CONVERGENCE IS NOT ENOUGH: AVERAGE-CASE PERFORMANCE OF NO-REGRET LEARNING DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning in games involves two main challenges, even in settings in which agents seek to coordinate: convergence to equilibria and selection of *good* equilibria. Unfortunately, solving the issue of convergence, which is the focus of state-of-the-art models, conveys little information about the quality of the equilibria that are eventually reached, often none at all. In this paper, we study a class of arbitrary-sized games in which *q-replicator* (QRD), a widely-studied class of no-regret learning dynamics that include gradient descent (GD), standard replicator dynamics (RD), and log-barrier dynamics as special cases, can be shown to converge *pointwise* to Nash equilibria. Turning to our main task, we provide both theoretical and experimental results on the *average case performance* of different learning dynamics in games. For example, in the case of GD, we show a tight average Price of Anarchy bound of 2, for a class of symmetric 2×2 potential games with unbounded Price of Anarchy (PoA). Furthermore, in the same class, we provide necessary and sufficient conditions so that GD outperforms RD in an average case analysis giving novel insights about two of the most widely applied dynamics in game theory. Finally, our experiments suggest that unbounded gaps between average case performance and PoA analysis are common, indicating a fertile area for future work.

1 INTRODUCTION

Multi-agent coordination often involves the solution of complex optimization problems. What makes these problems so hard, even when agents have common (Bard et al., 2020) or aligned interests (Dafoe et al., 2020; Dafoe et al., 2021), is that learning occurs on highly non-convex landscapes; thus, even if the learning dynamics equilibrate, their fixed points may include unnatural saddle points or even local minima of very poor performance (Dauphin et al., 2014). To address this issue, a large stream of recent work has focused on the convergence of optimization-driven (e.g., no-regret) learning dynamics to *good* limit points. Notable results include avoidance of saddle points and convergence of first order methods, e.g., *gradient descent*, to local optima (Ge et al., 2015; Lee et al., 2019; Mertikopoulos et al., 2019), point-wise or last-iterate convergence of various learning dynamics to (proper notions of) equilibria in zero-sum (competitive) games (Daskalakis & Panageas, 2019; Bailey & Piliouras, 2019; Cai et al., 2022) and convergence of no-regret learning to stable points in potential (cooperative) games ()HeliouCM17,PPP17,DBLP:journals/corr/abs-2203-12056,Leo22.

Even though these results seem to provide a sufficient starting point to reason about the *quality of the collective learning outcome*, unfortunately, this is far from being true. Non-trivial game settings routinely possess attracting points of vastly different performance, and this remains true, even if one is able to restrict attention to refined and highly robust notions of equilibria (Flokas et al., 2020).

Nevertheless, and despite the intense interest of the machine learning community to address the problem of *equilibrium selection*, there is a remarkable scarcity of work in this direction. To make matters worse, static, game-theoretic approaches to the problem (Harsanyi, 1973; Harsanyi & Selten, 1988; van Damme, 1987), offer little insight, often none at all, from a dynamic/learning perspective. In this case, the challenge is to show approximately optimal performance not for (almost) all initial conditions (which is not possible), but in expectation, i.e., for uniformly random chosen initial conditions (worst-case versus average-case analysis). This is a fundamentally hard problem since one has to couple the performance of equilibria to the relative size of their *regions of attraction*. However, regions of attraction are complex geometric manifolds that quickly become mathematically

intractable even in low-dimensional settings. Importantly, their analysis requires the combination of tools from machine learning, game theory, non-convex optimization and dynamical systems.

In terms of average case analysis of game theoretic dynamics in coordination/common interest games, the only other references that we know of are Zhang & Hofbauer (2015); Panageas & Piliouras (2016). In fact, Panageas & Piliouras (2016) is the key precursor to our work. Critically, whereas Panageas & Piliouras (2016) focuses exclusively on a single dynamics, i.e., replicator dynamics and bounding its average price of anarchy (APoA) in restricted instances of games such as Stag Hunt, we show how these techniques can be applied much more broadly by addressing novel challenges:

- **Axiomatic challenge:** Can we formally define the notion of Average Price of Anarchy for large classes of dynamics and games?
- **Analytical challenge:** Even if the definitions can be made robust how do we analyze these nonlinear dynamical systems given random initial conditions in the presence of multiple attractors?
- **Experimental/visualization challenge:** Can we develop novel custom visualization techniques as well as showcase that our experimental results have predictive power even in complex high dimensional settings?

Model and Contributions. To make progress in addressing these challenges, we study the *q-replicator dynamics* (QRD), one of the most fundamental and widely-studied classes of multi-agent learning dynamics that include *gradient descent*, *replicator* and *log-barrier* dynamics as special cases (A. Giannou, 2021). We start with our first motivating question which we answer affirmatively by proving *pointwise* convergence of all QRD dynamics to Nash equilibria (NEs) in almost all finite potential games. Potential games include multi-agent interactions in which coordination is desirable, congestion games and games of identical interests as important and widely-studied subclasses (Wang & Sandholm, 2002; Panait & Luke, 2005; Carroll et al., 2019; Dafoe et al., 2020).

The proof of point-wise convergence to NEs combines recent advances (Swenson et al., 2020)[1] with standard convergence techniques in the study of potential games, e.g., Palaiopanos et al. (2017b). Such techniques have been used to either establish convergence of QRD to NEs under the assumption of point-wise convergence (Mertikopoulos & Sandholm, 2016) or prove convergence to limit cycles of (restricted) equilibrium points (Mertikopoulos & Sandholm, 2018). However, whereas in previous works such results are the main focus, in our case they are only the starting point as they clearly not suffice to explain the disparity between the regularity of QRD in theory (bounded regret, convergence to Nash equilibria) and their conflicting performance in practice (agents' utilities after learning).

We then turn to our second question and the fundamental problem of equilibrium quality. While different QRD dynamics may reach the same asymptotically stable equilibria, this is only a minimal and definitely not sufficient condition to compare their performance. In particular, the *regions of attraction* of these common attracting equilibria, i.e., the sets of convergent initial conditions, can be very different for different QRD dynamics.

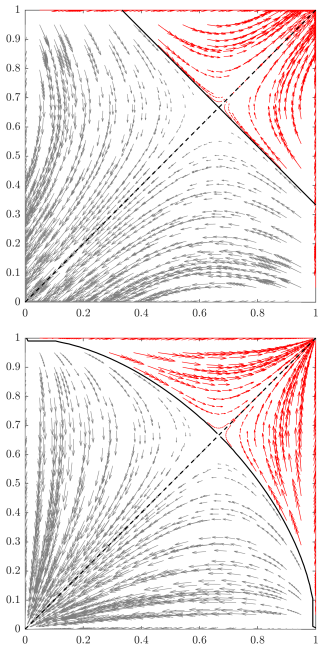

Figure 1: Vector fields of gradient descent (top) and replicator dynamics (bottom) for a game with payoff- and risk-dominant equilibrium at the bottom-left corner. The trajectories in the *region of attraction* of the good (bad) equilibrium are shown in gray (red). The black solid and dashed lines show the stable and unstable manifolds, respectively. In this case, gradient descent outperforms replicator dynamic.

[1]Specifically, Swenson et al. (2020) shows that all NEs in almost all potential games are regular in the sense of Harsanyi, i.e., they are isolated and highly robust Harsanyi (1973); van Damme (1987). *Almost all* refers to a set whose complement is a closed set with Lebesgue measure zero.

In our main technical contribution, we tackle this task by providing geometric insights into the shapes and sizes of the regions of attractions of different QRD dynamics. We show that in a class of two-agent potential games, gradient descent reaches the payoff-dominant (socially optimal) equilibrium more often than standard replicator whenever this equilibrium is also risk-dominant (less risky), see Figure 1. As an implication, we study a class of games in which the *Price of Anarchy* is unbounded, i.e., in which the worst-case equilibrium can be arbitrarily worse than the socially optimal outcome (Panageas & Piliouras, 2016), and derive a (tight) upper bound of 2 for the *Average Price of Anarchy* for the gradient descent dynamics and all instances of the class in which the risk- and payoff-dominant equilibria coincide. This is the first such tight result of its kind.

Conceptually, our methods provide a systematic approach to explore the design and hyperparameter space of learning dynamics and extend recent advances towards a taxonomy of learning dynamics in low-dimensional or general potential games (Panageas & Piliouras, 2016; Pangallo et al., 2022). More importantly, they signify the expressiveness in this task of performance measures that couple the likelihood of convergence to a certain outcome (region of attraction) with the performance of an algorithm at this outcome. From a practical perspective our findings admit a dual interperation. On the one hand, they provide concrete recommendations about the optimality of different QRD dynamics based on the features of the underlying game. On the other hand, they suggest, that even in the simplest possible classes of games, there is not a single optimal QRD dynamic to *beat them all*.

Intriguingly, the above results hinge on two interconnected, yet fundamentally different, theories. The first part (convergence), relies on the theory of Lyapunov analysis and the properties of dissipative systems, i.e., systems that lose momentum over time till they converge to a steady state.By contrast, the second part, i.e., the qualitative analysis of the different parametrizations of the QRD dynamics, relies on the existence of invariant functions that characterize stable and unstable areas in the state space of such systems (Palaiopanos et al., 2017a; Nagarajan et al., 2020). The existence of invariant functions, however, is a feature most often studied in conservative systems, a fundamentally orthogonal principle to the one of dissipation.

2 Preliminaries: Game-Theoretic and Behavioral Models

Game-theoretic model. A multi-agent finite potential game $\Gamma := \{\mathcal{N}, (\mathcal{A}_k, u_k)_{k\in\mathcal{N}}, \Phi\}$ denotes the interaction between a set $\mathcal{N} := \{1, \dots, n\}$ of agents. Each agent $k \in \mathcal{N}$ has a finite set of actions, $\mathcal{A}_k$, with size $|\mathcal{A}_k|$, and a reward function $u_k : \mathcal{A} \to \mathbb{R}$ where $\mathcal{A} := \prod_{k\in\mathcal{N}} \mathcal{A}_k$ is the set of all *pure action profiles* of Γ. Agents may use mixed actions or *choice distributions*, $x_k = (x_{ka_k})_{a_k\in\mathcal{A}_k} \in \mathcal{X}_k$, where x_{ka_k} is the probability with which agent k uses their action $a_k \in \mathcal{A}_k$ and $\mathcal{X}_k := \{x_k \in \mathbb{R}^{|\mathcal{A}_k|} \mid \sum_{a_k\in\mathcal{A}_k} x_{ka_k} = 1,\ x_{ka_k} \geq 0\}$ is the $(|\mathcal{A}_k| - 1)$-dimensional simplex. Given any mixed-action $x_k \in \mathcal{X}_k$, we will write $\mathrm{supp}(x_k) := \{a_k \in \mathcal{A}_k \mid x_{ka_k} > 0\}$ to denote the *support* of the action x_k, i.e., the set of all pure actions $a_k \in \mathcal{A}_k$ that are selected with a positive probability at x_k. Using conventional notation, we write $s = (s_k, s_{-k}) \in \mathcal{A}$ and $x = (x_k, x_{-k}) \in \mathcal{X} := \prod_{k\in\mathcal{N}} \mathcal{X}_k$ to denote the *joint pure and mixed action profiles* of Γ, where s_{-k} and x_{-k} are the vectors of pure and mixed actions, respectively, of all agents other than k. When time is relevant, we will use the index t for all the above, e.g., we will write $x_k(t)$ for agent k's choice distribution at time $t \geq 0$. The function $\Phi : \mathcal{A} \to \mathbb{R}$ is called a *potential function* of Γ and satisfies $u_k(s) - u_k(s'_k, s_{-k}) = \Phi(s) - \Phi(s'_i, s_{-i})$, for all $k \in \mathcal{N}$ and all $s, s' \in \mathcal{A}$. The agents' reward functions and the potential function extend naturally to mixed action profiles with $u_k(x) = \mathbb{E}_{s\sim x}[u_k(s)]$ and $\Phi(x) = \mathbb{E}_{s\sim x}[\Phi(s)]$.

Regular Nash and restricted equilibria. A *Nash equilibrium (NE)* of Γ is an action profile $x^* \in \mathcal{X}$ such that $u_k(x^*) \geq u_k(x_k, x^*_{-k})$, for all $k \in \mathcal{N}$ and for all $x \in \mathcal{X}$. By linearity of expectation, the above definition is equivalent to:

$$u_k(x^*) \geq u_k(a_k, x^*_{-k}), \quad \text{for all } a_k \in \mathcal{A}_k, \text{ and all } k \in \mathcal{N}, \tag{1}$$

where $u_k(a_k, x^*_{-k})$ denotes the reward of agent k when they play the pure action a_k, versus the mixed strategies x^*_{-k} for the rest of the agents. Let $\mathrm{NE}(\Gamma)$ denote the set of all NE of Γ. A NE is called symmetric if $x^*_1 = \ldots = x^*_n$, and is called *fully mixed* if $\mathrm{supp}(x^*) = \prod_{k\in\mathcal{N}} \mathrm{supp}(x^*_k) = \mathcal{A}$. A NE is called *regular* if it satisfies the following definition.

Definition 2.1 (Regular Nash equilibria (Harsanyi, 1973; Swenson et al., 2020))**.** *A Nash equilbrium, $x^* \in NE(\Gamma)$, is called* regular *if it is (i) quasi-strict, i.e., if for each player $k \in \mathcal{N}$, x^*_k assigns positive*

probability to all best responses of player k against x^*_{-k}~~all best responses of each player $k \in \mathcal{N}$ to x^*_{-k} are contained in x^*_k~~, and (ii) second-order non-degenerate, i.e., if the Hessian, $H(x^*)$, taken with respect to $\operatorname{supp}(x^*)$ *is non-singular.*

Finally, a *restriction* of Γ is a game $\Gamma' := \{\mathcal{N}, (\mathcal{A}'_k, u'_k)_{k\in\mathcal{N}}\}$, where $\mathcal{A}'_k \subseteq \mathcal{A}_k$ and $u'_k : \mathcal{A}' \to \mathbb{R}$ is the restriction of u_k to $\mathcal{A}' := \prod_{k\in\mathcal{N}} \mathcal{A}'_k$ for all $k \in \mathcal{N}$. An action-profile $x \in \mathcal{X}$ is called a *restricted equilibrium* of Γ if it is a Nash equilibrium of a restriction of Γ, cf. Mertikopoulos & Sandholm (2018). It is easy to see that all restrictions of a potential game $\Gamma := \{\mathcal{N}, (\mathcal{A}_k, u_k)_{k\in\mathcal{N}}, \Phi\}$ are potential games, whose potential functions are restrictions of Φ to the respective subspaces of $\mathcal{A}$.

