games

MDPI

Article

Bifurcation Mechanism Design—From Optimal Flat Taxes to Better Cancer Treatments

Ger Yang [1], David Basanta [2] and Georgios Piliouras [3,*]

[1] Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78705, USA; geryang@utexas.edu
[2] Integrated Mathematical Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA; david@cancerevo.org
[3] Engineering Systems and Design (ESD), Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372, Singapore
* Correspondence: georgios@sutd.edu.sg

Received: 25 February 2018; Accepted: 18 April 2018; Published: 26 April 2018

Abstract: Small changes to the parameters of a system can lead to abrupt qualitative changes of its behavior, a phenomenon known as bifurcation. Such instabilities are typically considered problematic, however, we show that their power can be leveraged to design novel types of mechanisms. *Hysteresis mechanisms* use transient changes of system parameters to induce a permanent improvement to its performance via optimal equilibrium selection. *Optimal control mechanisms* induce convergence to states whose performance is better than even the best equilibrium. We apply these mechanisms in two different settings that illustrate the versatility of bifurcation mechanism design. In the first one we explore how introducing flat taxation could improve social welfare, despite decreasing agent "rationality," by destabilizing inefficient equilibria. From there we move on to consider a well known game of tumor metabolism and use our approach to derive potential new cancer treatment strategies.

Keywords: game theory; cancer; economics; hysteresis

1. Introduction

The term bifurcation, which means splitting in two, is used to describe abrupt qualitative changes in system behavior due to smooth variation of its parameters. Bifurcations are ubiquitous and permeate all natural phenomena. Effectively, they produce discrete events (e.g., rain breaking out) out of smoothly varying, continuous systems (e.g., small changes to humidity or temperature). Typically, they are studied through bifurcation diagrams, multi-valued maps that prescribe how each parameter configuration translates to possible system behaviors (e.g., Figure 1).

Bifurcations arise in a natural way in game theory. Games are typically studied through their Nash correspondences, a multi-valued map connecting the parameters of the game (i.e., payoff matrices) to system behavior, in this case Nash equilibria. As we slowly vary the parameters of the game, typically the Nash equilibria will also vary smoothly, except at bifurcation points where, for example, the number of equilibria abruptly changes as some equilibria appear/disappear altogether. Such singularities may substantially impact both system behavior and system performance. For example, if the system state was at an equilibrium that disappeared during the bifurcation, then a turbulent transitionary period ensues where the system tries to reorganize itself at one of the remaining equilibria. Moreover, the quality of all remaining equilibria may be significantly worse than the original. Even more disturbingly, it is not a priori clear that the system will equilibrate at all. Successive bifurcations that lead to increasingly more complicated recurrent behavior is a standard route to chaos [1], which may have devastating effects on system performance.

Game theorists are particularly aware of the need to produce "robust" predictions, i.e., predictions that allow for deviations from an idealized exact specification of the parameters of the setting [2]. For example, ϵ-approximate Nash equilibria allow for the possibility of computational bounded agents, whereas ϵ-regret outcomes allow for persistently non-equilibrating behavior [3]. These approaches, however, do not really address the problem at its core as any solution concept defines a map from parameter space to behavioral space and no such map is immune to bifurcations. If pushed hard enough any system will destabilize. The question is what happens next?

Well, a lot of things *may* happen. It is intuitively clear that if we are allowed to play around arbitrarily with the payoffs of the agents then we can reproduce any game and no meaningful analysis is possible. Using payoff entries as controlling parameters is problematic for another reason. It is not clear that there exists a compelling parametrization of the payoff space that captures how real life decision-makers deviate from the Platonic ideal of the payoff matrix. Instead, we focus on another popular aspect of economic theory: agent "rationality".

We adopt a standard model of boundedly rational learning agents. Boltzmann Q-learning dynamics [4–6] is a well studied behavioral model in which agents are parameterized by a temperature/rationality term T. Each agent keeps track of the collective past performance of his/her actions (i.e., learns from experience) and chooses an action according to a Boltzmann/Gibbs distribution with parameter T. When applied to a multi-agent game, the behavioral fixed points of Q-learning are known as quantal response equilibria (QREs) [7]. Naturally, QREs depend on the temperature T. As $T \to 0$ players become perfectly rational, and play approaches a Nash equilibrium,[1] whereas as $T \to \infty$ all agents use uniformly random strategies. As we vary the temperature the QRE(T) correspondence moves between these two extremes producing bifurcations along the way at critical points where the number of QREs changes (Figure 1).

Our goal in this paper is to quantify the effects of these rationality-driven bifurcations to the social welfare of two-player two-strategy games. At this point, a moment of pause is warranted. Why is this a worthy goal? Games of small size (2×2 games in particular) are rarely seem like a subject worthy of serious scientific investigation. This, however, could not be further from the truth.

First, the correct way to interpret this setting is from the point of population games where each agent is better understood as a large homogeneous population (e.g., men and women, attackers and defenders, cells of phenotype A, and cells of phenotype B). Each of a handful of different types of users has only a few meaningful actions available to them. In fact, from the perspective of applied game theory, only such games with a small number of parameters are practically meaningful. The reason should be clear by now. Any game theoretic modeling of a real life scenario is invariably noisy and inaccurate. In order for game-theoretic predictions to be practically binding, they have to be robust to these uncertainties. If the system intrinsically has a large number of independent parameters, e.g., 20, then this parameter space will almost certainly encode a vast number of bifurcations, which invalidate any theoretical prediction. Practically useful models *need* to be small.

Secondly, game theoretic models applied for scientific purposes *are* often small. Specifically, the exact setting studied here with Boltzmann Q-learning dynamics applied in 2×2 games has been used to model the effects of taxation to agent rationality [9] (see Section 6.2 for a more extensive discussion) as well as to model the effects of treatments that trigger phase transitions to cancer dynamics [10] (see Section 6.1). Our approach yields insights to explicit open questions in both of these applications areas. In fact, direct application of our analysis can address similar inquiries for any other phenomenon modeled by Q-learning dynamics applied in 2×2 games.

---

[1] Mixed strategies in the QRE model are sometimes interpreted as frequency distributions of deterministic actions in a large population of users. This population interpretation of mixed strategies is standard and dates back to Nash [8]. Depending on context, we will use either the probabilistic interpretation or the population one.

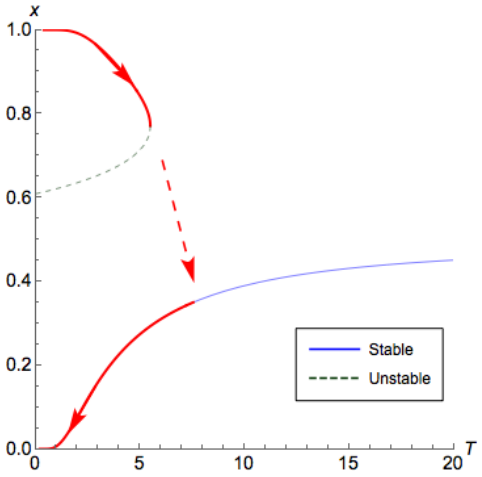

Figure 1. Bifurcation diagram for a 2×2 population coordination game. The x axis corresponds to the system temperature T, whereas the y axis corresponds to the projection of the proportion of the first population using the first strategy at equilibrium. For small T, the system exhibits multiple equilibria. Starting at $T = 0$, and by increasing the temperature beyond the critical threshold $T_C = 6$, and then bringing it back to zero, we can force the system to converge to another equilibrium.

Finally, the analysis itself is far from straightforward as it requires combining sets of tools and techniques that have so far been developed in isolation from each other. On one hand, we need to understand the behavior of these dynamical systems using tools from topology of dynamical systems, whose implications are largely qualitative (e.g., prove the lack of cyclic trajectories). On the other hand, we need to leverage these tools to quantify at which exact parameter values bifurcations occur and produce price-of-anarchy guarantees,which by definition are quantitative. As far as we know, this is the first instance of a fruitful combination of these tools. In fact, not only do we show how to analyze the effects of bifurcations to system efficiency, we also show how to leverage this understanding (e.g., knowledge of the geometry of the bifurcation diagrams) to design novel types of mechanisms with good performance guarantees.

Our Contribution

We introduce two different types of mechanisms: hysteresis and optimal control mechanisms.

Hysteresis mechanisms use transient changes to the system parameters to induce permanent improvements to its performance via optimal (Nash) equilibrium selection. The term hysteresis is derived from an ancient Greek word that means "to lag behind." It reflects a time-based dependence between the system's present output and its past inputs. For example, let's assume that we start from a game theoretic system of Q-learning agents with temperature $T = 0$ and assume that the system has converged to an equilibrium. By increasing the temperature beyond some critical threshold and then bringing it back to zero, we can force the system to provably converge to another equilibrium, e.g., the best (Nash) equilibrium (Figure 1, Theorem 4). Thus, we can ensure performance equivalent to that of the price of stability instead of the price of anarchy. One attractive feature of this mechanism is that from the perspective of the central designer it is relatively "cheap" to implement. Whereas typical mechanisms require the designer to continuously intervene (e.g., by paying the agents) to offset their greedy tendencies, this mechanism is transient with a finite amount of total effort from the perspective of the designer. Further, the idea that game theoretic systems have effectively systemic memory is rather interesting and could find other applications within algorithmic game theory.

Optimal control mechanisms induce convergence to states whose performance is better than even the best Nash equilibrium. Thus, we can at times even beat the price of stability (Theorem 5). Specifically, we show that by controlling the exploration/exploitation tradeoff, we can achieve strictly better states

than those achievable by perfectly rational agents. In order to implement such a mechanism, it does not suffice to identify the right set of agents' parameters/temperatures so that the system has some QRE whose social welfare is better than the best Nash. We need to design a trajectory through the parameter space so that this optimal QRE becomes the final resting point.

