotonic n-th order dynamics eliminate all pure (resp. mixed) strategies that are dominated by mixed (resp. pure) strategies.*
- *Payoff-monotonic n-th order dynamics eliminate all pure strategies that are dominated by pure strategies.*

If, in addition, players start at rest ($\dot{x}(0) = \ldots = x^{(n-1)}(0) = 0$), then the above conclusions hold with the characterization "dominated" replaced by "weakly dominated". Finally, the rate of extinction is exponential in t^n (or t^{n-1} for weakly dominated strategies) in the sense of (4.2)/(4.10).

Proof. The crucial point in the proof of Theorems 4.1 (resp. Theorem 4.3) is the lower bound for the n-th (resp. $(n-1)$-th) derivative of the difference $V_k(x) = D_{\mathrm{KL}}(q_k \parallel x_k) - D_{\mathrm{KL}}(q'_k \parallel x_k)$ which determines the rate of extinction of dominated (resp. weakly dominated) strategies. Thus, by replacing u by w in (4.5) (resp. (4.12)), and using the appropriate monotonicity condition for each case of dominance (pure/mixed by pure/mixed), our assertion follows along the same lines as Theorem 4.1 (resp. Theorem 4.3). $\square$

In the same spirit, we obtain the following counterpart to the higher order folk theorem for higher order payoff-monotonic dynamics:

Proposition 6.2. *The conclusions of Theorem 5.1 hold for all higher order ($n \geq 2$) payoff-monotonic dynamics.*

This proposition follows by replacing u with w in the proof of Theorem 5.1 and using the payoff-monotonicity condition (AM), so there are no qualitative differences between first and higher order payoff-monotonic dynamics. On the other hand, Proposition 5.4 and Theorem 5.5 hold for a much wider class of higher order dynamics:

Proposition 6.3. *The flow of the generalized learning dynamics* (GLD$_n$)*, $n \geq 2$, is volume-preserving; moreover, the same holds for* (GD$_n$) *w.r.t. a non-Euclidean volume form on X. Consequently, the conclusions of Theorem 5.5 hold for all higher order imitative dynamics of the form* (GD$_n$)*.*

Proof. Since the payoff observables $w_{k\alpha}$ do not depend on $\dot{x}$ and other higher order derivates, incompressibility stems from the equivalent first order formulation (5.4) of (GLD$_n$) with u replaced by w; the conclusions of Theorem 5.5 are then proved in the same way. $\qquad \square$

Interestingly, if $n = 1$ and the payoff observables $w_{k\alpha}$ do not depend on $x_{k\alpha}$, then the system (3.14) remains divergence-free and the conclusions of Proposition 6.3 continue to apply. For instance, this explains why the asymmetric replicator equation is divergence-free whereas its symmetric counterpart isn't: in the case of the former, we have $w_{k\alpha}(x) = u_{k\alpha}(x) = u_k(\alpha; x_{-k})$, a quantity which is independent of $x_{k\alpha}$; in the symmetric case however, if U denotes the payoff matrix of the game being played, then we would have $w_\alpha(x) = u_\alpha(x) = u(\alpha; x) = \sum_\beta U_{\alpha\beta} x_\beta$, showing that, in general, the symmetrized replicator dynamics are not divergence-free.

In view of the above, Proposition 6.3 clashes quite strongly with the first order regime. For instance, if we take Maynard Smith's payoff-adjusted variant of the replicator dynamics (whereby players divide (RD$_1$) by their average payoffs), then there exist games with asymptotically stable interior equilibria (for instance, see the Matching Pennies example of Weibull, 1995). In higher orders however, Proposition 6.3 shows that this can no longer be the case: the learning dynamics (GLD$_n$) endow orbits with a tangential acceleration component, and this acceleration carries them away from interior equilibria and towards the boundary of X. As a result, only vertices of X can be attracting in higher order dynamics of the general class (GD$_n$).

7. Concluding remarks.

The results in the present paper suggest that higher order considerations open the door to some intriguing new questions and directions in the study of learning and evolution in games. For one, the elimination of weakly dominated strategies is a key feature of higher order dynamics which puts them firmly apart from all their first order siblings; coupled with the survival of *iteratively* weakly dominated strategies, this provides a dynamic justification of the well-known $\mathbf{S}^\infty \mathbf{W}$ rationalizability process of Dekel and Fudenberg (1990) which cannot otherwise arise from first order considerations. Furthermore, the population interpretation of our higher order dynamics by means of "long-term" variants of existing revision protocols paves the way to a wide array of new classes of dynamics where the impossibility theorem of Hart and Mas-Colell (2003) no longer bars the way – a point also made by Shamma and Arslan (2005) in the context of derivative action fictitious play algorithms.