Behavioral-learning model. The evolution of the agents' choice distributions (or mixed actions) in the joint action space $\mathcal{X}$ is governed by the *q-replicator dynamics (QRD)* which are the parametric dynamics described by the system of differential equations (equations of motions) $\dot{x} := V_q(x)$, where $V_q : \mathcal{X} \to \mathbb{R}^{|\mathcal{A}|}$ is given by:

$$\dot{x}_{ka_k} = x^q_{ka_k} \left(u_k(a_k, x_{-k}) - \frac{\sum_{a_j\in\mathcal{A}_k} x^q_{ka_j} u_k(a_j, x_{-k})}{\sum_{a_j\in\mathcal{A}_k} x^q_{ka_j}} \right), \quad \text{for all } k \in \mathcal{N}, a_k \in \mathcal{A}_k, \quad \text{(QRD)}$$

for any $q \geq 0$. Special cases of the above dynamics are the projection or *gradient descent (GD)* dynamics, for $q = 0$, the the *standard replicator (RD)* dynamics, for $q = 1$, and the *log-barrier* or inverse update dynamics, for $q = 2$ (Mertikopoulos & Sandholm, 2018; A. Giannou, 2021).

3 POINTWISE CONVERGENCE OF QRD TO NASH EQUILIBRIA

Our results consist of two parts. In the first part, which is the subject of this section, we show convergence of QRD to Nash equilibria in a class of potential games, which we term *perfectly-regular potential games*, whose definition follows.

Definition 3.1 (Perfectly-regular potential games)**.** *A potential game Γ is called* regular *if it has only regular Nash equilibria. A regular potential game is called* perfectly-regular potential game *(PRPG) if all its restrictions are regular potential games, i.e., if they only possess regular Nash equilibria.*

Almost all potential games are PRPGs; this is a generalization of Swenson et al. (2020) who prove that almost all potential games are regular. Furthermore the PRPG class contains other important subclasses of games, e.g., congestion games, as well as games with identical reward functions, which are currently widely studied in the context of cooperative artificial intelligence (Wang & Sandholm, 2002; Panait & Luke, 2005; Carroll et al., 2019; Dafoe et al., 2020). The convergence result is stated formally in Theorem 3.2; its complete proof may be found in Appendix A.

Theorem 3.2 (pointwise convergence of QRD to NE in PRPGs)**.** *Given any perfectly-regular potential game (PRPG), Γ, and any interior initial condition $x(0) \in \operatorname{int}\mathcal{X}$, the q-replicator dynamics, defined as in equation QRD, converge pointwise to a Nash equilibrium x^* of Γ for any parameter $q \geq 0$. Furthermore, the set $\mathcal{Q}(\operatorname{int}\mathcal{X}) := \bigcup_{x_0\in\operatorname{int}\mathcal{X}} \{x^* \in \mathcal{X} \mid \lim_{t\to\infty} x(t) = x^*,\ x(0) = x_0\}$, i.e., the set of all limit points of interior initial conditions, is finite.*

Sketch of the proof. The proof of Theorem 3.2 proceeds in two steps, which utilize the properties that (i) PRPGs have a finite number of regular equilibria, and (ii) the probability of optimal actions *near* an equilibrium point is increasing in time with respect to the QRD. In the first step, we prove that for any initial condition, the sequence of joint action profiles $x(t)_{t\geq 0}$ that is generated by QRD for any $q \geq 0$ converges to a restricted equilibrium of a PRPG, Γ. This relies on the fact that the set of cluster (limit) points of the trajectory—also called the ω-limit set—is a finite, and in fact, as we show, a singleton (a single element set) for any PRPG. In turn, this follows from the fact that a PRPG provably contains only a finite number of restricted equilibria.

Having established convergence to restricted equilibria, in the second step, it remains to show that, in fact, any such limit point has to be a NE of Γ, i.e., we need to exclude convergence to restricted equilibria that are not NE of Γ. To establish this, we couple the structure of PRPGs, which ensures that there is a finite number of (regular) restricted equilibria, with the nature of QRD which guarantees that in the vicinity of a limit point, optimal actions, i.e., best responses, need to be played with

increasingly higher probability. Thus, all actions in the support of the limit choice distribution of each agent must be best responses against the actions of all other agents, which implies that all points that can be reached by QRD are NE of Γ. $\square$

In other words, Theorem 3.2 says that for *almost all potentials games* and *almost all initial conditions*, QRD converge to a NE of the game.An important implication of Theorem 3.2 is that, when one is reasoning about the quality of the collective learning outcome in cooperative multi-agent settings (as captured by PRPGs), they can restrict their attention to NE. However, reverting to off-the-shelf, static performance measures that compare the quality of different NE, we only obtain results that are meaningless, if not misleading, from a dynamic/learning perspective. The reason is that certain *bad* (or sometimes even *good*) NE may be reachable only from a very small set of initial conditions. Thus, we need to develop and argue about *average* performance measures that couple the outcome of the learning process (NE of an PRPG), with the likelihood that such an outcome is reached by the given learning dynamic (region of attraction of this NE). This is the subject of the next section.

4 Quality of the collective learning outcome

When static performance metrics fail. Having established that in the landscape of potential games, QRD converge *almost surely* to Nash equilibria, we next turn our attention to the main challenge of quantifying the quality of the collective learning outcome. In order to do that, one would first have to establish *appropriate* performance metrics. In a static regime, we can rely in a variety of meaningful metrics, e.g., the *Price of Anarchy (PoA)* (Koutsoupias & Papadimitriou, 1999; Christodoulou & Koutsoupias, 2005; Roughgarden, 2015), which is defined as the ratio between the *socially worst* NE of the game and the *socially optimal* outcome, where the social-optimality of an outcome $x \in \mathcal{X}$ is measured with respect to the *social welfare* $\mathrm{SW}(x) \coloneqq \sum_{k \in \mathcal{N}} u_k(x)$, i.e., the total reward of the agents. The PoA is a natural *static* metric that one may consider in a PRPG setup. After all, coordination is the essence of potential games, which typically model multi-agent settings where this is a desirable property. However, it is not difficult to find PRPGs where the PoA fails to provide any meaningful information about the game. Let us consider the following example:

Example 4.1 (A simple example of unbounded performance loss)**.** *Consider the parametric* 2×2*-PRPG,* Γ_w*, i.e., a* 2*-player* 2*-actions PRPG, with payoff functions* $u_{w,1}(s_1, s_2) = u_{w,2}(s_2, s_1) = A_w(s_1, s_2)$*, where the matrix* $A_w \in \mathbb{R}^{2\times 2}$ *is given by:*

$$A_w = \begin{pmatrix} 1 & 0 \\ 0 & w \end{pmatrix}, \quad 1 \leq w. \tag{2}$$

The games, Γ_w*, are already expressive enough to capture the aforementioned problem. In order to see this, observe that the NE that corresponds to* $x_1 = (1, 0)$ *and* $x_2 = (1, 0)$ *has social welfare equal to* $\mathrm{SW}(x) = 1 + 1 = 2$ *but the NE that corresponds to* $x'_1 = (0, 1)$ *and* $x'_2 = (0, 1)$ *has* $\mathrm{SW}(x') = w + w = 2w$*. Since* w *can take any value larger than* 1*, the difference in performance can be* arbitrary large *with respect to the PoA. Specifically, PoA*$(\Gamma_w) = \frac{\mathrm{SW}(x')}{\mathrm{SW}(x)} = w \to \infty$ *as* $w \to \infty$*.*[2]

4.1 Regions of attraction and average performance measures

While useful in static environments, the PoA metric fails to capture the dynamic nature of multi-agent learning. In particular, it does not provide an answer to the question:

> *How likely is it for the agents to reach a good or bad outcome given that the multi-agent system converges?*

To answer this question and argue about the collective performance of the game dynamics, we need to quantify the likelihood of each outcome when we the initial conditions of the system are randomly sampled. A region of attraction of a given outcome formalizes this notion.

[2]Note about notation: In a 2×2-game, one usually abuses notation and writes $x, y \in [0, 1]$ (instead of $(x, 1 - x)$ and $(y, 1 - y)$) to denote the mixed choice distributions of players 1 and 2, respectively. Then, all notions that we presented in Section 2 may be viewed as functions of x, y. For example, we could have written $\mathrm{SW}(1, 1)$ to denote the social welfare of the NE that corresponds to $x = y = 1$. In this section we are going to interchange between the two notations, but our choice is always going to be clear by the context.

Definition 4.2 (Regions of attraction)**.** *Let Γ be any game and assume that its joint action profile, $x \in \mathcal{X}$, is evolving according to the equations of motion $\dot{x} = f(x)$. Then for any $x^* \in \mathcal{X}$, the set $\mathit{RoA}_{f,\Gamma}(x^*) := \{x_0 \in \mathcal{X} \mid \lim_{t\to\infty} x(t) = x^*,\, x(0) = x_0\}$ is called the* region of attraction (RoA) *of x^* with respect to the dynamics f.*

In other words, the RoA of a point $x^* \in \mathcal{X}$ is the set of all initial conditions in $\mathcal{X}$ for which the dynamics asymptotically converge to x^*. Note that RoAs do not intersect. If we can determine the regions of attraction of some game dynamics, then given a certain static performance metric, e.g., the social welfare, we can define a corresponding *average-performance metric* that *weighs-in* all possible outcomes, in the sense of limit points, according to their likelihood of occurring with respect to the given dynamics. In order for this *average* to be *meaningful*, a minimum requirement, is that the dynamics converge for almost all, i.e., all but a measure zero, initial conditions. Formally, an average performance metric is defined as follows[3]:

Definition 4.3 (Average-performance metric)**.** *Let Γ be a multi-agent game and assume that its joint action profile, $x \in \mathcal{X}$, is evolving according to the equations of motion $\dot{x} = f(x)$. Let $\mathcal{X}_0 \subseteq \mathcal{X}$ be a set of initial conditions such that the set of convergence points $\mathcal{Q}(\mathcal{X}_0)$ is finite. Then, given a performance metric $g : \mathcal{X} \to \mathbb{R}$ of Γ, the average-performance of the dynamics governed by f in Γ with respect to the performance metric g and the set of initial condition $\mathcal{X}_0$, is given by*

$$\mathit{APM}_{g,\mathcal{X}_0}(f,\Gamma) := \sum_{x^*\in\mathcal{Q}(\mathcal{X}_0)} \mu(\mathit{RoA}_{f,\Gamma}(x^*)) \cdot g(x^*), \tag{APM}$$

where μ is a probability measure on $\mathcal{X}_0$.

In other words, an APM is the expected optimality of a random initialization of the dynamics in $\mathcal{X}_0 \subseteq \mathcal{X}$ with respect to some metric g. For instance, if the performance metric g is the social welfare, then the average-performance metric with respect to g measures the expected social welfare of the system for any random initialization in $\mathcal{X}_0$. The average-performance metric that we are going to use in the remainder of this section is the *Average Price of Anarchy (APoA)*. The APoA is an APM with respect to the social welfare, re-normalised such that the APoA is greater than equal to 1, with equality only if (almost) all the initial conditions converge to the socially optimal outcome of the system. Formally, given a multi-agent game Γ, equations of motion $\dot{x} = f(x)$ that describe the evolution of the agents actions in Γ, and a set of initial conditions $\mathcal{X}_0 \subseteq \mathcal{X}$ that consists of almost all $\mathcal{X}$, the APoA is given by the formula:

$$\mathrm{APoA}(f,\Gamma) = \frac{\max_{x\in\mathcal{X}} \mathrm{SW}(x)}{\mathrm{APM}_{\mathrm{SW},\mathcal{X}_0}(f,\Gamma)}. \tag{APoA}$$

Here, it is important to note that Definition 4.3 does not ensure that an APM is always a *meaningful* metric for the system. However, as long as one can prove that (i) the dynamics converge pointwise to some $x^* \in \mathcal{Q}(\mathcal{X}) \subseteq \mathrm{NE}(\Gamma)$ for almost all initial condition $x_0 \in \mathcal{X}$, and (ii) the set of limit points, $\mathcal{Q}(\mathcal{X})$, is finite —two conditions that are satisfied by any PRPG that evolves with respect to some QRD (cf. Theorem 3.2)—the APoA has an intuitive interpretation. Specifically, in this setup, the APoA is always bounded between the PoA and the *Price of Stability (PoS)* of the game, i.e., the ratio between the socially optimal outcome and the socially optimal NE.

4.2 The taxonomy of QRD in 2×2 PRPGs

To systematically evaluate and compare the performance of different QRD in perfectly-regular finite potential games, we address the case of symmetric 2×2 coordination games, i.e., games in which one can change the identities of the players without changing the payoff to the actions. Such games constitute one of the current *frontiers* in terms of classification of game-dynamics (Zhang & Hofbauer, 2015; Pangallo et al., 2022). Such games are trivially potential games and include games of identical payoffs as special cases. Omitted definitions and proofs of this section may be found in Appendix B.