2. Preliminaries

2.1. Game Theory Basics: 2×2 *Games*

In this paper, we focus on 2×2 games. We define it as a game with two players, and each player has two actions. We write the payoff matrices of the game for each player as

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \tag{1}$$

respectively. The entry a_{ij} denotes the payoff for Player 1 when s/he chooses action i and his/her opponent chooses action j; similarly, b_{ij} denotes the payoff for Player 2 when s/he chooses action i and his/her opponent chooses action j. We define x as the probability that the Player 1 chooses his/her first action, and y as the probability that Player 2 chooses his/her first action. We also define two row vectors $\boldsymbol{x} = (x, 1-x)^T$ and $\boldsymbol{y} = (y, 1-y)^T$ as the *strategy* for each player. For simplicity, we denote the i-th entry of vector $\boldsymbol{x}$ by x_i. We call the tuple (x, y) as the *system state* or the *strategy profile*.

An important solution concept in game theory is the *Nash equilibrium*, where each user cannot make profit by unilaterally changing his/her strategy, that is

Definition 1 (Nash equilibrium)**.** *A strategy profile* (x_{NE}, y_{NE}) *is a Nash equilibrium (NE) if*

$$x_{NE} \in \arg\max_{x \in [0,1]} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y}_{NE} \qquad y_{NE} \in \arg\max_{y \in [0,1]} \boldsymbol{y}^T \boldsymbol{B} \boldsymbol{x}_{NE}.$$

We call (x_{NE}, y_{NE}) a *pure* Nash equilibrium (PNE) if both $x_{NE} \in \{0, 1\}$ and $y_{NE} \in \{0, 1\}$. A Nash equilibrium assumes each user is fully rational. An alternative solution concept is the *quantal response equilibrium* [7], where it assumes that each user has bounded rationality:

Definition 2 (Quantal response equilibrium)**.** *A strategy profile* (x_{QRE}, y_{QRE}) *is a QRE with respect to temperature* T_x *and* T_y *if*

$$x_{QRE} = \frac{e^{\frac{1}{T_x}(\boldsymbol{A}\boldsymbol{y}_{QRE})_1}}{\sum_{j \in \{1,2\}} e^{\frac{1}{T_x}(\boldsymbol{A}\boldsymbol{y}_{QRE})_j}} \qquad 1 - x_{QRE} = \frac{e^{\frac{1}{T_x}(\boldsymbol{A}\boldsymbol{y}_{QRE})_2}}{\sum_{j \in \{1,2\}} e^{\frac{1}{T_x}(\boldsymbol{A}\boldsymbol{y}_{QRE})_j}}$$

$$y_{QRE} = \frac{e^{\frac{1}{T_y}(\boldsymbol{B}\boldsymbol{x}_{QRE})_1}}{\sum_{j \in \{1,2\}} e^{\frac{1}{T_y}(\boldsymbol{B}\boldsymbol{x}_{QRE})_j}} \qquad 1 - y_{QRE} = \frac{e^{\frac{1}{T_y}(\boldsymbol{B}\boldsymbol{x}_{QRE})_2}}{\sum_{j \in \{1,2\}} e^{\frac{1}{T_y}(\boldsymbol{B}\boldsymbol{x}_{QRE})_j}}.$$

Analogous to the definition of Nash equilibria, we can consider the QREs as the case where each player is not only maximizing the expected utility but also maximizing the entropy. We can see that the QREs are the solutions to maximizing the linear combination of the following program:

$$\boldsymbol{x}_{QRE} \in \arg\max_{\boldsymbol{x}} \left\{ \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y}_{QRE} - T_x \sum_j x_j \ln x_j \right\}$$

$$\boldsymbol{y}_{QRE} \in \arg\max_{\boldsymbol{y}} \left\{ \boldsymbol{y}^T \boldsymbol{B} \boldsymbol{x}_{QRE} - T_y \sum_j y_j \ln y_j \right\}.$$

This formulation has been widely seen in Q-learning dynamics literature (e.g., [9,11,12]). With this formulation, we can find that the two parameters T_x and T_y control the weighting between the utility and the entropy. We call T_x and T_y the *temperatures*, and their values define the level of irrationality. If T_x and T_y are zero, then both players are fully rational, and the system state is a Nash equilibrium. However, if both T_x and T_y are infinity, then each player is choosing his/her action according to a uniform distribution, which corresponds to the fully irrational players.

2.2. Efficiency of an Equilibrium

The performance of a system state can be measured via the *social welfare*. Given a system state (x, y), we define the social welfare as the sum of the expected payoff of all users in the system:

Definition 3. *Given a* 2×2 *game with payoff matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, and a system state* (x, y)*, the social welfare is defined as*

$$SW(x,y) = xy(a_{11}+b_{11}) + x(1-y)(a_{12}+b_{21}) + y(1-x)(a_{21}+b_{12}) + (1-x)(1-y)(a_{22}+b_{22}).$$

In the context of algorithmic game theory, we can measure the efficiency of a game by comparing the best social welfare with the social welfare of equilibrium system states. We call the strategy profile that achieves the maximal social welfare as the *socially optimal (SO)* strategy profile. The efficiency of a game is often described as the notion of the *price of anarchy* (PoA) and the *price of stability* (PoS). Given a set of equilibrium states S, we define the PoA/PoS as the ratio of the social welfare of the socially optimal state to the social welfare of the worst/best equilibrium state in S, respectively. Formally,

Definition 4. *Given a* 2×2 *game with payoff matrices* A *and* B*, and a set of equilibrium system states* $S \subseteq [0,1]^2$*, the price of anarchy (PoA) and the price of stability (PoS) are defined as*

$$PoA(S) = \frac{\max_{(x,y)\in[0,1]^2} SW(x,y)}{\min_{(x,y)\in S} SW(x,y)} \qquad PoS(S) = \frac{\max_{(x,y)\in[0,1]^2} SW(x,y)}{\max_{(x,y)\in S} SW(x,y)}.$$

3. Our Model

3.1. Q-Learning Dynamics

In this paper, we are particularly interested in the scenario when both players' strategies are evolving under *Q-learning dynamics*:

$$\dot{x}_i = x_i\left[(\boldsymbol{Ay})_i - \boldsymbol{x}^T\boldsymbol{Ay} + T_x\sum_j x_j \ln(x_j/x_i)\right] \qquad \dot{y}_i = y_i\left[(\boldsymbol{Bx})_i - \boldsymbol{y}^T\boldsymbol{Bx} + T_y\sum_j y_j \ln(y_j/y_i)\right]. \tag{2}$$

Q-learning dynamics has been studied because of its connection with multi-agent learning problems. For example, it has been shown in [13,14] that Q-learning dynamics captures the system evolution of a repeated game, where each player learns his/her strategy through Q-learning and Boltzmann selection rules. More details are provided in Appendix A.

An important observation on the dynamics of Equation (2) is that it demonstrates the exploration/exploitation tradeoff [14]. We can find that the right hand side of Equation (2) is composed of two parts. The first part $x_i[(\boldsymbol{Ay})_i - \boldsymbol{x}^T\boldsymbol{Ay}]$ is exactly the vector field of replicator dynamic [15]. Basically, the replicator dynamics drives the system to the state of higher utility for both players. As a result, we can consider this as a selection process in terms of population evolutionary, or an exploitation process from the perspective of a learning agent. Then, for the second part, $x_i[T_x\sum_j x_j \ln(x_j/x_i)]$, we show in the appendix that if the time derivative of $\boldsymbol{x}$ contains this part alone, this results in an increase of the system entropy.

The system entropy is a function that captures the randomness of the system. From the population evolutionary perspective, the system entropy corresponds to the variety of the population. As a result, this term can be considered as the mutation process. The level of the mutation is controlled by the temperature parameters T_x and T_y. Besides, in terms of the reinforcement learning, this term can be considered as an exploration process, as it provides the opportunity for the agent to gain information about the action that does not look the best so far.

3.2. Convergence of the Q-Learning Dynamics

By observing the Q-learning dynamics of Equation (2), we can find that the interior rest points for the dynamics are exactly the QREs of the 2×2 game. It is claimed in [16] (albeit without proof) that the Q-learning dynamics for a 2×2 game converges to interior rest points of probability simplexes for any positive temperature $T_x > 0$ and $T_y > 0$. We provide a formal proof in Appendix B. The idea is that, for positive temperatures, the system is dissipative and, by leveraging the planar nature of the system, it can be argued that it converges to fixed points.

3.3. Rescaling the Payoff Matrix

At the end of this section, we discuss the transformation of the payoff matrices that preserves the dynamics in Equation (2). This idea is proposed in [17,18], where the *rescaling* of a matrix is defined as follows

Definition 5 ([18]). *$\boldsymbol{A}'$ and $\boldsymbol{B}'$ is said to be a rescaling of $\boldsymbol{A}$ and $\boldsymbol{B}$ if there exist constants c_j, d_i, and $\alpha > 0$, $\beta > 0$ such that $a'_{ij} = \alpha a_{ij} + c_j$ and $b'_{ji} = \beta b_{ji} + d_i$.*

It is clear that rescaling the game payoff matrices is equivalent to updating the temperature parameters of the two agents in Equation (2). Therefore, it suffices to study the dynamics under the assumption that the 2×2 payoff matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are in the following *diagonal form*.

Definition 6. *Given 2×2 matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, their diagonal form is defined as*

$$\boldsymbol{A}_D = \begin{pmatrix} a_{11} - a_{21} & 0 \\ 0 & a_{22} - a_{12} \end{pmatrix} \quad \boldsymbol{B}_D = \begin{pmatrix} b_{11} - b_{21} & 0 \\ 0 & b_{22} - b_{12} \end{pmatrix}$$

Note that, although rescaling the payoff matrices to their diagonal form preserves the equilibria, it does not preserve the social optimality, i.e., the socially optimal strategy profile in the transformed game is not necessarily the socially optimal strategy profile in the original game.