Nevertheless, even before considering other classes of higher order dynamics, several important questions remain: For instance, are the higher order replicator dynamics consistent (e.g. as in Sorin, 2009)? What can we expect in symmetric, single-population environments (where payoffs are no longer multilinear) or with respect to setwise solution concepts – such as sets that are closed under better replies (Ritzberger and Weibull, 1995)? Finally, from the point of view of learning, our approach has been focused on continuous time with players being able to observe (or otherwise calculate) the payoffs associated to their mixed strategies. This last assumption is relatively harmless in a nonatomic population setting, but crucial from an atomic point of view; in particular, it is only natural to ask whether our results continue to apply in discrete-time environments with a finite number of players only being able to observe their in-game payoffs.

Appendix A. Asymptotic stability in terms of relative scores.

Our aim in this appendix will be to make precise sense of the asymptotic stability statement for the dynamics (ZD$_n$) in Section 5; also, to simplify notation, we will drop the index k and rely on context to resolve any ambiguities.

To begin with, let $\mathbb{R}' = \mathbb{R} \cup \{-\infty\}$ denote the real number line extended to one end by adjoining $-\infty$. Formally, we will use all extended real number operations for $-\infty$, and $\mathbb{R}'$ will be made into a topological space by defining the basic neighborhoods of $-\infty$ to be all sets of the form $U_a = \{x \in \mathbb{R}' : x < a\}$; by taking the product topology, we will then form the extended product space $Z = \mathbb{R}^{\mathcal{A}^*} \cup \{-\infty\}$ by adjoining the "point at negative infinity" $-\infty \equiv (-\infty, \dots, -\infty)$ to $\mathbb{R}^{\mathcal{A}^*}$. In this way, the reduced Gibbs map (GM*) may be extended to Z by mapping $-\infty$ to $q \in X$, and, by the topology of Z, this extension will be continuous.

To incorporate initial conditions at negative infinity for (ZD$_n$), we will work with the extended configuration space Z and the similarly extended phase space $\Omega' = \Omega \cup (\{-\infty\} \times Z \times \cdots \times Z)$, where Ω denotes the original phase space of (ZD$_n$). To extend the flow of (ZD$_n$) to all of Ω', we will define trajectories starting at $-\infty$ to have $z(t) = -\infty$ for all $t \geq 0$; then, using the extended real number operations for $-\infty$ if needed, we will have $z_\mu(t) = z_\mu(0) + \dot{z}_\mu(0) + \frac{1}{2}\ddot{z}_\mu(0)t^2 + \cdots + \int \cdots \int \Delta u_\mu(x(t_1)) \, dt_1 \cdots dt_n$ for all $t_n \equiv t > 0$ and for all $(z(0), \dot{z}(0), \dots) \in \Omega'$, where $x(s) = G^*(z(s))$. As a consequence of the above, it is then easy to see that the above collection of trajectories indeed defines a continuous flow on Ω' which reduces to the flow of (ZD$_n$) on Ω.

With regards to asymptotic stability, Theorem 5.1 shows that if all initial conditions $z_{k\mu}(0)$, $\dot{z}_{k\mu}(0)$ are uniformly small (say, less than some $a \in \mathbb{R}$), then we will have $z_{\mu}(t) \to -\infty$; moreover, an easy adaptation of the proof of Theorem 5.1 shows that the same will hold for all derivates $\dot{z}_{\mu}(t), \ddot{z}_{\mu}(t), \ldots$ of $z(t)$ as well. This shows that if $z(t)$ starts at a neighborhood of $\{-\infty\} \times \{-\infty\} \times \cdots$ in Ω', then it will converge there; completing the argument for Lyapunov stability as in the proof of Theorem 5.1, we thus obtain:

Proposition. *With respect to the extended real number topology defined above, the state* $\{-\infty\} \times \{-\infty\} \times \cdots$ *which corresponds to the strict equilibrium q is asymptotically stable in* (ZD$_n$).

References

Alvarez, F., 2000: On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, **38 (4)**, 1102–1119.

Attouch, H., X. Goudou, and P. Redont, 2000: The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, **2 (1)**, 1–34.

Balkenborg, D., J. Hofbauer, and C. Kuzmics, 2013: Refined best reply correspondence and dynamics. *Theoretical Economics*, **8**, 165–192.

Björnerstedt, J. and J. W. Weibull, 1996: Nash equilibrium and evolution by imitation. *The Rational Foundations of Economic Behavior*, K. J. Arrow, E. Colombatto, M. Perlman, and C. Schmidt, Eds., St. Martin's Press, New York, NY, 155–181.

Cantrell, C. D., 2000: *Modern mathematical methods for physicists and engineers*. Cambridge University Press, Cambridge, UK.

Daskalakis, C., P. W. Goldberg, and C. Papadimitriou, 2006: The complexity of computing a Nash equilibrium. *STOC '06: Proceedings of the 38th annual ACM symposium on the Theory of Computing*.

Dekel, E. and D. Fudenberg, 1990: Rational behavior with payoff uncertainty. *Journal of Economic Theory*, **52**, 243–267.