Representation of symmetric 2×2 PRPGs. Recall that a NE, x^*, of a symmetric potential Γ is called *payoff-dominant* if $u_k(x^*) \geq u_k(x')$ for all $x' \in \mathrm{NE}(\Gamma)$, and it is called *risk-dominant* if x^* is

[3]For this definition, recall that a probability measure μ on a compact space $\mathcal{X}$ is a σ-additive function from the powerset of $\mathcal{X}$ to $\mathbb{R}_+$ such that $\mu(\mathcal{X}) = 1$ and $\mu(\mathcal{X}') \geq 0$ for all $\mathcal{X}' \subseteq \mathcal{X}$.

unilaterally optimal against the uniform distribution of the rest of the agents. All symmetric 2×2 PRPGs can be conveniently represented by the parametric class of games $\Gamma_{w,\beta}$, with payoff functions $u_{w,\beta,1}(s_1, s_2) = u_{w,\beta,2}(s_2, s_1) = A_{w,\beta,s_1,s_2}$, where the matrix $A_{w,\beta} \in \mathbb{R}^{2\times 2}$ is given by:

$$A_{w,\beta} = \begin{pmatrix} 1 & 0 \\ \beta & w \end{pmatrix}, \quad \beta \leq 1 \leq w. \tag{3}$$

The game $\Gamma_{w,\beta}$ has the same NE as the original game, retains the payoff- and risk-dominance properties of its equilibrium points, and preserves the limiting behavior of any QRD (see Lemma B.1). Each game $\Gamma_{w,\beta}$ has three NE, two pure at $x = y = 0$ and $x = y = 1$, with social welfare $\mathrm{SW}(0, 0) = 2w$ and $\mathrm{SW}(1, 1) = 2$, respectively, as well as one fully-mixed NE at:

$$x^* = y^* = \alpha := \frac{w}{w + 1 - \beta}. \tag{4}$$

For convenience, we are going to refer to the first pure-NE as x_w. Note that x_w is payoff-dominant for any parametrization $\Gamma_{w,\beta}$, and it is also risk-dominant whenever $w > 1 - \beta$, or equivalently, whenever $\alpha > 0.5$. The first result of this section states that whenever the risk- and payoff-dominant equilibria of $\Gamma_{w,\beta}$ coincide, i.e., $\alpha \geq 0.5$, then the gradient descent dynamics, i.e., the 0-replicator dynamics, perform better (or equally in the generic case $\alpha = 0.5$) on average than the standard replicator dynamics with respect to the social welfare of their outcomes, i.e., they yield a smaller APoA. In any other instance of these games, i.e., for $\alpha < 0.5$, the RD perform better than GD with respect to the same metric.

Theorem 4.4 (Performance of QRD in symmetric 2×2 PRPG)**.** *Given any 2×2 symmetric PRPG, which, without any loss of generality, can be represented as an instance $\Gamma_{w,\beta}$, it holds that*

$$APM_{\mathrm{SW},\mathrm{int}\,\mathcal{X}}(V_0, \Gamma_{w,\beta}) \geq APM_{\mathrm{SW},\mathrm{int}\,\mathcal{X}}(V_1, \Gamma_{w,\beta}) \tag{5}$$

if and only if ~~whenever~~ the payoff-dominant equilibrium is also risk-dominant, with equality only when ~~if and only if~~ $\alpha = 0.5$, i.e., $w = 1 - \beta$, where V_0, V_1 are the equations of motion of the 0-replicator and 1-replicator dynamics, respectively equation QRD.

Interpretation of Theorem 4.4. The proof of Theorem 4.4 proceeds with a first order analysis of the manifolds that separate the regions of attractions of the two pure equilibria for the different dynamics (cf. Figures 2 and 3). When comparing the gradient descent (GD) dynamics and the replicator dynamics (RD), the main implication of this theorem is that the expected social welfare is optimized by GD whenever risk and payoff-dominant equilibria coincide and is optimized by RD when risk and payoff-dominant equilibria differ. More generally, this result may be interpreted in two ways. On the one hand, it provides a concrete recommendation on the optimal behavior of the agents (GD versus RD) based solely on the properties of the underlying game. On the other hand, it suggests that even in the low-dimensional setting of 2×2 potential games, there is not a uniform recommendation, and the optimal behavior largely depends on the features of the underlying game. As it turns out, in this case, the decisive feature is the *riskiness* of the payoff-dominant equilibrium.

Generalization to all QRD. Technically, the proof of Theorem 4.4 uses tools that are orthogonal to the Lyapunov analysis, and the theory of dissipation of dynamical systems, that we used to prove convergence to NE in section 3. It leverages the *constants of motion* or *invariant functions* (Nagarajan et al., 2020), i.e., quantities that remain constant along the trajectories of the learning dynamics. The rationale is that if one could identify such a function, then, by finding its value at the unique mixed equilibrium α of the game, they can determine all initial conditions that asymptotically converge to it: these will be all points at the same level set of the invariant function. The manifold, i.e., the geometric locus, of all the points that converge to the equilibrium, i.e., the *stable manifold* of α, is the one that separates the regions of attractions of the two pure NE of the game. Because of this property, we may also refer to the stable manifold of the mixed NE as the *separatrix* (Panageas & Piliouras, 2016). Note that, since the dynamics are also backward-invariant (Panageas & Piliouras, 2016; Mertikopoulos & Sandholm, 2018), their level-set will also contain a set of initial conditions that converge to it when moving backward in time. This points constitute the *unstable manifold* of α. In the following lemma we identify such an invariant for all QRD.

Lemma 4.5 (Invariant functions of QRD in 2×2 symmetric PRPGs)**.** *Given a 2×2 symmetric PRPG, $\Gamma_{w,\beta}$, whose agents evolve with respect to the q-replicator dynamics, the separable function $\Psi_q : (0, 1)^2 \to \mathbb{R}$ with $\Psi_q(x, y) := \psi_q(x) - \psi_q(y)$, where $\psi_q : (0, 1) \to \mathbb{R}$ is given by:*

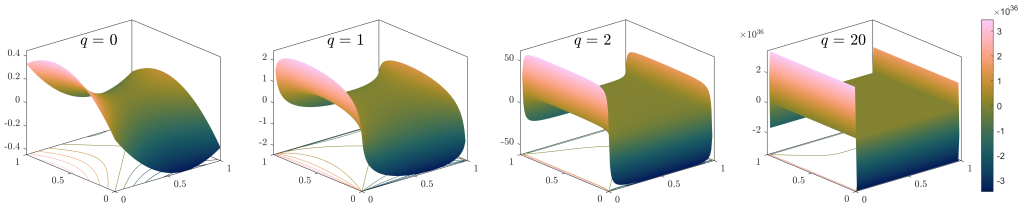

Figure 2: The invariant function, $\Psi_q(x, y)$, for all $x, y \in [0, 1]^2$ in the game $\Gamma_{w,\beta}$ for $w = 2$, $\beta = 0$, and various values of q: $q = 0$ (gradient descent), $q = 1$ (standard replicator), $q = 2$ (log-barrier), and $q = 20$. The invariant function becomes very steep at the boundary as q increases, taking both arbitrarily large negative (**dark**) and positive (light) values in the vicinity of the NE.

$$\psi_q(x) = \begin{cases} \dfrac{x^{2-q} + (1-x)^{2-q} - 1}{2-q} + \dfrac{1 - \alpha x^{1-q} - (1-\alpha)(1-x)^{1-q}}{1-q}, & q \neq 1, 2, \\ \alpha \ln(x) + (1-\alpha)\ln(1-x), & q = 1, \\ \ln(x) + \ln(1-x) + \dfrac{\alpha}{x} + \dfrac{1-\alpha}{1-x}, & q = 2, \end{cases} \tag{6}$$

remains constant along any trajectory $\{x(t), y(t)\}_{t \geq 0}$ *of the system. The function* $\Psi_q(x)$ *is continuous with respect to the parameter* q *at, both,* $q = 1$ *and* $q = 2$*, since* $\lim_{q \to 1} \Psi_q(x) = \Psi_1(x)$ *and* $\lim_{q \to 2} \Psi_q(x) = \Psi_2(x)$ *for all* $x \in (0, 1)$.

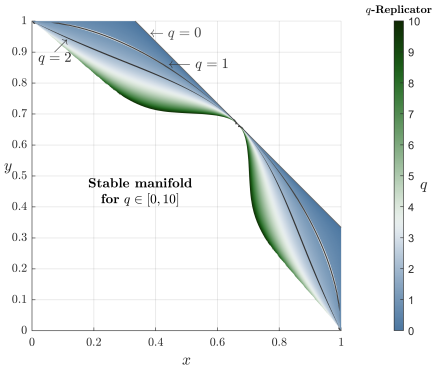

Figure 4: Stable manifold (separatrix) for all different values of $q \in [0, 10]$ (from blue to brown) in the $\Gamma_{w,\beta}$ game for $w = 2$ and $\beta = 0$. The manifolds for $q = 0$, $q = 1$, and $q = 2$ are shown in shades of black for reference (cf. Figure 3). The region of attraction of the payoff-dominant equilibrium (bottom-left corner) shrinks as q increases.

In Figure 2, we visualize the invariant function, $\Psi_q(x, y)$, for $x, y \in (0, 1)^2$ for various values of $q \in [0, 20]$. From the panels of Figure 2, it is also evident that $\Psi_q(x, y)$ acts as a handy tool to visualize the regions of attraction of the two pure NE of the game. Namely, at the unique mixed NE, i.e., at $x = y = \alpha$, the invariant function, Ψ_q, is equal to 0. The same holds for any point $(x, y) \in (0, 1)^2$ with $x = y$. Thus, we can factorize $\Psi_q(x, y)$ as $\Psi_q(x, y) = \Psi_{q,\text{Stable}}(x, y) \cdot (x - y)$ where $\Psi_{q,\text{Stable}}(x, y) = 0$ is precisely the geometric locus of all points $(x, y) \in (0, 1)^2$ such that $\lim_{t \to \infty} x(t) = \alpha$, and $y = x$ is the geometric locus of all points such $\lim_{t \to -\infty} x(t) = \alpha$. These two manifolds constitute the *stable* and *unstable* manifolds, respectively, of the q-replicator dynamics.

Since the invariant function $\Psi_q(x, y)$ takes the value 0 only at the stable and unstable manifolds, we can visualize the separatrix for different values of q by plotting the 0-level set of the invariant functions in Figure 2. These are depicted in Figure 3. As a sanity check, we also see from Figure 3 that the region of attraction of the payoff-dominant equilibrium for $q = 0$ (GD dynamics) is larger than the region of attraction for $q = 1$ (RD).

Empirical evidence for the monotonicity of the APM with respect to q. If we stack the stable manifolds (solid blue lines) in the panels of Figure 3, it becomes evident that the region of attraction of the payoff-dominant and risk-dominant equilibrium grows as q decreases to 0. This is depicted in Figure 4 for all values of $q \in [0, 10]$ (the progression of the surface remains essentially unchanged for larger q). Analogous plots (but with the results reversed as predicted by Theorem 4.4) can be generated for instances of $\Gamma_{w,\beta}$, in which the risk-dominant equilibrium is different from the payoff-dominant one, as well as, for 2×2 generic PRPGs (cf. section 4). In general, putting together Theorem 4.4 and the aforementioned visualizations, we have both *theoretical* and *empirical* evidence that the region of attraction of the payoff-dominant equilibrium in $\Gamma_{w,\beta}$ is decreasing (increasing) in q for $q \geq 0$ whenever this equilibrium is (is not) risk-dominant. Formal verification of the monotonicity of

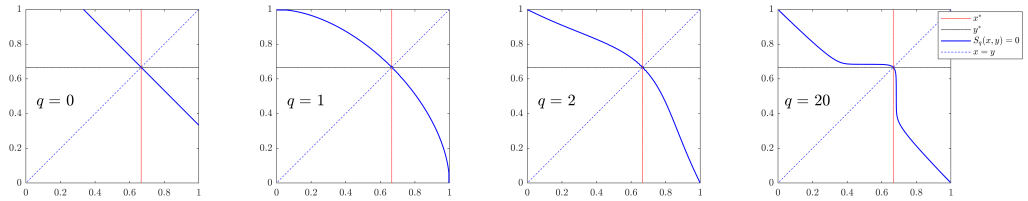

Figure 3: The stable manifolds, $\Psi_{q,\text{Stable}}(x, y) = 0$, (solid blue lines) for the same values of q and the same instance of $\Gamma_{w,\beta}$ as in Figure 2, in which the payoff- and risk- dominant NE is at the bottom left corner. For all q, the separatrix goes through the mixed NE at the intersection of the x^* (dashed red) and y^* (dashed black) coordinates. All panels also include the unstable manifold defined by $x - y = 0$ (dashed blue line). The region of attraction of the payoff-dominant NE is larger for all values of q; however, this is because this NE is also risk-dominant, cf. Theorem 4.4.

~~the stable manifolds~~ the regions of attractions with respect to q in the QRD parametrization remains open.

Application: APoA in 2×2 PRPGs. We conclude this section by providing a concrete result regarding the evaluation of the APoA average-performance measure in the class of 2×2 symmetric PRPGs, which showcases the practical importance of Theorem 4.4 and the invariant function approach.

Theorem 4.6. *The APoA of GD dynamics in all 2×2 symmetric PRPGs, $\Gamma_{w,\beta}$, is bounded by 2, i.e., APoA$(V_0, \Gamma_{w,\beta}) \leq 2$. Furthermore, this bound is tight.*

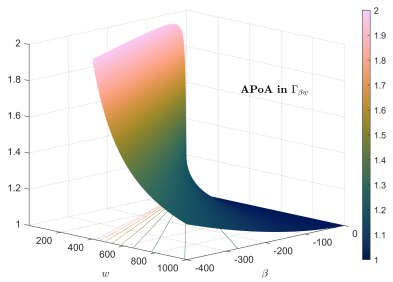

Figure 5: APoA of a 2×2 symmetric PRPG for the gradient descent dynamics and various values of β and w. The APoA is upper bounded by 2 (**dark** to light values) as shown in Theorem 4.6.