4. Hysteresis Effect and Bifurcation Analysis

4.1. Hysteresis effect in Q-Learning Dynamics: An Example

We begin our discussion with an example:

Example 1 (Hysteresis effect). *Consider a 2×2 game with reward matrices*

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \tag{3}$$

There are two PNEs in this game: $(x, y) = (0, 0)$ and $(1, 1)$. By fixing different T_y, we can plot different QREs with respect to T_x as in Figures 2 and 3, which we call the bifurcation diagrams. For simplicity, we only show the value of x in the figure, since, according to Equation (4), given x and T_y, the value of y can be uniquely determined. Assuming the system follows the Q-learning dynamics, as we slowly vary T_x, x tends to stay on the

line segment that is the closest to where it was originally corresponding to a stable but inefficient fixed point. We consider the following process:

1. *Where the initial state is* $(0.05, 0.14)$*, where* $T_x \approx 1$ *and* $T_y \approx 2$*, plot* x *versus* T_x *by fixing* $T_y = 2$ *in Figure 3.*
2. *Fix* $T_y = 2$ *and increase* T_x *to where there is only one QRE correspondence.*
3. *Fix* $T_y = 2$ *and decrease* T_x *back to* 1*. Now* $x \approx 0.997$*.*

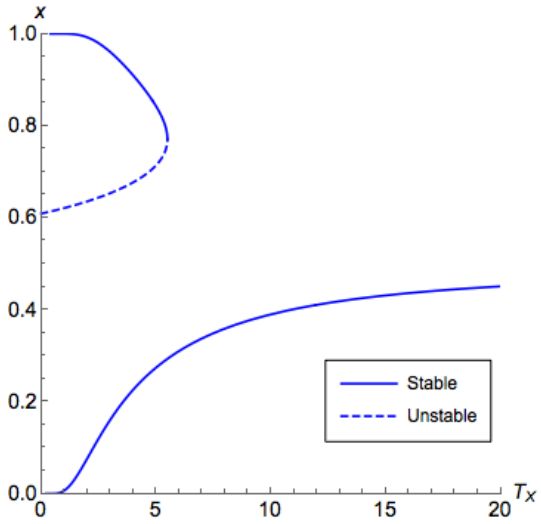

Figure 2. The bifurcation diagram for Example 1 with $T_y = 0.5$. The horizontal axis corresponds to the temperature T_x for the first (row) player and the vertical axis corresponds to the probability that the first player chooses the first action in equilibrium. There exist three branches (two stable and one unstable). For $x > 0.5$, there are two branches appearing in pairs, and they occur only when T_x is less than some value. For $x < 0.5$, there is a branch, which we call the principal branch, where the quantal response equilibrium (QRE) always exists for any $T_x > 0$.

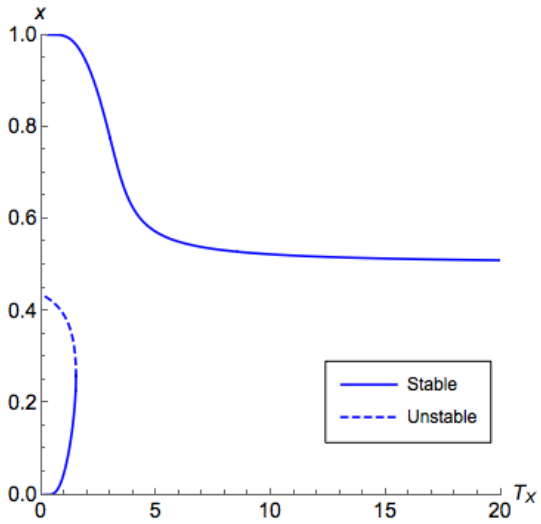

Figure 3. Bifurcation diagram for Example 1 with $T_y = 2$. The horizontal axis corresponds to the temperature T_x for the first (row) player and the vertical axis corresponds to the probability that the first player chooses the first action in equilibrium. Similar to Figure 2, there exist three branches (two stable and one unstable). However, unlike Figure 2, now the two branches appearing in pairs happen at $x < 0.5$, and the principal branch is at $x > 0.5$.

In the above example, we can find that, although at the end the temperature parameters are set back to their initial value, the system state ends up being an entirely different equilibrium. This behavior is known as the *hysteresis effect*. In this section, we would like to answer the question of *when this is going to happen*. Further, in the next section, we will show *how can we take advantage of this phenomenon*.

4.2. Characterizing QREs

We consider the bifurcation diagrams for QREs in 2×2 games. Without loss of generality, we consider a properly rescaled 2×2 game with payoff matrices in the diagonal form:

$$\boldsymbol{A}_D = \begin{pmatrix} a_X & 0 \\ 0 & b_X \end{pmatrix}, \quad \boldsymbol{B}_D = \begin{pmatrix} a_Y & 0 \\ 0 & b_Y \end{pmatrix}$$

We can also assume that the action indices are ordered properly and rescaled properly so that $a_X > 0$ and $|a_X| \geq |b_X|$. For simplicity, we assume $a_X = b_X$ and $b_X = b_Y$ do not hold at the same time. At QRE, we have

$$x = \frac{e^{\frac{1}{T_x} y a_X}}{e^{\frac{1}{T_x} y a_X} + e^{\frac{1}{T_x}(1-y) b_X}} \qquad y = \frac{e^{\frac{1}{T_y} x a_Y}}{e^{\frac{1}{T_y} x a_Y} + e^{\frac{1}{T_y}(1-x) b_Y}}. \tag{4}$$

Given T_x and T_y, there could be multiple solutions to Equation (4). However, we find that, if we know the equilibrium states, then we can recover the temperature parameters. We solve for T_x and T_y in Equation (4) and get

$$T_X^I(x,y) = \frac{-(a_X + b_X)y + b_X}{\ln(\frac{1}{x} - 1)} \qquad T_Y^I(x,y) = \frac{-(a_Y + b_Y)x + b_Y}{\ln(\frac{1}{y} - 1)}. \tag{5}$$

We call this the *first form of representation*, where T_x and T_y are written as functions of x and y. Here the capital subscripts for T_X and T_Y indicate that they are considered as functions. A direct observation of Equation (5) is that both of them are continuous function over $(0,1) \times (0,1)$ except for $x = 1/2$ and $y = 1/2$.

An alternative way to describe the QRE is to write T_x and y as a function of x and parameterize with respect to T_y in the following *second form of representation*. This will be the form that we use to prove many useful characteristics of QREs.

$$T_X^{II}(x, T_y) = \frac{-(a_X + b_X) y^{II}(x, T_y) + b_X}{\ln(\frac{1}{x} - 1)} \tag{6}$$

$$y^{II}(x, T_y) = \left(1 + e^{\frac{1}{T_y}(-(a_Y + b_Y)x + b_Y)}\right)^{-1}. \tag{7}$$

In this way, if we are given T_y, we are able to analyze how T_x changes with x. This helps us understand how to answer the question of what the QREs are given T_x and T_x in the system.

We also want to analyze the stability of the QREs. From dynamical system theory (e.g., [19]), a fixed point of a dynamical system is said to be asymptotically stable if all of the eigenvalues of its Jacobian matrix have a negative real part; if it has at least one eigenvalue with a positive real part, then it is unstable. It turns out that, under the second form representation, we are able to determine whether a segment in the diagram is stable or not.

Lemma 1. *Given T_y, the system state $\left(x, y^{II}(x, T_y)\right)$ is a stable equilibrium if and only if*

1. $\frac{\partial T_X^{II}}{\partial x}(x, T_Y) > 0$ *if* $x \in (0, 1/2)$;
2. $\frac{\partial T_X^{II}}{\partial x}(x, T_Y) < 0$ *if* $x \in (1/2, 1)$.

Proof. The given condition is equivalent to the case where both eigenvalues of the Jacobian matrix of the dynamics (2) are negative. □

Finally, we define the *principal branch*. In Example 1, we call the branch on $x \in (0.5, 1)$ the *principal branch* given $T_y = 2$, since, for any $T_x > 0$, there is some $x \in (0.5, 1)$ such that $T_X^{II}(x, T_y) = T_x$. Analogously, we can define it formally as in the following definition with the help of the second form representation.

Definition 7. *Given T_y, the region $(a, b) \subset (0, 1)$* contains the principal branch *of QRE correspondence if it satisfies the following conditions:*

1. *$T_X^{II}(x, T_y)$ is continuous and differentiable for $x \in (a, b)$.*
2. *$T_X^{II}(x, T_y) > 0$ for $x \in (a, b)$.*
3. *For any $T_x > 0$, there exists $x \in (a, b)$ such that $T_X^{II}(x, T_y) = T_x$.*

Further, for a region (a, b) that contains the principal branch, $x \in (a, b)$ is on the principal branch *if it satisfies the following conditions:*

1. *The equilibrium state $(x, y^{II}(x, T_y))$ is stable.*
2. *There is no $x' \in (a, b), x' < x$ such that $T_X^{II}(x', T_y) = T_X^{II}(x, T_y)$.*

4.3. Coordination Games

We begin our analysis with the class of coordination games, where we have all a_X, b_X, a_Y, and b_Y positive. Additionally, without loss of generality, we assume $a_X \geq b_X$. In this case, there is no dominant strategy for either player, and there are two PNEs.

Let us revisit Example 1, we can make the following observations from Figures 2 and 3:

1. Given T_y, there are three branches. One is the principal branch, while the other two appear in pairs and occur only when T_x is less than some value.
2. For small T_y, the principal branch goes toward $x = 0$; for a large T_y, the principal branch goes toward $x = 1$.