Flåm, S. D. and J. Morgan, 2004: Newtonian mechanics and Nash play. *International Game Theory Review*, **6 (2)**, 181–194.

Fraenkel, L. E., 1978: Formulæ for higher derivatives of composite functions. *Mathematical Proceedings of the Cambridge Philosophical Society*, **83 (2)**, 159–165.

Fudenberg, D. and C. Harris, 1992: Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, **57 (2)**, 420–441.

Fudenberg, D. and D. K. Levine, 1998: *The Theory of Learning in Games*, Economic learning and social evolution, Vol. 2. The MIT Press, Cambridge, MA.

Gilboa, I. and A. Matsui, 1991: Social stability and equilibrium. *Econometrica*, **59 (3)**, 859–867.

Hart, S. and A. Mas-Colell, 2003: Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, **93 (5)**, 1830–1836.

Hofbauer, J., 1995: Imitation dynamics for games. Tech. rep., University of Vienna.

Hofbauer, J., 1996: Evolutionary dynamics for bimatrix games: a Hamiltonian system? *Journal of Mathematical Biology*, **34**, 675–688.

Hofbauer, J. and W. H. Sandholm, 2009: Stable games and their dynamics. *Journal of Economic Theory*, **144**, 1710–1725.

Hofbauer, J. and W. H. Sandholm, 2011: Survival of dominated strategies under evolutionary dynamics. *Theoretical Economics*, **6 (3)**, 341–377.

Hofbauer, J. and K. Sigmund, 1988: *The Theory of Evolution and Dynamical Systems*. Cambridge University Press.

Hofbauer, J., S. Sorin, and Y. Viossat, 2009: Time average replicator and best reply dynamics. *Mathematics of Operations Research*, **34 (2)**, 263–269.

Hofbauer, J. and J. W. Weibull, 1996: Evolutionary selection against dominated strategies. *Journal of Economic Theory*, **71**, 558–573.

Landau, L. D. and E. M. Lifshitz, 1976: Statistical physics. *Course of Theoretical Physics*, Pergamon Press, Oxford, Vol. 5.

Lee, J. M., 2003: *Introduction to Smooth Manifolds*. No. 218 in Graduate Texts in Mathematics, Springer-Verlag, New York, NY.

Mertikopoulos, P. and A. L. Moustakas, 2010: The emergence of rational behavior in the presence of stochastic perturbations. *The Annals of Applied Probability*, **20 (4)**, 1359–1388.

Nash, J. F., 1950: Non-cooperative games. Ph.D. thesis, Princeton University.

Ritzberger, K. and K. Vogelsberger, 1990: The Nash field. Tech. rep., Institute for Advanced Studies, Vienna.

Ritzberger, K. and J. W. Weibull, 1995: Evolutionary selection in normal-form games. *Econometrica*, **63**, 1371–99.

Rustichini, A., 1999: Optimal properties of stimulus-response learning models. *Games and Economic Behavior*, **29**, 230–244.

Samuelson, L., 1993: Does evolution eliminate dominated strategies? *Frontiers of Game Theory*, MIT Press, Cambridge, MA.

Samuelson, L. and J. Zhang, 1992: Evolutionary stability in asymmetric games. *Journal of Economic Theory*, **57**, 363–391.

Sandholm, W. H., 2010: *Population Games and Evolutionary Dynamics*. Economic learning and social evolution, MIT Press, Cambridge, MA.

Sandholm, W. H., E. Dokumacı, and R. Lahkar, 2008: The projection dynamic and the replicator dynamic. *Games and Economic Behavior*, **64**, 666–683.

Shamma, J. S. and G. Arslan, 2005: Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Trans. Autom. Control*, **50 (3)**, 312–327.

Smith, M. J., 1984: The stability of a dynamic model of traffic assignment - an application of a method of Lyapunov. *Transportation Science*, **18**, 245–252.

Sorin, S., 2009: Exponential weight algorithm in continuous time. *Mathematical Programming*, **116 (1)**, 513–528.

Taylor, P. D. and L. B. Jonker, 1978: Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, **40 (1-2)**, 145–156.

Viossat, Y., 2011: Deterministic monotone dynamics and dominated strategies, http://arxiv.org/pdf/1110.6246v1.

Weibull, J. W., 1995: *Evolutionary Game Theory*. MIT Press, Cambridge, MA.

FRENCH NATIONAL CENTER FOR SCIENTIFIC RESEARCH (CNRS) AND ÉCOLE POLYTECHNIQUE, PARIS, FRANCE
   E-mail address: rida.laraki@polytechnique.edu
   URL: https://sites.google.com/site/ridalaraki/

FRENCH NATIONAL CENTER FOR SCIENTIFIC RESEARCH (CNRS) AND LABORATOIRE D'INFORMATIQUE DE GRENOBLE, GRENOBLE, FRANCE
   E-mail address: panayotis.mertikopoulos@imag.fr
   URL: http://mescal.imag.fr/membres/panayotis.mertikopoulos/home.html