The bound also holds for $\beta = 1 - w$, but in this case, there exists no risk-dominant equilibrium. The proof of Theorem 4.6 essentially proceeds by first order analysis of the function depicted in Figure 5 which, in turn, depends on the invariant function of the gradient descent dynamic. One way to see that this bound is tight, is to set $\beta = 1 - w + \epsilon$, for a small $\epsilon > 0$ and let w increase (cf. Figure 5). In combination Theorem 4.4 and Theorem 4.6 imply that the APoA of the RD (QRD with $q = 1$), is *not* upper bounded by 2 whenever $\alpha < 0.5$, i.e., whenever the risk- and payoff-dominant equilibria are different. However, for the case $\alpha > 0.5$, the separatrices for all $q \geq 0$ as visualized in Figure 4, (empirically) imply that similar bounds hold for all values of q.[4] In appendix C, we run simulations of q-replicator dynamics which provide evidence that the statement of Theorem 4.4 and the bound of Theorem 4.6 continue to hold in PRPGs of higher dimensions, i.e., beyond the 2×2 setting.

5 Conclusions

In this paper, we studied the class of q-replicator dynamics (QRD), and showed that all QRD converge pointwise to Nash equilibria in perfectly-regular potential games, a class of games that encompasses almost all potential games, i.e., the standard models of multi-agent coordination. The convergence of QRD in these settings is remarkably robust, occurring regardless of the number of agents or actions and for all possible parametrizations of QRD. From the perspective of equilibrium selection and quality, however, convergence provides little information, often none at all. Turning to this challenging problem, we provided geometric insights into the reasons why different dynamics exhibit fundamentally different performance despite their convergence to the very same set of attracting points. Our techniques leverage two intertwined, yet orthogonal to each other elements of dynamical systems theory: dissipation (Lyapunov theory) and conservation (invariant functions).

[4] To avoid confusion, in Figure 4, we visualize the stable manifolds for the case in which GD are the dynamics with the *largest* region of attraction, i.e., has the lowest APoA. The case $\alpha < 0.5$, in which the manifolds are simply mirrored on the $y = 1 - x$ diagonal, is in Appendix B.

REFERENCES

P. Mertikopoulos A. Giannou, E.-V. Vlatakis-Gkaragkounis. On the Rate of Convergence of Regularized Learning in Games: From Bandits and Uncertainty to Optimism and Beyond. Part of Advances in Neural Information Processing Systems 34 pre-proceedings (NeurIPS 2021), 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bf40f0ab4e5e63171dd16036913ae828-Abstract.html.

J. Bailey and G. Piliouras. Fast and Furious Learning in Zero-Sum Games: Vanishing Regret with Non-Vanishing Step Sizes. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/3501672ebc68a5524629080e3ef60aef-Paper.pdf.

N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, I. Dunning, S. Mourad, H. Larochelle, M. G. Bellemare, and M. Bowling. The Hanabi challenge: A new frontier for AI research. *Artificial Intelligence*, 280:103216, 2020. ISSN 0004-3702. doi: 10.1016/j.artint.2019.103216.

Y. Cai, A. Oikonomou, and W. Zheng. Tight Last-Iterate Convergence of the Extragradient Method for Constrained Monotone Variational Inequalities. *CoRR*, abs/2204.09228, 2022. URL https://doi.org/10.48550/arXiv.2204.09228.

M. Carroll, R. Shah, M. K. Ho, T. Griffiths, S. Seshia, P. Abbeel, and A. Dragan. On the Utility of Learning about Humans for Human-AI Coordination. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32, 2019.

G. Christodoulou and E. Koutsoupias. The price of anarchy of finite congestion games. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pp. 67–73, New York, NY, USA, 2005. Association for Computing Machinery. doi: 10.1145/1060590.1060600.

A. Dafoe, E. Hughes, Y. Bachrach, T. Collins, K. R. McKee, J. Z. Leibo, K. Larson, and T. Graepel. Open Problems in Cooperative AI. *arXiv e-prints*, December 2020.

A. Dafoe, Y. Bachrach, G. Hadfield, E. Horvitz, K. Larson, and T. Graepel. Cooperative AI: machines must learn to find common ground. *Nature*, 7857:33–36, 2021. doi: 10.1038/d41586-021-01170-0.

C. Daskalakis and I. Panageas. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. In Avrim Blum (ed.), *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, 2019.

Y. N. Dauphin, R. Pascanu, Ç. Gülçehre, K. H. Cho, S. Ganguli, and Y. Bengio. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pp. 2933–2941, 2014.

L. Flokas, E. V. Vlatakis-Gkaragkounis, T. Lianeas, P. Mertikopoulos, and G. Piliouras. No-Regret Learning and Mixed Nash Equilibria: They Do Not Mix. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

R. Ge, F. Huang, C. Jin, and Y. Yuan. Escaping From Saddle Points - Online Stochastic Gradient for Tensor Decomposition. In Peter Grünwald, Elad Hazan, and Satyen Kale (eds.), *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, pp. 797–842. JMLR.org, 2015.

J. C. Harsanyi. Oddness of the number of equilibrium points: A new proof. *International Journal of Game Theory*, 2(1):235–250, Dec 1973. ISSN 1432-1270. doi: 10.1007/BF01737572. URL https://doi.org/10.1007/BF01737572.

J. C. Harsanyi and R. Selten. *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press., 1988.

Jun Honda. Equilibrium selection for symmetric coordination games with an application to the minimum-effort game. CiteSeerX, 2012.

E. Koutsoupias and C. Papadimitriou. Worst-Case Equilibria. In Christoph Meinel and Sophie Tison (eds.), *STACS 99*, pp. 404–413, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg. doi: 10.1007/3-540-49116-3_38.

J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht. First-order methods almost always avoid strict saddle points. *Math. Program.*, 176(1-2):311–337, 2019.

S. Leonardos and G. Piliouras. Exploration-exploitation in multi-agent learning: Catastrophe theory meets game theory. *Artificial Intelligence*, 304:103653, 2022. ISSN 0004-3702. doi: 10.1016/j.artint.2021.103653.

P. Mertikopoulos and W. H. Sandholm. Learning in Games via Reinforcement and Regularization. *Mathematics of Operations Research*, 41(4):1297–1324, 2016. doi: 10.1287/moor.2016.0778.

P. Mertikopoulos and W. H. Sandholm. Riemannian game dynamics. *Journal of Economic Theory*, 177:315–364, 2018. ISSN 0022-0531. doi: 10.1016/j.jet.2018.06.002.

P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar, and G. Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

S. G. Nagarajan, D. Balduzzi, and G. Piliouras. From chaos to order: Symmetry and conservation laws in game dynamics. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7186–7196. PMLR, 13–18 Jul 2020. URL `https://proceedings.mlr.press/v119/nagarajan20a.html`.

G. Palaiopanos, I. Panageas, and G. Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5872–5882, 2017a.

G. Palaiopanos, I. Panageas, and G. Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017b. URL `https://proceedings.neurips.cc/paper/2017/file/e93028bdc1aacdfb3687181f2031765d-Paper.pdf`.

I. Panageas and G. Piliouras. Average case performance of replicator dynamics in potential games via computing regions of attraction. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pp. 703–720, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450339360. doi: 10.1145/2940716.2940784.

L. Panait and S. Luke. Cooperative Multi-Agent Learning: The State of the Art. *Autonomous Agents and Multi-Agent Systems*, 11(3):387–434, Nov 2005. doi: 10.1007/s10458-005-2631-2.

M. Pangallo, J. B.T. Sanders, T. Galla, and J. D. Farmer. Towards a taxonomy of learning dynamics in 2×2 games. *Games and Economic Behavior*, 132:1–21, 2022. ISSN 0899-8256. doi: 10.1016/j.geb.2021.11.015.

T. Roughgarden. Intrinsic Robustness of the Price of Anarchy. *J. ACM*, 62(5), nov 2015. ISSN 0004-5411. doi: 10.1145/2806883.

B. Swenson, R. Murray, and S. Kar. Regular potential games. *Games and Economic Behavior*, 124: 432–453, 2020. ISSN 0899-8256. doi: 10.1016/j.geb.2020.09.005.

E. van Damme. *Stability and Perfection of Nash Equilibria*. Springer-Verlag, Berlin, Heidelberg, 1987. ISBN 0387171010.

X. Wang and T. Sandholm. Reinforcement Learning to Play an Optimal Nash Equilibrium in Team Markov Games. In *Proceedings of the 15th International Conference on Neural Information Processing Systems*, NIPS'02, pp. 1603–1610, Cambridge, MA, USA, 2002. MIT Press.

B. Zhang and J. Hofbauer. Equilibrium selection via replicator dynamics in 2×2 coordination games. *International Journal of Game Theory*, 44(2):433–448, May 2015. ISSN 1432-1270. doi: 10.1007/s00182-014-0437-7.

SUPPLEMENTARY MATERIAL

CONVERGENCE IS NOT ENOUGH: AVERAGE-CASE PERFORMANCE OF NO-REGRET LEARNING DYNAMICS

A MISSING PROOFS AND MATERIALS: SECTION 3

A finite potential game, Γ, defined as in section 2, is called *perfectly-regular* if all its restrictions are regular, i.e., if all its restrictions have only regular Nash equilibria (cf. Definition 3.1). We will write R_Γ to denote the set of all restrictions of Γ as defined in section 2. We start by showing that the class of perfectly-regular potential games (PRPG) is well-defined, since restrictions of potential games are also potential games. Let us also recall the formal definition of regular Nash equilibria:

Definition 2.1 (Regular Nash equilibria (Harsanyi, 1973; Swenson et al., 2020))**.** *A Nash equilbrium, $x^* \in NE(\Gamma)$, is called* regular *if it is (i) quasi-strict, i.e., if for each player $k \in \mathcal{N}$, x_k^* assigns positive probability to all best responses of player k against x_{-k}^*~~all best responses of each player $k \in \mathcal{N}$ to x_{-k}^* are contained in x_k^*~~, and (ii) second-order non-degenerate, i.e., if the Hessian, $H(x^*)$, taken with respect to* $\text{supp}(x^*)$ *is non-singular.*

Lemma A.1. *A restriction of a potential game is also a potential game.*

Proof. Let Γ' a restriction of a potential game Γ and potential function $\Phi : \mathcal{A} \to \mathbb{R}$. Take $\Phi' : \mathcal{A}' \to R$ to be the restriction of Φ to $\mathcal{A}' \subseteq \mathcal{A}$. Then for all $k \in \mathcal{N}$, and $s, s' \in \mathcal{A}'$ we have that:

$$u'_k(s) - u'_k(s'_k, s_{-k}) = u_k(s) - u_k(s'_k, s_{-k}) = \Phi(s) - \Phi(s'_k, s_{-k}) = \Phi'(s) - \Phi'(s'_k, s_{-k}),$$

where the second equality follows from the definition of potential games (cf. section 2). Hence, Φ' is a potential function, and therefore Γ' a potential game. □

Using the recent results of Swenson et al. (2020), it is not difficult to show that perfectly-regular potential games are generic, and have a finite number of restricted equilibria. These are the statements of Lemma A.2 and Lemma A.3, respectively.

Lemma A.2. *Almost all finite potential games are perfectly-regular.*

Proof. Let RPG, and PRPG denote the sets of regular potential, and perfectly-regular potential games, respectively. Let also Γ be a random finite potential game. Since Γ is finite, there exist 2^m distinct restrictions of Γ, where $m := \sum_{k \in \mathcal{N}} |\mathcal{A}_k|$. Then, by Lemma A.1, we have that any restriction Γ' of Γ is also a random finite potential game, and therefore $\Pr(\Gamma' \in \text{RPG}) = 1$, with respect to the Lebesgue measure Swenson et al. (2020). It follows that:

$$\begin{aligned}
\Pr(\Gamma \in \text{PRPG}) &= 1 - \Pr(\Gamma \notin \text{PRPG}) \\
&= 1 - \Pr(\bigcup_{\Gamma' \in R_\Gamma} \Gamma' \notin \text{RPG}) \\
&\geq 1 - \sum_{\Gamma' \in R_\Gamma} \Pr(\Gamma' \notin \text{RPG}) \\
&= 1,
\end{aligned}$$

where the last equality follows from the fact that $|R_\Gamma|$ is finite. □

Lemma A.3. *Every perfectly-regular finite potential game has a finite number of restricted equilibria.*

Proof. Let Γ be a perfectly-regular finite potential game, and let Γ' be one of its restrictions. By the definition of a perfectly-regular potential game, we have that Γ' is a regular potential game. Furthermore, since Γ is finite and $\mathcal{A}'_k \subseteq \mathcal{A}_k$ for all $k \in \mathcal{N}$, it follows that Γ' is finite. But then Γ' is a finite regular potential game and as such it has a finite number of Nash equilibria, i.e., $\text{NE}(\Gamma') < \infty$

Swenson et al. (2020). Finally, since each restricted equilibrium is a Nash equilibrium of a restrictions of Γ, it follows that there exist at most:

$$\sum_{\Gamma' \in R_\Gamma} |\text{NE}(\Gamma')| \leq 2^m \max_{\Gamma' \in R_\Gamma} |\text{NE}(\Gamma')| < \infty$$

restricted equilibria of Γ. Therefore, the number of restricted equilibria of Γ is finite. □

We, now, consider the q-replicator dynamics of a finite potential game Γ, given by the dynamical system of equations in equation QRD:

$$\dot{x}_{ka_k} = x^q_{ka_k} \left(u_k(a_k, x_{-k}) - \frac{\sum_{a_j \in \mathcal{A}_k} x^q_{ka_j} u_k(a_j, x_{-k})}{\sum_{a_j \in \mathcal{A}_k} x^q_{ka_j}} \right), \quad \text{for all } k \in \mathcal{N}, a_k \in \mathcal{A}_k$$

Our goal for the remainder of this section is to prove that for any interior initial condition $x(0) \in \mathcal{X}$ the dynamics in equation QRD converge pointwise to a Nash equilibrium of a given perfectly-regular finite potential game.[5] The proof proceeds in two parts: First, we prove that the dynamics converge to a restricted equilibrium of the game for any initial condition. Second, we prove that for any *interior* initial condition, the dynamics are bound to deviate from any rest-point that is not a Nash equilibrium, and therefore, they have to converge to a Nash equilibrium.