Now, we are going to show that these observations are generally true in coordination games. The proofs in this section are deferred to Appendix D, where we will provide a detailed discussion on the proving techniques.

The first idea we are going to introduce is the *inverting temperature*, which is the threshold of T_y in Observation (2). We define it as

$$T_I = \max\left\{0, \frac{b_Y - a_Y}{2\ln(a_X/b_X)}\right\}.$$

We note that T_I is positive only if $b_Y > a_Y$, which is the case where two players have different preferences. When $T_y < T_I$, as the first player increases his/her rationality from fully irrational, i.e., T_x decreases from infinity, s/he is likely to be influenced by the second player's preference. If T_y is greater than T_I, then the first player prefers to follow his/her own preference, making the principal branch move toward $x = 1$. We formalize this idea in the following theorem:

Theorem 1 (Direction of the principal branch)**.** *Given a* 2×2 *coordination game, and given* T_y*, the following statements are true:*

1. *If* $T_y > T_I$*, then* $(0.5, 1)$ *contains the principal branch.*
2. *If* $T_y < T_I$*, then* $(0, 0.5)$ *contains the principal branch.*

The second idea is the *critical temperature*, denoted as $T_C(T_y)$, which is a function of T_y. The critical temperature is defined as the infimum of T_x such that, for any $T_x > T_C(T_y)$, there is a unique QRE correspondence under (T_x, T_y). Generally, there is no close form for the critical temperature. However, we can still compute it efficiently, as we show in Theorem 2. Another interesting value of T_y we should point out is $T_B = \frac{b_Y}{\ln(a_X/b_X)}$, which is the maximum value of T_y that QREs not on the principal branch are presenting. Intuitively, as T_y goes beyond T_B, the first player ignores the decision of the second player and turns his/her face to what s/he thinks is better. We formalize the idea of T_C and T_B in the following theorem:

Theorem 2 (Properties about the second QRE)**.** *Given a* 2×2 *coordination game, and given* T_y*, the following statements are true:*

1. *For almost every* $T_x > 0$*, all QREs not lying on the principal branch appear in pairs.*
2. *If* $T_y > T_B$*, then there is no QRE correspondence in* $x \in (0, 0.5)$*.*
3. *If* $T_y > T_I$*, then there is no QRE correspondence for* $T_x > T_C(T_y)$ *in* $x \in (0, 0.5)$*.*
4. *If* $T_y < T_I$*, then there is no QRE correspondence for* $T_x > T_C(T_y)$ *in* $x \in (0.5, 1)$*.*
5. $T_C(T_y)$ *is given as* $T_X^{II}(x_L, T_y)$*, where* x_L *is the solution to the equality*

$$y^{II}(x, T_y) + x(1-x)\ln\left(\frac{1}{x} - 1\right)\frac{\partial y^{II}}{\partial x}(x, T_y) = \frac{b_X}{a_X + b_X}.$$

6. x_L *can be found using binary search.*

The next aspect of the QRE correspondence is their stability. According to Lemma 1, the stability of the QREs can also be inspected with the advantage of the second form representation by analyzing $\frac{\partial T_X^{II}}{\partial x}$. We state the results in the following theorem:

Theorem 3 (Stability)**.** *Given a* 2×2 *coordination game, and given* T_y*, the following statements are true:*

1. *If* $a_Y \geq b_Y$*, then the principal branch is continuous.*
2. *If* $T_y < T_I$*, then the principal branch is continuous.*
3. *If* $T_y > T_I$ *and* $a_Y < b_Y$*, then the principal branch may not be continuous.*
4. *If* T_x *is fixed, for the pairs of QREs not lying on the principal branch, the one with the lowest distance to* $x = 0.5$ *is unstable, while the other one is stable.*

Note that Part 3 in Theorem 3 infers that there is potentially an unstable segment between segments of the principal branch. This phenomenon is illustrated in Figures 4 and 5. Though this case is weaker than other cases, this does not hinder us from designing a controlling mechanism as we are going to do in Section 5.3.

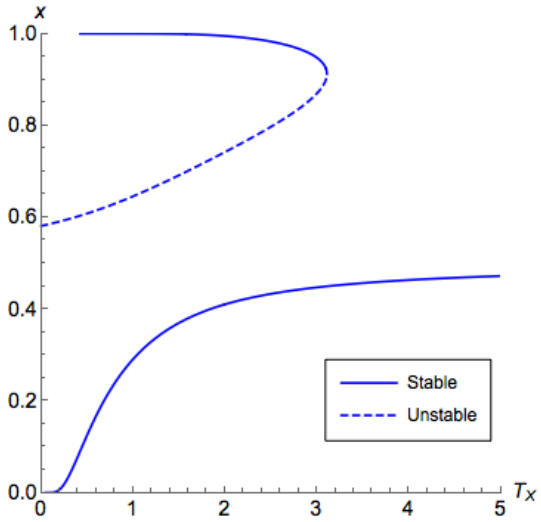

Figure 4. Bifurcation diagram for a coordination game with $a_Y < b_Y$ and a low T_y. The horizontal axis corresponds to the temperature T_x for the first (row) player and the vertical axis corresponds to the probability that the first player chooses the first action in equilibrium. We can find that the principal branch is contained in $x < 0.5$.

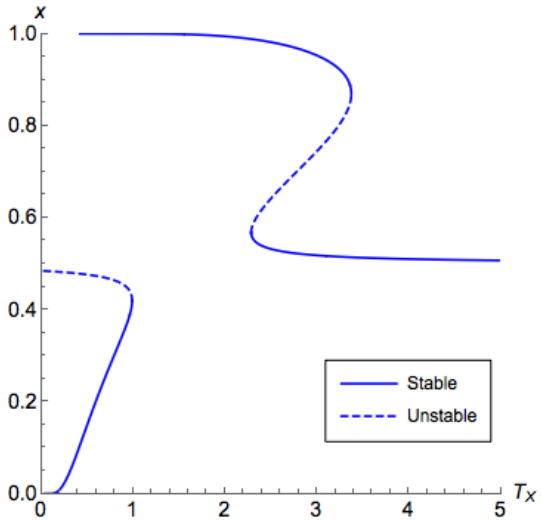

Figure 5. Bifurcation diagram for a coordination game with $a_Y < b_Y$ and a high T_Y. The horizontal axis corresponds to the temperature T_x for the first (row) player and the vertical axis corresponds to the probability that the first player chooses the first action in equilibrium. We can find that the principal branch is contained in $x > 0.5$. In addition, there is a non-stable segment on the principal branch.

4.4. Non-Coordination Games

Due to space constraints, the analysis for non-coordination games is deferred to Appendix C.

5. Mechanism Design

In this section, we aim to design a systematic way to improve the social welfare in a 2×2 game by changing the temperature parameters. We focus our discussion on the class of coordination games. Recall that any 2×2 game has more than one PNE if and only if its diagonal form is a coordination game. This means that, in a coordination game, given any temperature parameters, there could be more than one equilibrium correspondences. In this case, we are not guaranteed to achieve the socially

optimal equilibrium state even if we set the system to the *correct* temperatures due to the hysteresis effects that we have discussed in the previous section. Therefore, the main task for us in this section is to determine when and how we can get to the socially optimal equilibrium state. In Section 5.3, we consider the case when the socially optimal state is one of the PNEs. Since rescaling the payoff matrices to their diagonal form does not preserve the social optimality, in Section 5.1, we generalize our discussion to the case when the social optimal state does not coincide with any PNE.

5.1. Hysteresis Mechanism: Select the Best Nash Equilibrium via QRE Dynamics

First, we consider the case when the socially optimal state is one of the PNEs. The main task for us in this case is to determine when and how we can get to the socially optimal PNE. In Example 1, by sequentially changing T_x, we move the equilibrium state from around $(0,0)$ to around $(1,1)$, which is the social optimum state. We formalize this idea as the *hysteresis mechanism* and present it in Theorem 4. The hysteresis mechanism mainly takes advantage of the hysteresis effect we have discussed in Section 4—that we use transient changes of system parameters to induce permanent improvements to system performance via optimal equilibrium selection.

Theorem 4 (Hysteresis Mechanism)**.** *Consider a* 2×2 *game that satisfies the following properties:*

1. *Its diagonal form satisfies* $a_X, b_X, a_Y, b_Y > 0$.
2. *Exactly one of its pure Nash equilibrium is the socially optimal state.*

Without loss of generality, we can assume $a_X \geq b_X$. *Then there is a mechanism to control the system to the social optimum by sequentially changing* T_x *and* T_y *if (1)* $a_Y \geq b_Y$ *and (2) the socially optimal state is* $(0,0)$ *do not hold at the same time.*

Proof. First, note that, if $a_Y \geq b_Y$, by Theorem 1, the principal branch is always in the region $x > 0.5$. As a result, once T_y is increased beyond the critical temperature, the system state will no longer return to $x < 0.5$ at any positive temperature. Therefore, $(0,0)$ cannot be approached from any state in $x > 0.5$ through the QRE dynamics.

On the other hand, if $a_Y \geq b_Y$ and the socially optimal state is the PNE $(1,1)$, then we can approach that state by first getting onto the principal branch. The mechanism can be described as

(C1) (a) Raise T_x to some value above the critical temperature $T_C(T_y)$.
(b) Reduce T_x and T_y to 0.

Though in this case the initial choice of T_y does not affect the result, if the social designer is taking the costs from assigning large T_x and T_y values into account, s/he is going to trade off between T_C and T_y since a typically smaller T_y induces a larger T_C.

Next, consider $a_Y < b_Y$. If we are aiming for state $(0,0)$, then we can undergo the following procedure:

(D1) (a) Keep T_y at some value below $T_I = \frac{b_Y - a_Y}{2\ln(a_X/b_X)}$. Now the principal branch is at $(0, 0.5)$.
(b) Raise T_x to some value above the critical temperature $T_C(T_y)$.
(c) Reduce T_x to 0.
(d) Reduce T_y to 0.

On the other hand, if we are aiming for state $(1,1)$, then the following procedure suffices:

(D2) (a) Keep T_y at some value above $T_I = \frac{b_Y - a_Y}{2\ln(a_X/b_X)}$. Now the principal branch is at $(0.5, 1)$.
(b) Raise T_x to some value above the critical temperature $T_C(T_y)$.
(c) Reduce T_x to 0.
(d) Reduce T_y to 0.

Note that, in the last two steps, only by reducing T_y after T_x keeps the state around $x = 1$. We recommend that the interested reader refers to Figure 11 for Case (D1) and Figure 12 for Case (D2) for more insights. □

5.2. Efficiency of QREs: An Example

A question that arises with the solution concept of QRE is *whether QRE improves social welfare*? Here we show that the answer is *yes*. We begin with an example to illustrate:

Example 2. *Consider a standard coordination game with the payoff matrices of the form*

$$A = \begin{pmatrix} \epsilon & 1 \\ 0 & 1+\epsilon' \end{pmatrix} \quad B = \begin{pmatrix} 1+\epsilon & 0 \\ 1 & \epsilon' \end{pmatrix} \tag{8}$$

where $\epsilon > \epsilon' > 0$ are some small numbers. Note that, in this game, there are two PNEs, $(x, y) = (1, 1)$ and $(x, y) = (0, 0)$, with social welfare values $1 + 2\epsilon$ and $1 + 2\epsilon'$, respectively. We can see that for small ϵ and ϵ' values, the socially optimal state is $(x, y) = (1, 0)$, with social welfare value 2. In this case, the state $(x, y) = (1, 1)$ is the PNE with the best social welfare. However, we are able to achieve a state with better social welfare than any NE through QRE dynamics. We illustrate the social welfare of the QREs with different temperatures of this example in Figure 6. In this figure, we can see that, at PNE, which is the point $T_x = T_y = 0$, the social welfare is $1 + 2\epsilon$. However, we are able to increase the social welfare by increasing T_y. We will show in Section 5.3 a general algorithm for finding particular temperature as well as a mechanism, which we refer to as the optimal control mechanism, *that drives the system to the desired state.*

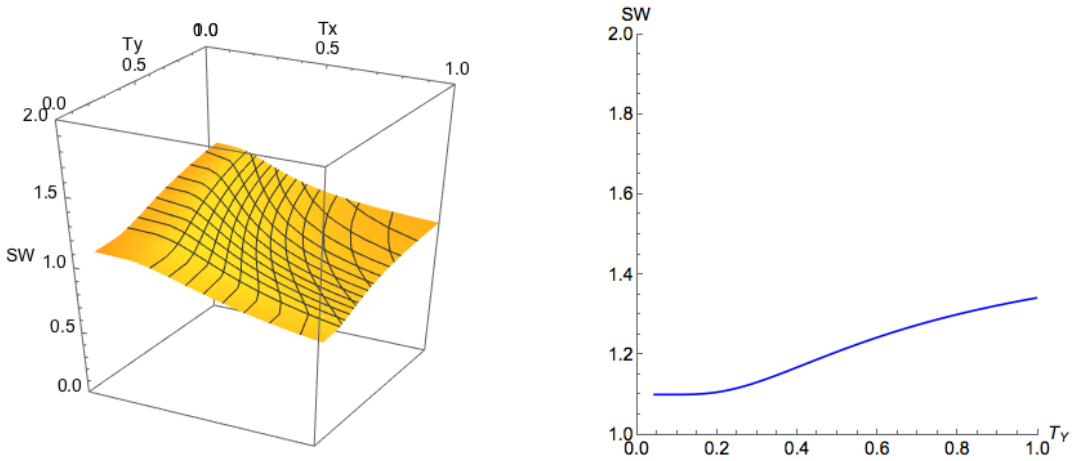

Figure 6. The left figure is the social welfare on the principal branch for Example 2, and the right figure is an illustration when $T_X = 0$. We can see that by increasing T_y, we can obtain an equilibrium with a social welfare higher than that of the best Nash equilibrium (which is $T_x = T_y = 0$).

5.3. Optimal Control Mechanism: Better Equilibrium with Irrationality

Here, we show a general approach to improve the PoS bound for coordination games from Nash equilibria by QREs and Q-learning dynamics. We denote $QRE(T_x, T_y)$ as the set of QREs with respect to T_x and T_y. Further, denote QRE as the set of the union of $QRE(T_x, T_y)$ over all positive T_x and T_y. Additionally, denote the set of pure Nash equilibria system states as NE. Since the set NE is the limit of $QRE(T_x, T_y)$ as T_x and T_y approach zero, we have the bounds:

$$PoA(QRE) \geq PoA(NE), \quad PoS(QRE) \leq PoS(NE).$$

Then, we define *QRE-achievable states*:

Definition 8. *A state $(x, y) \in [0, 1]^2$ is a QRE-achievable state if for every $\epsilon > 0$, there is a positive finite T_x and T_y and (x', y') such that $|(x', y') - (x, y)| < \epsilon$ and $(x', y') \in QRE(T_x, T_y)$.*

Note that, with this definition, pure Nash equilibria are QRE-achievable states. However, the socially optimal states are not necessarily QRE-achievable. For example, we illustrate in Figure 7 the set of QRE-achievable states for Example 2. We can find that the socially optimal state, $(x, y) = (1, 0)$, is not QRE-achievable. Nevertheless, it is easy to see from Figures 7 and 8 that we can achieve a higher social welfare at $(x, y) = (1, 0.5)$, which is a QRE-achievable state. Formally, we can describe the set of QRE-achievable states as the positive support of T_X^I and T_Y^I:

$$S = \left\{ \left\{ x \in \left[\frac{1}{2}, 1 \right], y \in \left[\frac{b_X}{a_X + b_X}, 1 \right] \right\} \cup \left\{ x \in \left[0, \frac{1}{2} \right], y \in \left[0, \frac{b_X}{a_X + b_X} \right] \right\} \right\}$$
$$\cap \left\{ \left\{ x \in \left[\frac{b_Y}{a_Y + b_Y}, 1 \right], y \in \left[\frac{1}{2}, 1 \right] \right\} \cup \left\{ x \in \left[0, \frac{b_Y}{a_Y + b_Y} \right], y \in \left[0, \frac{1}{2} \right] \right\} \right\}.$$

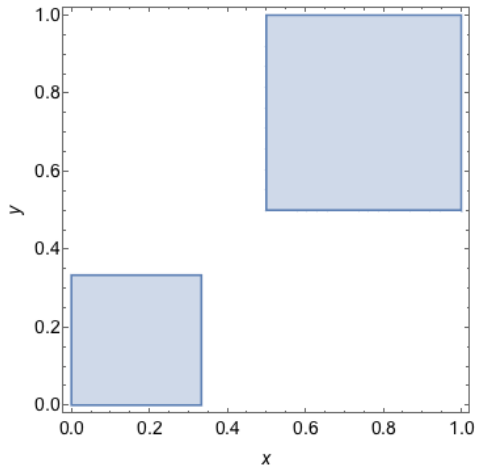

Figure 7. Set of QRE-achievable states for Example 2. A point (x, y) represents a mixed strategy profile where the first agent chooses its first strategy with probability x and the second agent chooses its first strategy with probability y. The grey areas depict the set of mixed strategy profiles (x, y) that can be reproduced as QRE states for Example 2, i.e., these are outcomes for which there are temperature parameters (T_x, T_y) for which the (x, y) mixed strategy profile is a QRE.

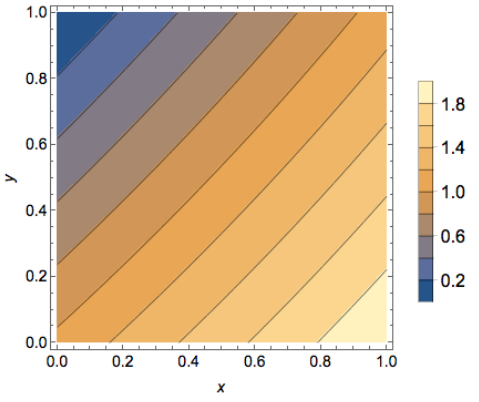

Figure 8. Social welfare for all states in Example 2. A point (x, y) represents a mixed strategy profile where the first agent chooses its first strategy with probability x and the second agent chooses its first strategy with probability y. The color of the point (x, y) corresponds to the social welfare of that mixed strategy profile with states of higher social welfare corresponding to lighter shades. The optimal state is $(1, 0)$, whereas the worst state is $(0, 1)$.

An example for the region of a game with $a_Y \geq b_Y$ is illustrated in Figure 7. For the case $a_Y < b_Y$, we demonstrate it in Figure 9.

In the following theorem, we propose the *optimal control mechanism* for a general process to achieve an equilibrium that is better than the PoS bound from Nash equilibria.

Theorem 5 (Optimal Control Mechanism). *Given a* 2×2 *game, if it satisfies the following property:*

1. *Its diagonal form satisfies* $a_X, b_X, a_Y, b_Y > 0$.
2. *None of its pure Nash equilibrium is the socially optimal state.*

Without loss of generality, we can assume $a_X \geq b_X$. *Then*

1. *there is a stable QRE-achievable state whose social welfare is better than any Nash equilibrium;*
2. *there is a mechanism to control the system to this state from the best Nash equilibrium by sequentially changing* T_x *and* T_y.