Let us begin by proving the first of the two claims. Recall that the ω-limit set of a sequence $(x(t))_{t \geq 0} \subseteq \mathcal{X}$, that is generated by the QRD, is defined as:

$$\omega(x(t)) \coloneqq \bigcap_{t \geq 0} \text{cl}\{x(t') \mid t' > t\}$$

where $\text{cl}\, S$ denotes the closure of a set S.

Lemma A.4. *Given a perfectly-regular finite potential game Γ, every ω-limit set, with respect to the q-replicator dynamics, is a singleton $\{x^*\}$, where $x^* \in \mathcal{X}$ is a rest-point of the dynamics. Specifically, x^* is a Nash equilibrium, if $q = 0$, or a restricted equilibrium, if $q > 0$. Furthermore, the set $\mathcal{Q}(\mathcal{X}) \coloneqq \bigcup_{x_0 \in \mathcal{X}} \{x^* \in \mathcal{X} \mid \lim_{t \to \infty} x(t) = x^*,\, x(0) = x_0\}$, i.e., the set of all limit points, is finite.*

Proof. Let Γ be a perfectly-regular finite potential game. Since Γ is a potential game, by the result of Mertikopoulos & Sandholm (2018), we have that every ω-limit set consists entirely of rest-points of the dynamics. In particular, these are Nash equilibria of Γ, if $q = 0$, or restricted equilibria of Γ, if $q > 0$. However, since Γ is perfectly-regular—it suffices for it to be regular for the case of $q = 0$—it follows by Lemma A.3 that every ω-limit set is a finite set. Consider now, the ω-limit set of an orbit $(x(t))_{t \geq 0}$ of the dynamics for some arbitrary initial condition $x(0) = x_0$. Since $x(t)$ is continuous, the ω limit set $\omega(x(t))$ is the decreasing intersection of compact, connected sets and, therefore, it is connected. Since the ω-limit set is finite, the above implies that, in fact it has to be a singleton $\{x^*\}$, where x^* is a Nash equilibrium if $q = 0$ Mertikopoulos & Sandholm (2018), or a restricted equilibrium if $q > 0$, respectively. Finally, from the above, we have that the set of all limit points, $\mathcal{Q}(\mathcal{X})$ is a subset of the restricted equilibria of Γ; therefore, since Γ is a perfectly-regular finite potential game, we have, by Lemma A.3, that the set of restricted equilibria of Γ and, consequently, $\mathcal{Q}(\mathcal{X})$ are finite. □

To prove Theorem 3.2 (restated bellow, for completeness), it suffices to exclude convergence to restricted equilibria that are not NE of the original game, Γ. To establish that, we will show that as the QRD approach a limit point x^*_k, the probability x_{ka_k} of non-optimally performing actions must go to zero. Thus, all actions in $\text{supp}\, x^*_k$ must be a best response against x^*_{-k} for all agents $k \in \mathcal{N}$ which implies that x^*_k is a NE of Γ.

Theorem 3.2 (pointwise convergence of QRD to NE in PRPGs)**.** *Given any perfectly-regular potential game (PRPG), Γ, and any interior initial condition $x(0) \in \text{int}\, \mathcal{X}$, the q-replicator dynamics, defined as in equation QRD, converge pointwise to a Nash equilibrium x^* of Γ for any parameter $q \geq 0$. Furthermore, the set $\mathcal{Q}(\text{int}\, \mathcal{X}) \coloneqq \bigcup_{x_0 \in \text{int}\, \mathcal{X}} \{x^* \in \mathcal{X} \mid \lim_{t \to \infty} x(t) = x^*,\, x(0) = x_0\}$, i.e., the set of all limit points of interior initial conditions, is finite.*

[5] Recall that the *interior* of the set $\mathcal{X}$, $\text{int}\, \mathcal{X}$ is the set of all joint choice distributions $x \in \mathcal{X}$ with full support, i.e., $x_{ka_k} > 0$ for all $a_k \in \mathcal{A}_k$ and for all $k \in \mathcal{N}$; all points of $\mathcal{X}$ that are not in the interior, are called *boundary points*.

Proof. If $q = 0$, the statement follows directly from Lemma A.4. So we only need to consider the q-Replicator Dynamics for $q > 0$. Let Γ be a perfectly-regular finite potential game, and let $(x(t))_{t\geq 0}$ be a trajectory of the q-replicator dynamics with initial condition $x(0) = x_0 \in \operatorname{int}\mathcal{X}$. Since $q > 0$, we know that the support of $x(t)$ remains constant for all $t \in \mathbb{R}$ Mertikopoulos & Sandholm (2018). Thus, since $x(0) \in \operatorname{int}\mathcal{X}$, it follows that $x(t)$ remains in the interior of $\mathcal{X}$ for all $t \geq 0$, i.e., $x_{ka_k}(t) > 0$ for all $k \in \mathcal{N}$ and for all $a_k \in \mathcal{A}_k$. Furthermore, by Lemma A.4, we have that the limit $\lim_{t\to\infty} x(t)$ exists and is a rest-point of the dynamics, say x^*.

Assume that x^* is not a Nash equilibrium, i.e., that x^* fails to satisfy equation 1. That implies that there exists a player $k \in \mathcal{N}$ and an action $a_k \in \mathcal{A}_k$ with $x^*_{ka_k} = 0$, i.e., a_k is not in the support of x^*_k, but it satisfies $u_k(a_k, x^*_{-k}) > u_k(x^*)$. Note, that since $\dot{x}^*_{ka_k} = 0$ for all $a_k \in \mathcal{A}_k$, we know from equation QRD that:

$$u_k(a_k, x^*_{-k}) = \frac{\sum_{a_j\in\mathcal{A}_k}(x^*_{ka_j})^q u_k(a_j, x^*_{-k})}{\sum_{a_j\in\mathcal{A}_k}(x^*_{ka_j})^q}, \tag{7}$$

for all $a_k \in \operatorname{supp}(x^*_k)$, and thus:

$$u_k(a_k, x^*_{-k}) > u_k(x^*) = \sum_{a_j\in\mathcal{A}_k} x^*_{ka_j} u_k(a_j, x^*_{-k}) = \frac{\sum_{a_j\in\mathcal{A}_k}(x^*_{ka_j})^q u_k(a_j, x^*_{-k})}{\sum_{a_j\in\mathcal{A}_k}(x^*_{ka_j})^q}, \tag{8}$$

where the last equality follows directly from equation equation 7 and the fact that x^*_k is in the simplex, i.e., $\sum_{a_k\in\mathcal{A}_k} x^*_{ka_k} = 1$. Fix $\epsilon > 0$, and consider the set:

$$B_\epsilon := \left\{ x \in \mathcal{X} \mid u_k(a_k, x_{-k}) > \frac{\sum_{a_j\in\mathcal{A}_k} x^q_{ka_j} u_k(a_j, x_{-k})}{\sum_{a_j\in\mathcal{A}_k} x^q_{ka_j}} + \epsilon \right\}.$$

By continuity, B_ϵ is *open* and by equation equation 8, it contains x^*—given ϵ small enough. Since $x(t)$ converges to x^* as $t \to \infty$, there exists a time $t_\epsilon \geq 0$ such that $x(t) \in B_\epsilon$ for all $t > t_\epsilon$. Therefore, for each $t > t_\epsilon$ we have that:

$$\dot{x}_{ka_k}(t) = (x_{ka_k}(t))^q \left(u_k(a_k, x_{-k}) - \frac{\sum_{a_j\in\mathcal{A}_k}(x_{ka_j}(t))^q u_k(a_j, x_{-k})}{\sum_{a_j\in\mathcal{A}_k}(x_{ka_j}(t))^q} \right) > \epsilon(x_{ka_k}(t))^q > 0,$$

where the last inequality follows because $x(t) \in \operatorname{int}\mathcal{X}$ for all $t \geq 0$. Finally, by integrating with respect to time, we have that for all $t > t_\epsilon$:

$$x_{ka_k}(t) = \int_{t'=t_\epsilon}^{t} \dot{x}_{ka_k}(t')\, dt + x_{ka_k}(t_\epsilon) > x_{ka_k}(t_\epsilon) > 0.$$

Therefore, by the continuity of $x(t)$, we have that:

$$x^*_{ka_k} = \lim_{t\to\infty} x_{ka_k}(t) \geq x_{ka_k}(t_\epsilon) > 0,$$

which is a contradiction to our assumption that $x^*_{ka_k} = 0$, which is a direct consequence of x^* *not* being a NE of the game; thus, x^* has to be a Nash equilibrium of Γ. □

B Additional Visualizations and Missing Proofs: Section 4

B.1 Visualizations: invariant functions and separating manifolds

In this part, we provide systematic, and essentially exhaustive, visualizations of the stable manifolds (separatrices) in the $\Gamma_{w,\beta}$ class.

Different payoff- and risk-dominant NE in $\Gamma_{w,\beta}$. The main differences in the class $\Gamma_{w,\beta}$ occur between games in which the payoff- and risk-dominant equilibria coincide and games in which they differ. Recall that Figure 2 shows the invariant function in a $\Gamma_{w,\beta}$ instance, where $w = 2$ and $\beta = 0$, i.e., in which the payoff- and risk- dominant equilibria coincide. In Figure 6, we provide an instance in which the payoff- and risk-dominant equilibria differ. Similar to Figure 3, Figure 7 depicts the separating manifolds (stable manifolds or separatrices) of the regions of attractions of the two pure

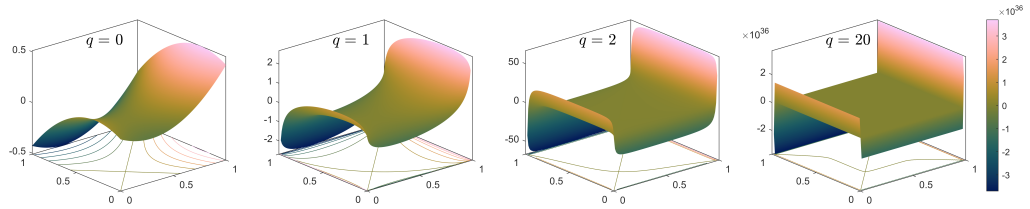

Figure 6: The invariant function, $\Psi_q(x, y)$, for all $x, y \in [0, 1]^2$ in the game $\Gamma_{w,\beta}$ for $w = 2$, $\beta = -4$, and the same values of q as in Figure 2: $q = 0$ (gradient descent dynamics), $q = 1$ (standard replicator dynamics), $q = 2$ (log-barrier dynamics), and $q = 20$. The invariant function again becomes very steep at the boundary as q increases, taking both arbitrarily large negative (**dark**) and positive (light) values in the vicinity of the NE.

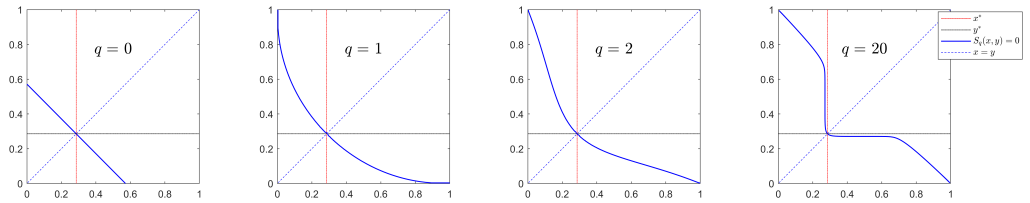

Figure 7: The stable manifolds, $\Psi_{q,\text{Stable}}(x, y) = 0$, (solid blue lines) for the same values of q and instance of $\Gamma_{w,\beta}$ as in Figure 6, in which the payoff-dominant NE is at the bottom-left corner and the risk-dominant NE is at the upper-right corner. For all q, the separatrix goes through the mixed NE at the intersection of the x^* (dashed red) and y^* (dashed black) coordinates. All panels also include the unstable manifold defined by $x - y = 0$ (dashed blue line). The region of attraction of the payoff-dominant NE is now smaller for all q, because this NE is not risk-dominant, cf. Theorem 4.4.

NE. These are precisely the zero-level sets of the invariant functions shown in Figure 6. As we can see, in this case, the region of attraction of the payoff-dominant equilibrium, w is smaller than the region of the, now, risk-dominant equilibrium, 1. Intuitively, when a NE becomes risk-dominated, its region of attraction shrinks, even if this NE is payoff-dominant. This is because, for a mixed choice of distributions, the risk-dominant NE yields a higher utility and is that more "attractive" for the dynamics. This trade-off between high reward at a certain state (e.g., w, w) and high risk if that state is not reached (e.g., $\beta, 0$, with $\beta < 0$), also explains why socially optimal, but otherwise risky, outcomes, e.g., the adoption of revolutionizing technology or a social norm that challenges the status quo, are never reached in real-life situations.

Stable manifolds for all $q \geq 0$. In a similar vein to Figure 4, we next depict the separatrices, stacked for all values of $q \in [0, 10]$, for different parameterizations of the $\Gamma_{w,\beta}$ class (Figure 8. In all panels of Figure 8, parameter w is equal to 2. We obtain qualitatively equivalent plots for any $w > 1$ and β small enough. The main takeaways from the (essentially exhaustive) visualizations in the panels of Figure 8 are that (i) the region of attraction of the risk-dominant equilibrium is larger for all $q \geq 0$ regardless of whether this equilibrium is payoff-dominant or not, (ii) the region of attraction of the payoff-dominant equilibrium may become arbitrarily small as this equilibrium becomes arbitrarily risky. In particular, observation (ii), suggests that in this case, it is hopeless to bound any static or average performance measure. This became more transparent with the APoA analysis in the previous Section of the Appendix (cf. Theorem 4.6 in the main paper). We conclude this part with some visualizations of the stable and unstable manifolds in a 2×2 non-symmetric PRPG.