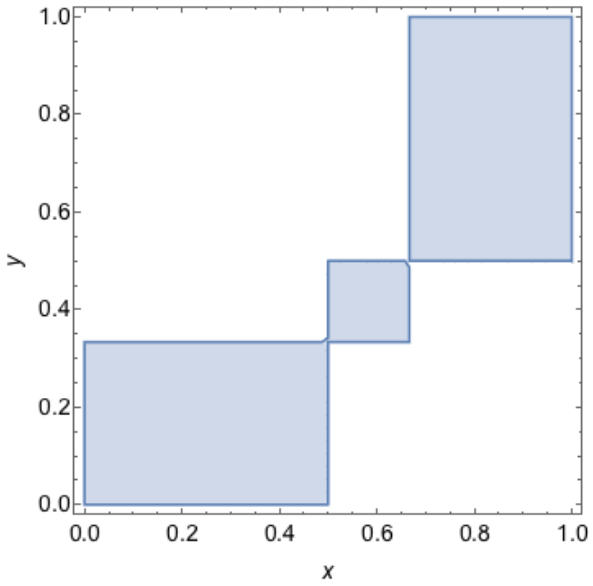

Figure 9. Set of QRE-achievable states for a coordination game with $a_Y < b_Y$. A point (x, y) represents a mixed strategy profile where the first agent chooses its first strategy with probability x and the second agent chooses its first strategy with probability y. The grey areas depict the set of mixed strategy profiles (x, y) that can be reproduced as QRE states a coordination game with $a_Y < b_Y$, i.e., these are outcomes for which there exists temperature parameters (T_x, T_y) for which the (x, y) mixed strategy profile is a QRE.

Proof. Note that, given those properties, there are two PNEs $(0, 0)$ and $(1, 1)$. Since we know neither of them is socially optimal, the socially optimal state must be either $(0, 1)$ or $(1, 0)$.

First, consider $a_Y \geq b_Y$. In this case, we know from Theorem 3 that all $x \in (0.5, 1)$ states belong to a principal branch for some $T_y > 0$ and are stable, while for $x < 0.5$ not all of them are stable. We illustrate the region of stable QRE-achievable states in Figure 10. By Theorems 2 and 3, we can infer that the states near the border $x = 0$ are stable. As a result, we can claim that the following states are what we are aiming for:

(A1) If $(1, 1)$ is the best NE and $(0, 1)$ is the SO state, then we select $(0.5, 1)$.
(A2) If $(1, 1)$ is the best NE and $(1, 0)$ is the SO state, then we select $(1, 0.5)$.
(A3) If $(0, 0)$ is the best NE and $(0, 1)$ is the SO state, then we select $\left(0, \frac{b_X}{a_X + b_X}\right)$.

(A4) If $(0,0)$ is the best NE and $(1,0)$ is the SO state, then we select $\left(\frac{b_Y}{a_Y+b_Y}, 0\right)$.

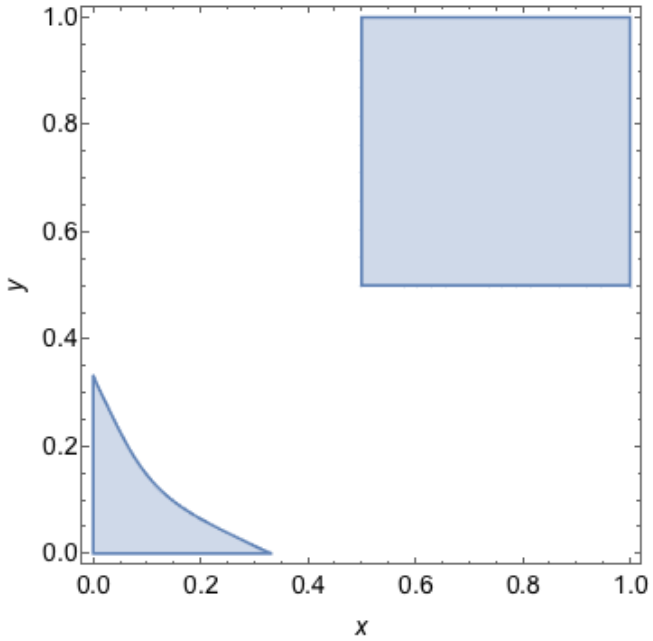

Figure 10. Stable QRE-achievable states for a coordination game with $a_Y > b_Y$. A point (x, y) represents a mixed strategy profile, where the first agent chooses its first strategy with probability x and the second agent chooses its first strategy with probability y. The grey areas depict the set of mixed strategy profiles (x, y) that can be reproduced as *stable* QRE states a coordination game with $a_Y > b_Y$, i.e., these are outcomes for which there are temperature parameters (T_x, T_y) for which the (x, y) mixed strategy profile is a *stable* QRE.

It is clear that these choices of states improve the social welfare. It is known that for the class of games we are considering, the price of stability is no greater than 2. In fact, in Cases A1 and A2, we reduce this factor to $4/3$. Additionally, in Cases A3 and A4, we reduce this factor to $\left(\frac{1}{2} + \frac{b_X/2}{a_X+b_X}\right)^{-1}$.

The next step is to show the mechanism to drive the system to the desired state. Due to symmetry, we only discuss Cases A1 and A3, where Cases A2 and A4 can be done analogously. For Case A1, the state corresponds to the temperatures $T_x \to \infty$ and $T_y \to 0$. For any small $\delta > 0$, we can always find the state $(0.5 + \delta, 1 - \delta)$ on the principal branch of some T_y. This means that we can achieve this state from any initial state, not only from the NEs. With the help of the first form representation of the QREs in Equation (5), given any QRE-achievable system state (x, y), we are able to recover them to corresponding temperatures through T_X^I and T_Y^I. The mechanism can be described as follows:

(A1) (a) From any initial state, raise T_x to $T_X^I(0.5 + \delta, 1 - \delta)$.
(b) Decrease T_y to $T_Y^I(0.5 + \delta, 1 - \delta)$.

For Case A3, the state we selected is not on the principal branch. This means that we cannot increase the temperatures too much; otherwise, the system state will move to the principal branch and will never return. We assume initially the system state is at (δ, δ) for some small $\delta > 0$, which is some state close to the best NE. Additionally, we can assume the initial temperatures are $T_x = T_X^I(\delta, \delta)$ and $T_y = T_Y^I(\delta, \delta)$. Our goal is to arrive at the state $\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$ for some small $\delta_1 > 0$ and $\delta_2 > 0$ such that $\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$ is stable. We present the mechanism in the following:

(A3) (a) From the initial state (δ, δ), move T_x to $T_X^I\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$.

(b) Increase T_y to $T_Y^I\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$.

Here, note that Step (b) should not proceed before Step (a) because, if we increase T_y first, then we risk leaving the principal branch.

Next, consider the case where $a_Y < b_Y$. Similarly to the previous case, we know from Theorems 2 and 3 that states near the borders $x = 0, 0.5, 1$ and $y = 0, 0.5, 1$ are basically stable states. Hence, we can claim the following results:

(B1) If $(1,1)$ is the best NE and $(0,1)$ is the SO state, then we select $\left(\frac{b_Y}{a_Y+b_Y}, 1\right)$.

(B2) If $(1,1)$ is the best NE and $(1,0)$ is the SO state, then we select $(1, 0.5)$.

(B3) If $(0,0)$ is the best NE and $(0,1)$ is the SO state, then we select $\left(0, \frac{b_X}{a_X+b_X}\right)$.

(B4) If $(0,0)$ is the best NE and $(1,0)$ is the SO state, then we select $(0.5, 0)$.

It is clear that these choices of states create improvement on the social welfare. An interesting result for this case is that basically these desired states can be reached from any initial state. Due to symmetry, we demonstrate the mechanisms for Cases (B3) and (B4), and the remaining ones can be done analogously.

For Case (B3), we are aiming for the state $\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$ for some small $\delta_1 > 0$ and $\delta_2 > 0$. We propose the following mechanism:

(B3) **Phase 1:** Getting to the principal branch.

(a) From any initial state, fix T_y at some value less than $T_I = \frac{b_Y - a_Y}{2\ln(a_X/b_X)}$.

(b) Increase T_x above the critical temperature $T_C(T_y)$.

(c) Decrease T_x to $T_x^I\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$.

Phase 2: Staying at the current branch.

(a) Increase T_y to $T_Y^I\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$.

This process is illustrated in Figures 11 and 12. In Phase 1, as we are keeping low T_y, meaning the second player is of more rationality. As the first player getting more rational, s/he is more likely to be influenced by the second player's preference, and eventually getting to a Nash equilibrium. In phase 2, we make the second player more irrational to increase the social welfare. The level of irrationality we add in phase 2 should be capped to prevent the first player to deviate his/her decision.

For Case (B4), since our desired state is on the principal branch, the mechanism will be similar to Case (A1).

(B4) (a) From any initial state, raise T_x to $T_X^I(0.5 + \delta, \delta)$.

(b) Decrease T_y to $T_Y^I(0.5 + \delta, \delta)$.

□

As a remark, in Cases (A3) and (A4), if we do not start from (δ, δ) but from some other states on the principal branch, we can instead aim for state $(0.5, 1)$. This state is not better than the best Nash equilibrium, but still makes improvements over the initial state. The process can be modified as

(A3') (a) From any initial state, raise T_x to $T_X^I(0.5 + \delta, 1 - \delta)$ (above $T_C(T_y)$).

(b) Reduce T_y to $T_Y^I(0.5 + \delta, 1 - \delta)$.

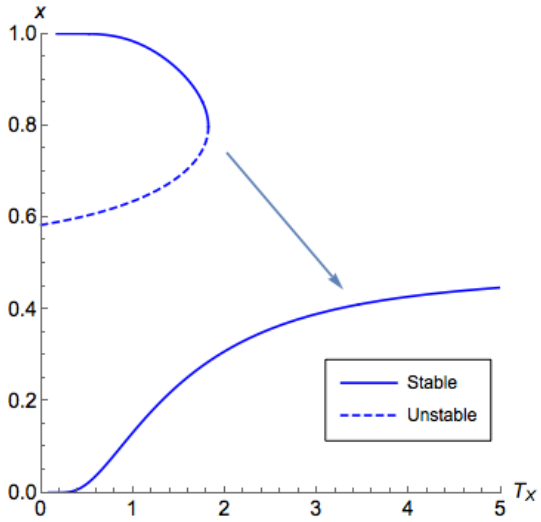

Figure 11. Illustration for Phase 1 in Case (B3), where we keep low T_Y but increase T_X and then decrease T_X back to a small value. In this phase, the equilibrium state moves from the branch where $x \in (0.7, 1.0)$ to the principal branch (the branch where $x < 0.5$).

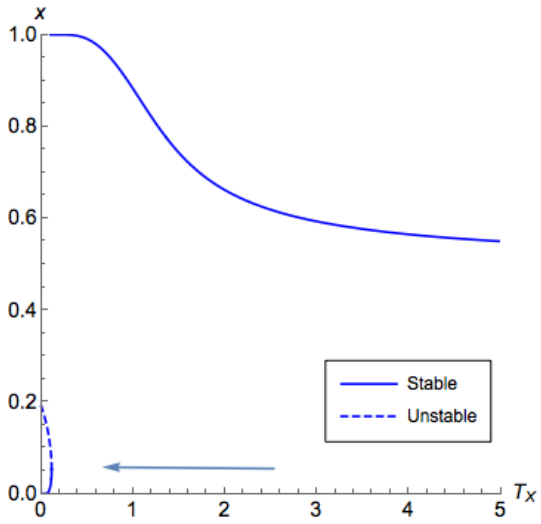

Figure 12. Illustration for Phase 2 in Case (B3). In this phase, we increase T_Y to $T_Y^I\left(\delta_1, \frac{b_X}{a_X+b_X} - \delta_2\right)$. The principal branch switches from $x < 0.5$ to $x > 0.5$ and the equilibrium state stays on the branch $x < 0.5$ (the branch pointed out by the blue arrow) only if T_X is low.

6. Applications

6.1. Evolution of Metabolic Phenotypes in Cancer

Evolutionary game theory (EGT) has been instrumental in studying evolutionary aspects of the somatic evolution that characterizes cancer progression. As opposed to conventional game theory, in evolutionary game theory, the strategies are fixed for the player and constitute its phenotype. Tumors are very heterogeneous, and frequency-dependent selection is a driving force in somatic evolution. While evolutionary outcomes can change depending on initial conditions or on the exact features and microenvironment of the relevant tumor phenotypes, evolutionary game theory can explain why certain clonal populations, usually the more aggressive and faster proliferating ones, emerge and overtake the previous ones. Tomlinson and Bodmer were the first to explore the role of cell–cell interactions in cancer using EGT [20]. This pioneering work was followed by others that built

on those initial ideas to study the role of key aspects of cancer evolution, such as the role of space [21] treatment [22,23] or metabolism [10,24].

Work by Kianercy and colleagues [10] shows how microenvironmental heterogeneity impacts somatic evolution. Kianercy and colleagues show how the tumor's genetic instability adapts to the heterogeneous microenvironment (with regard to oxygen concentration) to better tune metabolism to the dynamic microenvironment. While evolutionary dynamics can help a tumor population evolve to acquire all relevant mutations to become an aggressive cancer [25], they also help them become treatment-resistant, which leads to treatment failure as well as increased toxicity for the patient, which can result in patient death. Researchers such as Axelrod and colleagues [26] have speculated that tumor cells do not need to acquire all the hallmarks of cancer to become an aggressive cancer but that the cooperation between different cells with different abilities might allow the tumor as a whole to acquire all the hallmarks. A few years ago, Hanahan and Weinberg updated their original research to include disregulated metabolisms as one of the hallmarks of cancer [27]. Here we suggest that cooperation between cells with different metabolic needs and abilities could allow the tumor to grow faster but also present a new therapeutical target that could be clinically exploited. Namely, this cooperation, as described by Kinaercy and colleagues, allows for hypoxic cells to benefit from the presence of oxygenated non-glycolytic cells with modest glucose requirements, whereas cells with aerobic metabolism can benefit from the lactic acids that are the byproduct of anaerobic metabolism (see Figure 13).

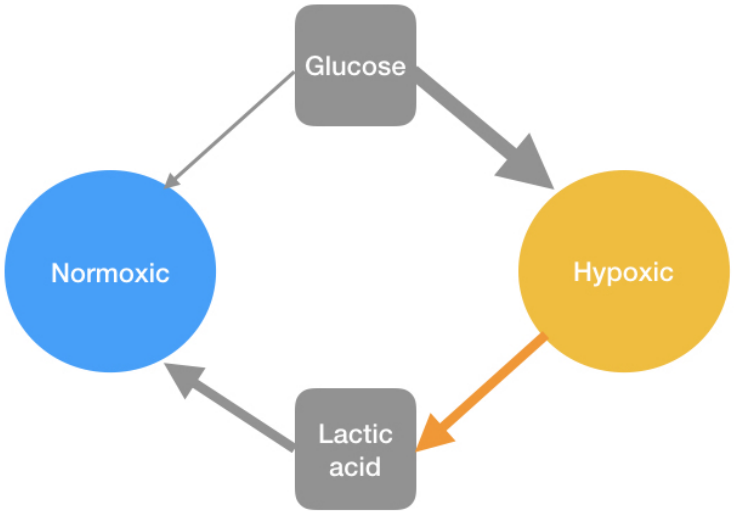

Figure 13. Interaction diagram between different type of cells. The hypoxic cells can benefit from the presence of oxygenated non-glycolytic cells with modest glucose requirements, whereas cells with aerobic metabolism can benefit from the lactic acids that are the byproduct of anaerobic metabolism.

By targeting this cooperation, a tumor's growth and progression could be disrupted using novel microenvironmental pH normalizers. What our work suggests is that small perturbations could return the system back to a state different from the one it started so that the microenvironmental impact does not need to be too substantial for the therapy to have an impact. The work we have described here supports the hypothesis that hysteresis would allow us to apply treatments for a short duration of time with the aim of changing the nature of the game instead of killing tumor cells. This would have the combined advantages of reducing toxicity and side effects and decreasing selection for resistant tumor phenotypes and thus reducing the emergence of resistance to the treatment. For instance, treatments

that aim to reduce the acidity of the environment [28] would impact not only acid producing cells but also the acid-resistant normoxic ones.

Our techniques (the hysteresis mechanism and the optimal control mechanism) can be applied to the cancer game [10] with two types of tumor phenotypic strategies: hypoxic cells and oxygenated cells (Table 1). These cells inhabit regions where oxygen could be either abundant or lacking. In the former, oxygenated cells with regular metabolism thrive but in the latter, hypoxic cells whose metabolism is less reliant on the presence of oxygen (but more on the presence of glucose) have higher fitness.

Table 1. Payoff matrix for the cancer game in [10], where $L > G_o/2$. This 2×2 game represents the tumor metabolic symbiosis rewards (ATP generation). The row agent represents hypoxic cells, and the column one represents oxygenated cell energy generation values based on their collective actions. Specifically, oxygenated cells can use both glucose and lactate for energy generation, whereas the hypoxic cells can use only glucose. Empirical data as discussed in [10] suggests that $L > Go/2$.

Hypoxic/Oxygenated	**Glucose**	**Lactate**
Glucose	$G_h/2, G_o/2$	G_h, L
Lactate	$0, G_o$	$0, 0$

6.2. Taxation

A direct application for the solution concept of QRE is to analyze the effect of taxation, which has been discussed in [9]. Unlike Nash equilibria, for QREs, if we multiply the payoff matrix by some factor α, the equilibrium does change. This is because, by multiplying α, effectively we are dividing the temperature parameters by α. This means that, if we charge taxes to the players with some flat tax rate $\alpha - 1$, the QREs will differ. Formally, we define the base temperature T_0 as the temperature when no tax is applied for both players. Then, we can define the *tax rate* for each player as $\alpha_x = 1 - T_0/T_x, \alpha_y = 1 - T_0/T_y$, respectively.

We demonstrate how the hysteresis mechanism can be applied in a 2×2 game via taxation with Example 1. Recall that in Example 1, we have two types of agents. We can consider these two types of agents as corresponding to two different sectors of the economy (e.g., aircraft manufacturing versus car manufacturing), which need to coordinate on their choice between two different competing technologies that are related to both sectors (e.g., 3D-printing). We can consider the row player as being the aircraft manufacturer and the column player as being the car manufacturer, with payoff matrices specified in Table 2. By assuming both players are of bounded rationality with temperature 1, we assume the base temperatures for both players are $T_0 = 1$. In this game, the equilibrium where both players choose Technology 1 has greater social welfare than the equilibrium where both players choose Technology 2. Consider the situation where, initially, the system is in an equilibrium state where both players choose Technology 2 with high probability. Then, with taxation, we have shown in the previous sections that we are able to increase the social welfare via the hysteresis mechanism or the optimal control mechanism. Here, we demonstrate how the simplified process that we have described in Example 1 can improve the social welfare in this game (see Figure 2 for the bifurcation diagram of this game):

1. The initial state is $(0.05, 0.14)$, where the row agent chooses Technology 1 with probability 0.05 and the column agent chooses Technology 1 with probability 0.14. This is an equilibrium state when we impose the tax rate $\alpha_x \approx 0$ to the row agent and the tax rate $\alpha_y \approx 0.5$ to the column agent (where $T_x \approx 1$ and $T_y \approx 2$).
2. Fix the tax rate for the column agent at $\alpha_y = 0.5$ (where $T_y = 2$) and increase the tax rate for the row agent to $\alpha_x = 0.8$ (where $T_x = 5$). Under this assignment of tax rates, there is only one QRE correspondence.