Non-symmetric PRPGs. Consider the identical-interest PRPG, $\text{ID}_{w,\beta}$, with identical payoff functions $u_{w,\beta,1}(s_1, s_2) = u_{w,\beta,2}(s_1, s_2) = A_{w,\beta,s_1,s_2}$, where the payoff matrix $A_{w,\beta} \in \mathbb{R}^{2\times 2}$ is given by:

$$A_{w,\beta} = \begin{pmatrix} 1 & 0 \\ \beta & w \end{pmatrix}, \quad \beta \leq 1 \leq w.$$

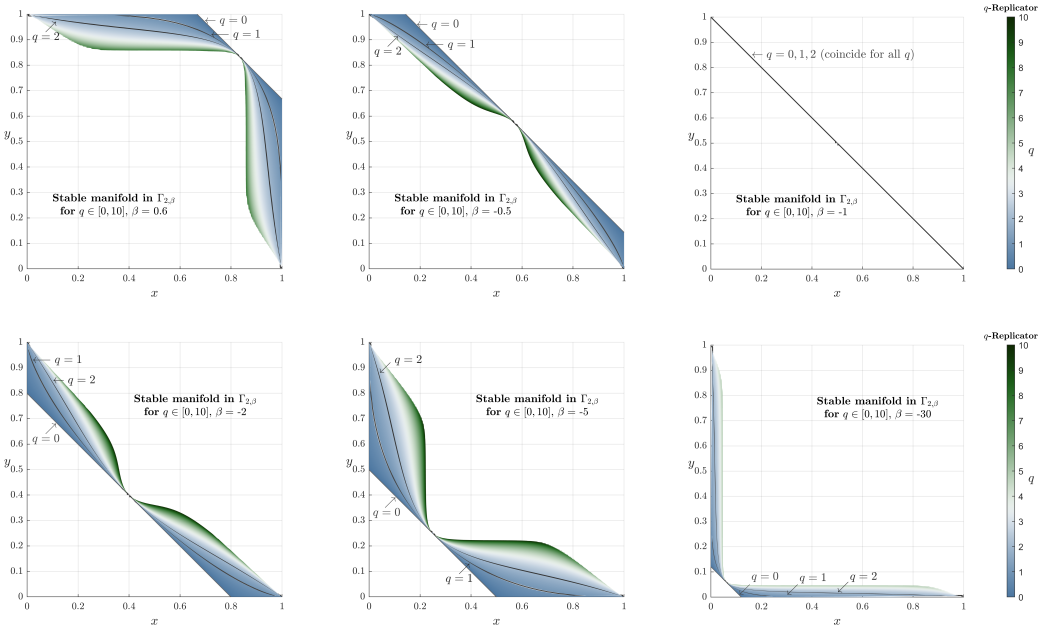

Figure 8: Stable manifolds (separatrices) for all different values of $q \in [0, 10]$ (from blue to brown) in different parameterizations of the $\Gamma_{w,\beta}$ game for $w = 2$ and varying β. In all panels, the manifolds for $q = 0$, $q = 1$, and $q = 2$ are shown in shades of black for reference. The region of attraction of the payoff-dominant equilibrium (bottom-left corner) shrinks as q increases when this equilibrium is also risk-dominant ($\alpha > 0.5$) and increases with q when this equilibrium is *not* risk-dominant ($\alpha < 0.5$). In fact, as β decreases, the payoff-dominant becomes increasingly more "risky" and its region of attraction becomes arbitrarily small.

The game $\text{ID}_{w,\beta}$ has the same pure NE as the games $\Gamma_{w,\beta}$, namely $x = y = 0$, with payoff w, and $x = y = 1$, with payoffs 1 for both players, but now the mixed NE is not symmetric and it is given by:

$$x^* = \frac{w - \beta}{w + 1 - \beta} \quad \text{and} \quad y^* = \frac{w}{w + 1 - \beta}.$$

In Figure 9, we visualize the stable *and* unstable manifolds (unlike the panels in Figure 8, where we only visualized the stable manifolds, since in that case, the unstable manifolds were always equal to the diagonal, $x = y$) for all values of $q \in [0, 10]$ in an instance of $\text{ID}_{w,\beta}$ with $w = 2$ and $\beta = -2$. In this case, the separating (stable) manifolds do not increase (decrease) monotonically with q as it is evident from the overlapping (equally) colored regions. Thus, one will require a different approach to estimate whether the size of the regions of attraction of the payoff-dominant equilibrium follow a certain monotonicity pattern, which again may change depending on whether this equilibrium is also risk-dominant or not. In the context of the current paper, Figure 9 highlights that (i) the geometry of the regions of attraction is highly complex under different algorithms (parametrizations of QRD) even for low-dimensional, identical interest games, and (ii) given this complexity, the findings in the class $\Gamma_{w,\beta}$ become even more surprising and intriguing. Extending the current analysis to further classes of games and developing potentially novel tools to address the geometry of these classes constitute straightforward, yet far-reaching, directions for future research.

B.2 Missing Materials and Proofs

We begin this section by showing that any 2×2 symmetric PRPGs is equivalent to a game $\Gamma_{w,\beta}$ as defined in subsection 4.2. The only non-generic games we are going to exclude fro the reformulation are games that are *dominance-solvable*,and therefore, their analysis is trivial and outside our scope. Recall that $\Gamma_{w,\beta}$ is a 2×2 symmetric PRPG with payoff functions $u_{w,\beta,1}(s_1, s_2) = u_{w,\beta,2}(s_2, s_1) = A_{w,\beta,s_1,s_2}$, where the payoff matrix $A_{w,\beta} \in \mathbb{R}^{2\times 2}$ is given by:

$$A_{w,\beta} = \begin{pmatrix} 1 & 0 \\ \beta & w \end{pmatrix}, \quad \beta \leq 1 \leq w.$$

For this part, it is also instructive to consult Pangallo et al. (2022), which provides a taxonomy of 2×2 games. Consider an arbitrary 2×2 symmetric PRPG with payoff with payoff functions $u_1(s_1, s_2) = u_2(s_2, s_1) = B_{s_1,s_2}$, where the payoff matrix $B \in \mathbb{R}^{2\times 2}$ is given by:

$$B = \begin{pmatrix} a & c \\ b & d \end{pmatrix}, \tag{9}$$

and set without any loss of the generality $d \geq a \geq b$—that may always be done by possibly re-indexing the agents' actions. If $c > d$, then the game is dominance-solvable (cf. Table 1 of Pangallo et al. (2022)) and as such the dynamics of the game are trivial. Hence, we may narrow our scope to payoff matrices that satisfy $d \geq c$. By our assumption that the game is a PRPG, and hence there it contains a finite number of Nash equilibria, we may exclude games where the above inequalities are *not strict*. All in all, we are going to assume without any loss of the generality that the following conditions hold:

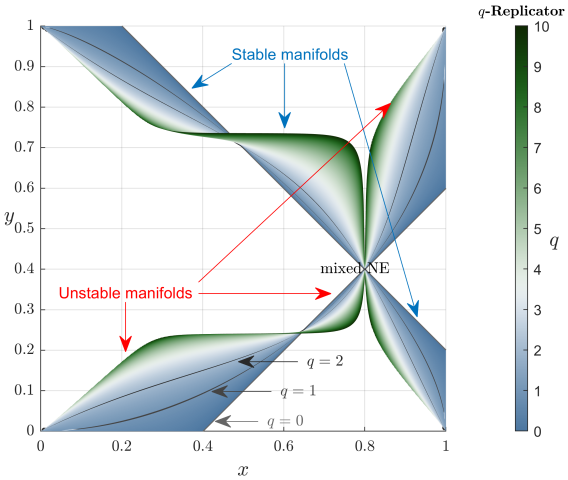

Figure 9: Stable and unstable manifolds for all $q \in [0, 10]$ in an instance of the identical interest game $\mathrm{ID}_{w,\beta}$ with $w = 2$ and $\beta = -2$. The unique mixed NE is not symmetric and lies at $x^* = 0.8$, while $y^* = 0.4$. The main difference with the symmetric games in $\Gamma_{w,\beta}$ is that the regions of attraction of the payoff-dominant equilibrium (bottom-left corner) are not increasing (nor decreasing) in q anymore.

$$d > a > b \quad \text{and} \quad d > c. \tag{10}$$

Let also $(x, 1-x)$ and $(y, 1-y)$ with $x, y \in [0, 1]$ denote the choice distributions of the two agents, adopting the common notation for the statespace of game dynamics in 2×2 games. Thus, by slightly abusing notation, the choice distributions can be conveniently represented by single variables, x and y for agents 1 and 2, respectively. Such games have three Nash equilibria: two pure at $x = y = 1$ and $x = y = 0$, with payoffs a and d, respectively, for both players, as well as one fully mixed at:

$$x^* = y^* = \frac{d-c}{a-b+d-c}, \tag{11}$$

with payoff $(x^*)^2 a + x^*(1-x^*)(b+c) + (1-x^*)^2 d$ for both players. Recall that, by definition, the Nash equilibrium $x = y = 0$ is always *payoff-dominant*—due to the possible re-indexing of the actions—and it is *risk-dominant* if $d - c > a - b$.

Lemma B.1. *Any 2×2 symmetric PRPG, Γ, with payoff functions $u_1(s_1, s_2) = u_2(s_2, s_1) = B_{s_1,s_2}$, where the payoff matrix $B \in \mathbb{R}^{2\times 2}$ is as in equation 9, can be equivalently represented by a game $\Gamma_{w,\beta}$. The game $\Gamma_{w,\beta}$ has the same NE as the original game, retains the payoff- and risk-dominance properties of its equilibrium points, and preserves the limiting behavior of any QRD.*

Proof. We begin by presenting the equations of motion of the q-replicator dynamics as functions of x and y. For the first agent that is:

$$\begin{aligned}
\dot{x} &= x^q \left(u_1(a_1, y) - \frac{x^q u_1(a_1, y) + (1-x)^q u_1(a_2, y)}{x^q + (1-x)^q} \right) \\
&= \frac{x^q (1-x)^q}{x^q + (1-x)^q} (u_1(a_1, y) - u_1(a_2, y)) \\
&= \frac{x^q (1-x)^q}{x^q + (1-x)^q} (ay + c(1-y) - by - d(1-y)) \\
&= \frac{x^q (1-x)^q}{x^q + (1-x)^q} [(a - b + d - c)y - (d - c)] \\
&= \frac{x^q (1-x)^q}{x^q + (1-x)^q} \kappa \cdot (y - y^*),
\end{aligned} \tag{12}$$

where $\kappa := a - b + d - c$. Similarly, we may derive the equation of motions for the second agent as $\dot{y} = \frac{y^q(1-y)^q}{y^q+(1-y)^q}\kappa \cdot (x - x^*)$. Here, (x^*, y^*) is the mixed Nash equilibrium of the game as given in equation 11, and holds $y^* = x^*$. Thus, apart from the variables x and y and the hyperparameter q which is exogenously given, the q-replicator dynamics depend on the payoffs of the game Γ only through κ and $x^* = \frac{d-b}{\kappa}$. It follows that any transformation that preserves the value of x^* and scales κ by a constant may only scale $\dot{x}$ and $\dot{y}$ by the same constant; that is, may only affect the convergence rate of the dyanmics, but *not* their limiting behavior. Starting from an arbitrary payoff matrix B as given in equation 9 let us assume, without any loss of the generality, that the conditions in equation 10 apply. Notice that, since by the aforementioned assumptions we have that $d > c$, we may set some $\delta \in \mathbb{R}$ such that $0 < a + \delta < d - c$. Accordingly, we consider the following sequence of transformations: (T1) Add δ to the first column, (T2) subtract c from the second column, and (T3) divide by $a + \delta$. These lead to:

$$B = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \xrightarrow{\text{(T1)}} \begin{pmatrix} a+\delta & c \\ b+\delta & d \end{pmatrix} \xrightarrow{\text{(T2)}} \begin{pmatrix} a+\delta & 0 \\ b+\delta & d-c \end{pmatrix} \xrightarrow{\text{(T3)}} \begin{pmatrix} 1 & 0 \\ \frac{b+\delta}{a+\delta} & \frac{d-c}{a+\delta} \end{pmatrix} =: A_{\frac{d-c}{a+\delta}, \frac{b+\delta}{a+\delta}}.$$

Notice that $A_{\frac{d-c}{a+\delta}, \frac{b+\delta}{a+\delta}}$ is the payoff matrix of a parametric game $\Gamma_{w,\beta}$, where $w := \frac{d-c}{a+\delta}$ and $\beta := \frac{b+\delta}{a+\delta}$. Observe, that (T1), (T2) and (T3) leave x^* unaltered and only scale κ by a constant $\frac{1}{a+\delta}$; that is, the limiting behavior of the q-replicator dynamics is preserved. Furthermore, the risk-dominance of the equilibrium points is preserved, because:

$$d - c > a - b \quad \text{if and only if} \quad \frac{d-c}{a+\delta} > 1 - \frac{b+\delta}{a+\delta}$$

Finally, the payoff-dominance of the Nash equilibrium $x = y = 0$ is also preserved because, by the definition of δ, we have that $\frac{d-c}{a+\delta} > 1$. □

Next, we are going to construct the invariant functions of $\Gamma_{w,\beta}$ with respect to the q-replicator dynamics, which are given by Lemma 4.5 that we restate below for completeness. Recall that (α, α) is defined to be the equilibrium point of the a game $\Gamma_{w,\beta}$; that is, $x^* = y^* = \alpha$.

Lemma 4.5 (Invariant functions of QRD in 2×2 symmetric PRPGs)**.** *Given a 2×2 symmetric PRPG, $\Gamma_{w,\beta}$, whose agents evolve with respect to the q-replicator dynamics, the separable function $\Psi_q : (0,1)^2 \to \mathbb{R}$ with $\Psi_q(x,y) := \psi_q(x) - \psi_q(y)$, where $\psi_q : (0,1) \to \mathbb{R}$ is given by:*

$$\psi_q(x) = \begin{cases} \dfrac{x^{2-q} + (1-x)^{2-q} - 1}{2-q} + \dfrac{1 - \alpha x^{1-q} - (1-\alpha)(1-x)^{1-q}}{1-q}, & q \neq 1, 2, \\ \alpha \ln(x) + (1-\alpha)\ln(1-x), & q = 1, \\ \ln(x) + \ln(1-x) + \dfrac{\alpha}{x} + \dfrac{1-\alpha}{1-x}, & q = 2, \end{cases} \tag{6}$$

remains constant along any trajectory $\{x(t), y(t)\}_{t \geq 0}$ of the system. The function $\Psi_q(x)$ is continuous with respect to the parameter q at, both, $q = 1$ and $q = 2$, since $\lim_{q \to 1} \Psi_q(x) = \Psi_1(x)$ and $\lim_{q \to 2} \Psi_q(x) = \Psi_2(x)$ for all $x \in (0,1)$.

Proof. To prove the statement, we will show that the time derivative of $\Psi_q(x(t), y(t))$ is equal to zero. Let us begin by constructing the derivative of $\psi(x)$. For $q \neq 1, 2$ we have that:

$$\psi_q'(x) = \frac{x}{x^q} - \frac{1-x}{(1-x)^q} - \frac{\alpha}{x^q} + \frac{1-\alpha}{(1-x)^q} = \frac{x-\alpha}{x^q} + \frac{x-\alpha}{(1-x)^q} = \frac{(x-\alpha)[(1-x)^q + x^q]}{x^q(1-x)^q}.$$

Similarly, for $q = 1$ we have that:

$$\psi_1'(x) = \frac{\alpha}{x} - \frac{1-\alpha}{1-x} = \frac{\alpha - x}{x(1-x)},$$

and, for $q = 2$, we have that:

$$\psi_2'(x) = \frac{1}{x} - \frac{1}{1-x} - \frac{\alpha}{x^2} + \frac{1-\alpha}{(1-x)^2} = \frac{(x-\alpha)[x^2 + (1-x)^2]}{x^2(1-x)^2}.$$

That is the derivative of $\psi(x)$ has the general form:

$$\psi_q'(x) = \lambda \cdot \frac{(x-\alpha)[(1-x)^q + x^q]}{x^q(1-x)^q},$$

for all $q \geq 0$, where $\lambda \in \{1, -1\}$. Notice, that the choice for λ is purely stylistic because the invariance of a function is not affected by scalar transformations. Using equation 12 from the proof of Lemma B.1 we have that:

$$\begin{aligned}\dot{\Psi}_q(x,y) &= \frac{\partial \Psi_q(x,y)}{\partial x}\dot{x} - \frac{\partial \Psi_q(x,y)}{\partial y}\dot{y} \\ &= \psi_q'(x)\dot{x} - \psi_q'(y)\dot{y} \\ &= \lambda \cdot \kappa \cdot [(x-\alpha)(y-\alpha) - (y-\alpha)(x-\alpha)] = 0\end{aligned}$$

□

Before we proceed with the proof of the main theorem of this section (Theorem 4.4), we need to provide the formal definition of the stable and unstable manifolds of the mixed NE of $\Gamma_{w,\beta}$.

Definition B.2 (Stable and unstable manifolds of $\Gamma_{w,\beta}$under QRD)**.** *Let $\Psi_q : [0,1]^2 \to \mathbb{R}$ with $\Psi_q(x,y) = \Psi_{q,\text{Stable}}(x,y) \cdot (x-y)$ for all $x, y \in [0,1]$ denote the invariant function of the q-replicator dynamics for the 2×2 symmetric PRPG, $\Gamma_{w,\beta}$. The* unstable manifold *of the mixed NE $(x,y) = (\alpha,\alpha)$ under the q-replicator dynamics is the curve $\mathcal{M}_{\text{Unstable}} := \{(x,y) \in (0,1)^2 \mid x = y\}$; that is, the set of points for which $\lim_{t\to-\infty} x(t) = \lim_{t\to-\infty} y(t) = \alpha$. Analogously, the* stable manifold *of the mixed NE is the curve $\mathcal{M}_{\text{Stable}} := \{(x,y) \in (0,1)^2 \mid \Psi_{q,\text{Stable}}(x,y) = 0\}$; that is, the set of points for which $\lim_{t\to\infty} x(t) = \lim_{t\to\infty} y(t) = \alpha$.*

For the rest of this section, we are going need to have at hand a n explicit form of the stable manifold $\mathcal{M}_{\text{Stable}}$ of the mixed NE of $\Gamma_{w,\beta}$ with respect to the 0-replicator dynamics, or GD. In that regard, we now going to construct that manifold. Recall that from equation 4.5 the invariant function of any QRD in $\Gamma_{w,\beta}$ is given by $\Psi_0(x,y) := \psi_0(x) - \psi_0(y) = 0$, where:

$$\psi_0(x) = \frac{x^2 + (1-x)^2 - 1}{2} + 1 - \alpha x - (1-\alpha)(1-x) = x^2 - 2\alpha x + \alpha.$$

Therefore, $\Psi_0(x,y) = x^2 - 2\alpha x + \alpha - y^2 - 2\alpha y + \alpha = (x-y)(x+y-2\alpha) = 0$ and the stable manifold (cf. subsection 4.2) satisfies $\Psi_{q,\text{Stable}}(x,y) = x + y - 2\alpha = 0$. In other words, the stable manifold is the line segment:

$$\mathcal{M}_{\text{Stable}} = \{(x,y) \in (0,1)^2 \mid y = 2\alpha - x\}. \tag{13}$$

Now, let us proceed with the proof of the main theorem.

Theorem 4.4 (Performance of QRD in symmetric 2×2 PRPG)**.** *Given any 2×2 symmetric PRPG, which, without any loss of generality, can be represented as an instance $\Gamma_{w,\beta}$, it holds that*

$$APM_{\text{SW,int}\,\mathcal{X}}(V_0, \Gamma_{w,\beta}) \geq APM_{\text{SW,int}\,\mathcal{X}}(V_1, \Gamma_{w,\beta}) \tag{5}$$

if and only if ~~whenever~~ the payoff-dominant equilibrium is also risk-dominant, with equality only when ~~if and only if~~ $\alpha = 0.5$, i.e., $w = 1 - \beta$, where V_0, V_1 are the equations of motion of the 0-replicator and 1-replicator dynamics, respectively equation QRD.

Proof. Let us first recall that, by equation equation 13 and Definition B.2, the stable and unstable manifolds of gradient descent (GD) is given by the lines $y = 2\alpha - x$ and $y = x$, respectively, where $\alpha = \frac{w}{w+1-\beta}$, while the stable and unstable manifolds of the standard replicator dynamics are given as solutions to:

$$\Psi_1(x,y) = \psi_1(x) - \psi_1(y) = \alpha \ln\left(\frac{x}{y}\right) + (1-\alpha)\ln\left(\frac{1-x}{1-y}\right) = 0,$$

where the line $y = x$ corresponds to a solution. We are going to to prove that the *single* remaining solution of the previous equation, although it cannot be expressed explicitly, satisfies $y \leq 2\alpha - x$, if

$\alpha > \frac{1}{2}$, $y \geq 2\alpha - x$, if $\alpha < \frac{1}{2}$ (with equality in both cases only if $x = \alpha$), and $y = 2\alpha - x$ otherwise; hence, the statements of Theorem 4.4 follow naturally.

It is not difficult to show that $\psi_1'(x) = \frac{\alpha - x}{x(1-x)}$ (cf. proof of Lemma 4.5), and $\psi_1''(x) = -\frac{x^2 - 2\alpha x + \alpha}{x^2(1-x)^2} < 0$. Therefore, ψ_1 is a strictly concave function with maximum at $x = \alpha$. To proceed, it will be useful to define the implicit function $y : (0,1) \to (0,1)$ such that $y(\alpha) = \alpha$, and $\forall x \in (0,1) \setminus \{\alpha\}$: $y(x) \neq \alpha$, and $\psi_1(y(x)) = \psi_1(x)$. By applying the Intermediate Value Theorem (IVT) on ψ_1 in the intervals $(0, \alpha)$, and $(\alpha, 1)$, we can verify that y is a well-defined bijective function. Note that $y = y(x)$ has to correspond to the remaining solution of $\Psi_1(x, y)$.

Without any loss of the generality, let us consider the case of $\alpha > \frac{1}{2}$. Since ψ_1' is strictly decreasing in $(0,1)$ ($\psi_1'' < 0$), we have that $\psi_1'(x) > 0$ for all $x \in (0, \alpha)$, and $\psi_1'(x) < 0$ for all $x \in (\alpha, 1)$. We begin by proving that for all $x \in (0, 1-\alpha)$ it holds that $|\psi_1'(\alpha - x)| < |\psi_1'(\alpha + x)|$, i.e., $\psi_1'(\alpha - x) < -\psi_1'(\alpha + x)$. Specifically, we have the following series of equivalences:

$$
\begin{aligned}
\psi_1'(\alpha - x) < -\psi_1'(\alpha + x) &\iff \frac{x}{(\alpha - x)(1 - \alpha + x)} < \frac{x}{(\alpha - x)(1 - \alpha - x)} \\
&\iff x(\alpha - x)(1 - \alpha - x) < x(\alpha - x)(1 - \alpha + x) \\
&\iff 2x^2(1 - 2\alpha) < 0 \\
&\iff \alpha > \frac{1}{2},
\end{aligned}
$$

which holds by assumption. Next, by taking advantage of the above, we can prove that, for all $x \in (0, 1-\alpha)$, it holds $\psi_1(\alpha - x) > \psi_1(\alpha + x)$; that is:

$$
\begin{aligned}
\psi_1(\alpha - x) &= \int_0^{\alpha - x} \psi_1'(t)\, dt \\
&= \int_0^{\alpha} \psi_1'(t)\, dt + \int_{\alpha - x}^{\alpha} -\psi_1'(t)\, dt \\
&> \int_0^{\alpha} \psi_1'(t)\, dt + \int_{\alpha}^{\alpha + x} \psi_1'(t)\, dt \\
&= \psi_1(\alpha + x)
\end{aligned}
$$

Furthermore, since ψ_1 is monotonically decreasing in $(\alpha, 1)$ ($\psi_1' < 0$ in $(\alpha, 1)$), we have that, for all $x \in (0, 1-\alpha)$, and for all $t \in [\alpha + x, 1)$, it holds $\psi_1(\alpha - x) > \psi_1(t)$. However, by the IVT, we have that there exists $t^* \in (\alpha, 1)$ such that $\psi_1(\alpha - x) = \psi_1(t^*)$. Therefore, it must hold that $t^* \in (\alpha, \alpha + x)$. Then, since y is a bijective, we have that $y(\alpha - x) = t^* < \alpha + x$, i.e., for all $x \in (2\alpha - 1, \alpha)$ we have that $y(x) < 2\alpha - x$. Similarly, we have that $y(\alpha + x) < \alpha - x$, i.e., for all $x \in (\alpha, 1)$, it holds that $y(x) < 2\alpha - x$. Finally, notice that for all $x \in (0, 2\alpha - 1)$, we trivially have that $2\alpha - x > 1 > y(x)$; therefore, for all $x \in (0,1) \setminus \{\alpha\}$: $y(x) < 2\alpha - x$. We remark that the case of $\alpha < \frac{1}{2}$ follows identical arguments, while the case $\alpha = \frac{1}{2}$ is trivial. □

Technically, a direct implication of the proof of Theorem 4.4 is provided in Lemma B.3 which may be of independent interest. Recall from the proof of Theorem 4.4 that the solutions to $\Psi_1(x, y) = 0$ are the functions $y = x$ and $y : (0,1) \to (0,1)$ such that $y(\alpha) = \alpha$, and for all $x \in (0,1) \setminus \{\alpha\}$ it holds that $y(x) \neq \alpha$, and $\psi_1(y(x)) = \psi_1(x)$.

Lemma B.3 (Curvature of the stable manifold of RD)**.** *Consider the* 1*-replicator dynamics (RD) in the parametric game $\Gamma_{w,\beta}$. The stable manifold, $\mathcal{M}_{\text{Stable}}$ of RD in $\Gamma_{w,\beta}$ is given by the curve $y = y(x)$. If the payoff-dominant equilibrium, $x = y = 0$, is also risk-dominant, then y is strictly concave. Conversely, if the non-payoff dominant equilibrium, $x = y = 1$, is risk-dominant, then y is strictly convex. Otherwise, $y(x) = 1 - x$.*

Proof. By differentiating both sides of the implicit function $\psi_1(y(x)) = \psi_1(x)$ with respect to x, we get that $\psi_1'(y(x))y'(x) = \psi_1'(x)$, i.e., $y'(x) = \frac{\psi_1'(x)}{\psi_1'(y(x))}$. Notice that, since y is bijective, ψ_1' is monotonic ($\psi_1'' < 0$), and $\psi_1'(y(\alpha)) = \psi_1'(\alpha) = 0$, the above equality is well-defined for all

$x \in (0,1) \setminus \{\alpha\}$. Hence, we have that:

$$\begin{aligned} y'' &= \frac{\psi_1''(x)[\psi_1'(y)]^2 - \psi_1'(x)\psi_1''(y)y'}{[\psi_1'(y)]^2} \\ &= \frac{\psi_1''(x)[\psi_1'(y)]^2 - [\psi_1'(x)]^2\psi_1''(y)}{[\psi_1'(y)]^3} \\ &= \frac{y^3(1-y)^3\alpha(1-\alpha)(x-y)(y+x-2\alpha)}{(\alpha-y)^3x^2(1-x)^2y^2(1-y)^2}, \end{aligned}$$

where the dependency of y to x is implied for compactness. Hence, we have that $y''(x) < 0$ if and only if $\frac{(x-y)(y+x-2\alpha)}{\alpha-y} < 0$. However, by the Intermediate Value Theorem (IVT) applied on ψ_1 in $(0,\alpha)$, and $(\alpha,1)$, and the definition of y, it follows, trivially, that $x \leq y$ if, and only if, $\alpha \leq y$, with equality in both inequalities only if $x = \alpha$. Therefore, $\frac{x-y}{\alpha-y} > 0$, $\forall x \in (0,1) \setminus \{\alpha\}$; hence $y''(x) < 0$ if, and only if, $y + x - 2\alpha < 0$, which by the proof of Theorem 4.4 is equivalent to $\alpha > \frac{1}{2}$. This concludes the proof for the first statement. The second statement follows in a similar manner by requesting $y''(x) > 0$, while the last statement is trivial. □

Application: APoA in 2×2 PRPGs. We conclude this section by providing the proof of Theorem 4.6, which we restate below for convenience. Recall that from this point forward, we focus on symmetric 2×2 PRPGs such that payoff- and risk-dominant equilibria coincide, as in such settings, one can prove particularly strong, tight bounds on ApoA. This showcases the practical importance of Theorem 4.4 and the invariant function approach.

Theorem 4.6. *The APoA of GD dynamics in all 2×2 symmetric PRPGs, $\Gamma_{w,\beta}$, is bounded by 2, i.e., APoA$(V_0, \Gamma_{w,\beta}) \leq 2$. Furthermore, this bound is tight.*

Proof. Let $\Gamma_{w,\beta}$ be a 2×2 symmetric PRPG, where the payoff-dominant equilibrium, $x = y = 0$, is also risk-dominant, i.e., $\beta > 1 - w$, or equivalently $\alpha > 0.5$, where $x^* = y^* = \alpha$ is the mixed NE of $\Gamma_{w,\beta}$. Recall that, by equation 13, the stable manifold of the mixed NE of $\Gamma_{w,\beta}$ with respect to GD is the line segment:

$$\ell : y = 2\alpha - x \quad \text{for } x \in (\max\{0, 2\alpha - 1\}, \min\{2\alpha, 1\}).$$

Since $\alpha > 0.5$, we have that $2\alpha - 1 \geq 0$ and $2\alpha \geq 1$; therefore, the extreme points of ℓ are $(0, 2\alpha)$ and $(2\alpha, 0)$. That implies the the RoA of $(1,1)$ is the triangle with extreme points at $(1,1)$, $(2\alpha - 1, 1)$, and $(1, 2\alpha - 1)$. Since that is a right triangle, with both its base and its height equal to $2(1-\alpha)$, the Lebesque measure of RoA$(0,0)$ is $\mu(\text{RoA}(1,1)) = 2(1-\alpha)^2$; subsequently, $\mu(\text{RoA}(0,)) = 1 - \mu(\text{RoA}(1,1)) = 1 - 2(1-\alpha)^2$. We may calculate the APoA of GD in $\Gamma_{w,\beta}$ as a function of w and β. Specifically, when $\alpha > 0.5$, i.e., $\beta < 1 - w$, we have that:

$$\begin{aligned} \text{APoA}(w,\beta) &\coloneqq \text{APoA}(\text{GD}, \Gamma_{w,\beta}) \\ &= \frac{\max_{x,y\in[0,1]} \text{SW}(x,y)}{\text{APM}_{\text{SW},[0,1]^2}(\text{GD}, \Gamma_{w,\beta})} \\ &= \frac{\text{SW}(0,0)}{\text{SW}(0,0)\cdot\mu(\text{RoA}(0,0)) + \text{SW}(1,1)\cdot\mu(\text{RoA}(1,1))} \\ &= \frac{w}{w\mu(\text{RoA}(0,0)) + \mu(\text{RoA}(1,1))} \\ &= \frac{w}{w[1-2(1-\alpha)^2] + 2(1-\alpha)^2} \\ &= \frac{w}{w\left[1 - 2\left(\frac{1-\beta}{w+1-\beta}\right)^2\right] + 2\left(\frac{1-\beta}{w+1-\beta}\right)^2} \\ &= \frac{w(w+1-\beta)^2}{w(w+1-\beta)^2 - 2(w-1)(1-\beta)^2}. \end{aligned}$$

We may, now, perform a first-order analysis in APoA(w,β); that is, for all $\beta \geq 1 - w$, we have that:

$$\frac{\partial \text{APoA}(w,\beta)}{\partial \beta} = \frac{-4w^2(w-1)(w+1-\beta)(1-\beta)}{[w(w+1-\beta)^2 - 2(w-1)(1-\beta)^2]^2} \leq 0.$$

From the above, it follows that $\text{APoA}(w, \beta) \leq \text{APoA}(w, 1-w)$; that is:

$$\text{APoA}(w, \beta) \leq \frac{4w^3}{4w^3 - 2(w-1)w^2} = \frac{2w^3}{w^3 + w^2} < 2,$$

where the last inequality follows by letting $w \to \infty$. Notice that this bound is tight. It is not difficult to see that if $\alpha < 0.5$, $\text{APoA}(w, \beta)$ is unbounded. □

C Numerical Experiments Beyond 2×2 PRPGs

In this part, we present results from simulations of the q-replicator dynamics in PRPGs of higher dimensions, i.e., beyond the 2×2 setting. The purpose of the current simulations is twofold. First, we want to test whether the theoretical prediction of Theorem 4.4 extends to larger 2-agent, symmetric games (games in which each of the 2 agents has more than two actions), i.e., whether gradient descent dynamics have better/worse average performance than the standard replicator dynamics when payoff-dominance and risk-dominance coincide/differ. Second, we want to test the theoretical prediction of Theorem 4.6, i.e., whether the Average Price of Anarchy (APoA) remains bounded by 2 (or if not, whether it changes according to some pattern that depends on the size of the input) in larger 2-agent, symmetric PRPGs in which payoff-dominance and risk-dominance coincide.

Recall that Theorem 3.2, i.e., pointwise convergence of QRD to NEs, holds for *arbitrary* PRPGs which means that such experiments are possible (and meaningful) in the first place. However, the *game-theoretic* interpretation of the results may not be as straightforward in larger dimensions as it was in the 2×2 case. The reason is that the notion of risk-dominance does not admit a straightforward rigorous generalization to arbitrary games. However, in the most natural class of games in which equilibrium selection is typically studied, risk-dominance can still be defined in an intuitive way Leonardos & Piliouras (2022). This is the class of *diagonal games* and their perturbations. A 2-agent, symmetric, diagonal game can be described by a payoff matrix, U, that has non-zero (and in our case, non-negative) elements only on the diagonal, i.e., it is of the form,

$$U = \begin{pmatrix} u_1 & 0 & \dots & 0 \\ 0 & u_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_n \end{pmatrix} := \text{diag}(u_1, u_2, \dots, u_n)$$

with $0 < u_1 < u_2 < \dots < u_n$. In these games, u_n is the payoff of the payoff-dominant equilibrium for each player. Any failure to coordinate at a pure equilibrium (point on the diagonal) results in a payoff of zero. Thus, all (pure) equilibria are in a sense *equally* risky.

C.1 Experimental Setup

We run experiments in random 2-agent, symmetric diagonal PRPGs (D-PRPGs) of dimensions $n = 2, 3, \dots, 20$ (size of each agent's action space). In each game, the payoffs $u_1, u_2, \dots, u_n$ are selected (pseudo-)randomly and satisfy the following properties: (i) the lowest diagonal payoff, u_1, is at least as large as some predefined positive constant (set equal to $1e-12$ for the experiments), (ii) the highest (diagonal) payoff, u_n, is equal to the dimension, n, of the game, i.e., $u_n = 2, 3, \dots$ and 20 respectively, and (iii) $u_2, \dots, u_{n-1}$ are in ascending order strictly between u_1 and u_n with randomly selected distances between them. For each dimension, we sample 100 random games and run the gradient descent and standard replicator dynamics for 1000 initial conditions till convergence.

C.2 Numerical Results

The outputs of the simulations of the above experiments are summarized in Figure 10. Four instances, for $n = 3, 5, 10$ and 20 are described for reference in Table 1.

These outputs provide indications for the following:

1. The gradient descent dynamics (continue to) outperform the replicator dynamics in all diagonal games in terms of average performance. The result holds not only for the aggregate average metrics reported in Figure 10 and Table 1, but also for each individual game that was sampled.

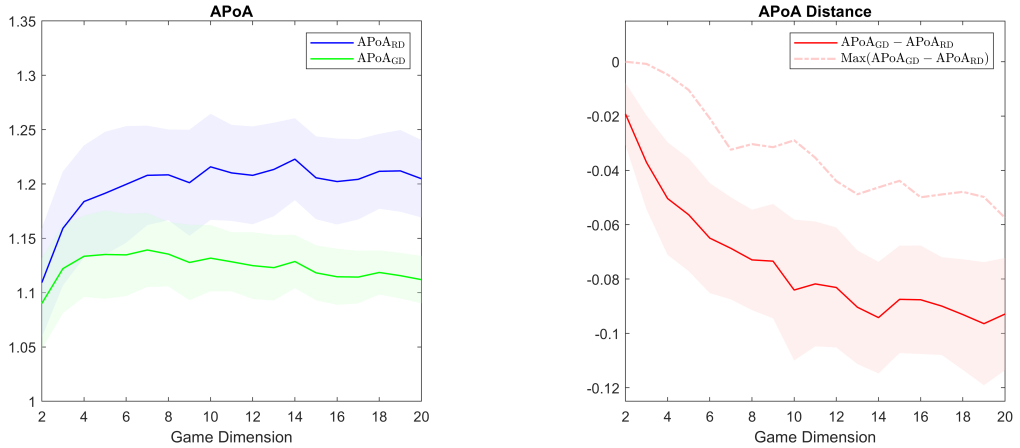

Figure 10: Numerical results regarding the APoA metric of replicator dynamics (RD) and gradient descent (RD) in diagonal PRPGs of dimensions $n = 2, 3, \ldots, 20$. The left panel shows the APoA of each dynamic together with its standard deviation over 100 randomly sampled games at each dimension. The panel on the right shows their difference (solid line enveloped by the standard deviation shaded region) and the difference of the maximum APoA of the dynamics observed over all the 100 sampled games at each dimension. The GD dynamics throughout outperform the RD dynamics suggesting that an extension of Theorem 4.4 may be possible to larger dimensional D-PRPGs.

	APoA $\pm$ std		Maximum APoA	
	RD	GD	RD	GD
$n = 3$	1.159 ± 0.052	1.122 ± 0.041	1.234	1.190
$n = 5$	1.191 ± 0.057	1.135 ± 0.040	1.300	1.223
$n = 10$	1.216 ± 0.049	1.131 ± 0.030	1.377	1.202
$n = 20$	1.205 ± 0.036	1.112 ± 0.022	1.284	1.215

Table 1: Numerical results regarding the APoA metric of replicator dynamics (RD) and gradient descent (GD) in diagonal PRPGs of dimension $n = 3, 5, 10$ and 20. The second column reports the average APoA and its standard deviation over 100 random diagonal games for each dimension and the last column reports the maximum APoA observed in all these instances. Note that the reported maximum APoA should only be interpreted as a rough indication of the actual maximum APoA (i.e., of the maximum APoA in the whole population of games of a specific dimension). Accurately estimating the latter requires a larger sample of games (currently equal to 100 for each dimension).

This provides evidence that the result of Theorem 4.4 may extend to larger dimensions, at least in the case of diagonal games.

2. The APoA never exceeded the theoretical bound of 2 (and in fact, it was much lower than that as the Maximum APoA column suggests) in all sampled games. This indicates that the theoretical bound of 2 for the gradient descent dynamics (cf. Theorem 4.6) possibly extends to larger dimensions as well. Moreover, this bound still holds numerically for the replicator dynamics (as was the case when $n = 2$).

Summing up, the experiments in the diagonal games provide preliminary evidence that the theoretical results of Theorem 4.4 and Theorem 4.6 that were established in the case of 2×2 games, seem to scale well also in larger dimensions. These results provide a promising starting point for the analysis of average performance measures for both larger classes of games (including arbitrary PRPGs of larger dimensions) and larger classes of dynamics (with arbitrary regularizers) and, intriguingly, suggest that our rigorous guarantees in low-dimensional games may admit similar counterparts even in higher/arbitrary dimensional games, at least of certain structure.

On Risk-dominance in Higher Dimensions In diagonal games (D-PRPGs), it is straightforward to make an equilibrium more *risky*. This is achieved simply by replacing the zero entries in the corresponding line of matrix U by some negative number, e.g.,

$$U_{\text{risky},n} = \begin{pmatrix} u_1 & 0 & \dots & 0 \\ 0 & u_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -10 & -10 & \dots & u_n \end{pmatrix} := \text{diag}(u_1, u_2, \dots, u_n; \text{risk}_n = -10)$$

In this case, the payoff-dominant equilibrium with payoff u_n to each player becomes more risky, since a failure to coordinate on it results to a negative payoff of -10 for the agent who selected the corresponding action. Proceeding in a similar fashion, one may replace the zero entries with an (arbitrarily large) negative element in all lines of the matrix except for the first one. In analogy to the $U_{\text{risky},n}$ notation, we will denote such games by $U_{\text{risky}} := \text{diag}(u_1, u_2, \dots, u_n; \text{risk} = -r)$, where $r > 0$ is the *risk constant* (equal to -10 in the example above). In this way, all equilibria become more risky except for the payoff-dominated one, i.e., the equilibrium with payoffs u_1 to each agent which corresponds to the action profile in which every agent selects their first action.[6]

Concerning our experiments, the outcome of Figure 10 (see also Table 1) is reversed in games of the form U_{risk}, i.e., in games in which the payoff superior equilibria were more risky (not reported here).

[6]An alternative interesting approach to generalize the notion of risk-dominance in arbitrary games is via pairwise comparisons of actions as proposed by Honda (2012).