3. Fix the tax rate for the column agent at $\alpha_y = 0.5$ (where $T_y = 2$) and decrease the tax rate α_x for the row agent back to 0 (where $T_x = 1$). Now $x \approx 0.997$, where both agents choose Action 1 with high probability.

Table 2. Payoff matrix for a coordination game between two agents where neither of the two pure Nash Pareto dominates the other. States where both agents play the first strategy (Technology 1) are nearly socially optimal and they can be selected via a bifurcation argument.

Sector A/Sector B	Technology 1	Technology 2
Technology 1	10, 2	0, 0
Technology 2	0, 0	5, 4

In [9], they considered three approaches—"anarchy," "socialism," and "market"—of how the taxes can be dynamically adjusted by the society, depending on whether the taxes are determined in a decentralized manner, by an external regulator, or through bargaining, respectively. The concept of our mechanisms is a variant of the "socialism" scheme since in our model the mechanism, who can be thought as an external regulator, determines the tax rates. Our mechanisms are systematic approaches that optimize an objective where, in [9], the trajectories toward maximizing expected utilities are considered.

7. Connection to Previous Works

Recently, there has been a growing interplay between game theory, dynamical systems, and computer science. Examples include the integration of replicator dynamics and topological tools [29–31] in algorithmic game theory, and Q-learning dynamics [5] in multi-agent systems [6]. Q-learning dynamics has been studied extensively in game settings, e.g., by Sato et al. in [13] and Tuyls et al. in [14]. In [12], Q-learning dynamics is considered as an extension of replicator dynamics driven by a combination of payoffs and entropy. Recent advances in our understanding of evolutionary dynamics in multi-agent learning can be found in the survey in [32].

We are particularly interested in the connection between Q-learning dynamics and the concept of QRE [7] in game theory. In [11], Cominetti et al. study this connection in traffic congestion games. The hysteresis effect of Q-learning dynamics was first identified in 2012 by Wolpert et al. [9]. Kianercy et al. in [16] observed the same phenomenon and provided discussions on bifurcation diagrams in 2×2 games. The hysteresis effect has also been highlighted in recent follow-up work by [10] as a design principle for future cancer treatments. It was also studied in [33] in the context of minimum-effort coordination games. However, our current understanding is still mostly qualitative and in this work we have pushed towards a more practically applicable, quantitative, and algorithmic analysis.

Analyzing the characteristics of various dynamical systems has also been attracting the attention of computer science community in recent years. For example, besides the Q-learning dynamics, the (simpler) replicator dynamics has been studied extensively due to its connections [30,34,35] to the multiplicative weight update (MWU) algorithm in [36].

Much attention has also been devoted to biological systems and their connections to game theory and computation. In recent work by Mehta et al. [37], the connection with genetic diversity was discussed in terms of the complexity of predicting whether genetic diversity persists in the long run under evolutionary pressures. This paper builds upon a rapid sequence of related results [38–43]. The key result is [39,40], where it was made clear that there is a strong connection between studying replicator dynamics in games and standard models of evolution. Follow-up works show how dynamics that incorporate errors (i.e., mutations) can be analyzed [44] and how such mutations can have a critical effect on ensuring survival in the presence of dynamically changing environments. Our paper makes progress along these lines by examining how noisy dynamics can introduce, for example, bifurcations.

We were inspired by recent work by Kianercy et al. establishing a connection between cancer dynamics and cancer treatment and studying Q-learning dynamics in games. This is analogous to the connections [39,40,45] between MWU and evolution detailed above. It is our hope that by starting off a quantitative analysis of these systems we can kickstart similarly rapid developments in our understanding of the related questions.

8. Conclusions

In this paper, we perform a quantitative analysis of bifurcation phenomena connected to Q-learning dynamics in 2×2 games. Based on this analysis, we introduce two novel mechanisms, the hysteresis mechanism and the optimal control mechanism. Hysteresis mechanisms use transient changes to the system parameters to induce permanent improvements to its performance via optimal (Nash) equilibrium selection. Optimal control mechanisms induce convergence to states whose performance is better than the best Nash equilibrium, showing that by controlling the exploration/exploitation tradeoff, we can achieve strictly better states than those achievable by perfectly rational agents.

We believe that these new classes of mechanisms could lead to interesting new questions within game theory. Importantly they could also lead to a more thorough understanding of cancer biology and how treatments could be designed not to kill tumor cells but to induce transient changes in the game with long-lasting consequences, impacting the equilibrium in ways that would be therapeutically useful.

Author Contributions: G.Y. worked on the analysis, experiments, figures and writeup, D.B. worked on the writeup, G.P. proposed the research direction and worked on the analysis and writeup.

Acknowledgments: Georgios Piliouras would like to acknowledge SUTD grant SRG ESD 2015 097 and MOE AcRF Tier 2 Grant 2016-T2-1-170 and an NRF 2018 Fellowship (NRF-NRFF2018-07). Ger Yang is supported in part by NSF grant numbers CCF-1216103, CCF-1350823, CCF-1331863, and CCF-1733832. David Basanta is partly funded by an NCI U01 (NCI) U01CA202958-01. Part of the work was completed while Ger Yang and Georgios Piliouras were visiting scientists at the Simons Institute for the Theory of Computing.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. From Q-Learning to Q-Learning Dynamics

In this section, we provide a quick sketch on how we can get to the Q-learning dynamics from Q-learning agents. We start with an introduction to the Q-learning rule. Then, we discuss the multi-agent model when there are multiple learners in the system. The goal of this section is to identify the dynamics of the system in which there are two learning agents playing a 2×2 game repeatedly over time.

Appendix A.1. Q-Learning Introduction

Q-learning [4,5] is a value-iteration method for solving the optimal strategies in Markov decision processes. It can be used as a model where users learn about their optimal strategy when facing uncertainties. Consider a system that consists of a finite number of states and there is one player who has a finite number of actions. The player is going to decide his/her strategy over an infinite time horizon. In Q-learning, at each time t, the player stores a value estimate $Q_{(s,a)}(t)$ for the payoff of each state–action pair (s, a). S/he chooses his/her action a_{t+1} that maximizes the Q-value $Q_{(s_t,\cdot)}(t)$ for time $t+1$, given the system state is s_t at time t. In the next time step, if the agent plays action a_{t+1}, s/he will receive a reward $r(t+1)$, and the value estimate is updated according to the rule:

$$Q_{(s_t,a_{t+1})}(t+1) = (1-\alpha)Q_{(s_t,a_{t+1})}(t) + \alpha(r(t+1) + \gamma \max_{a'} Q_{(s_{t+1},a')}(t))$$

where α is the step size, and γ is the discount factor.

Appendix A.2. Joint-Learning Model

Next, we consider the joint learning model as in [16]. Suppose there are multiple players in the system that are learning concurrently. Denote the set of players as P. We assume the system state is a function of the action each player is playing, and the reward observed by each player is a function of the system state. Their learning behaviors are modeled as simplified models based on the Q-learning algorithm described above. More precisely, we consider the case where each player assumes the system is only of one state, which corresponds to the case where the player has very limited memory and has discount factor $\gamma = 0$. The reward observed by player $i \in P$ given s/he plays action a at time t is denoted as $r_a^i(t)$. We can write the updating rule of the Q-value for agent i as follows:

$$Q_a^i(t+1) = Q_a^i(t) + \alpha[r_a^i(t) - Q_a^i(t)].$$

For the selection process, we consider the mechanism that each player $i \in P$ selects his/her action according to the Boltzmann distribution with temperature T_i:

$$x_a^i(t) = \frac{e^{Q_a^i(t)/T_i}}{\sum_{a'} e^{Q_{a'}^i(t)/T_i}} \tag{A1}$$

where $x_a^i(t)$ is the probability that agent i chooses action a at time t. The intuition behind this mechanism is that we are modeling the irrationality of the users by the temperature parameter T_i. For small T_i, the selection rule corresponds to the case of more rational agents. We can see that for $T_i \to 0$, (A1) corresponds to the best-response rule, that is, each agent selects the action with the highest Q-value with probability one. On the other hand, for $T_i \to \infty$, we can see that Equation (A1) corresponds to the selection rule of selecting each action uniformly at random, which models the case of fully irrational agents.

Appendix A.3. Continuous-Time Dynamics

This underlying Q-learning model has been studied in the previous decades. It is known that if we take the time interval to be infinitely small, this sequential joint learning process can be approximated as a continuous-time model ([13,14]) that has some interesting characteristics. To see this, consider the 2×2 game as we have described in Section 2.1. The expected payoff for the first player at time t given s/he chooses action a can be written as $r_a^x(t) = [\boldsymbol{Ay}(t)]_a$; similarly, the expected payoff for the second player at time t given s/he chooses action a is $r_a^y(t) = [\boldsymbol{Bx}(t)]_a$. The continuous-time limit for the evolution of the Q-value for each player can be written as

$$\begin{aligned} \dot{Q}_a^x(t) &= \alpha[r_a^x(t) - Q_a^x(t)] \\ \dot{Q}_a^y(t) &= \alpha[r_a^y(t) - Q_a^y(t)]. \end{aligned}$$

Then, we take the time derivative of Equation (A1) for each player to obtain the evolution of the strategy profile:

$$\begin{aligned} \dot{x}_i &= \frac{1}{T_x} x_i \left(\dot{Q}_i^x - \sum_k x_k \dot{Q}_k^x \right) \\ \dot{y}_i &= \frac{1}{T_y} y_i \left(\dot{Q}_i^y - \sum_k y_k \dot{Q}_k^y \right). \end{aligned}$$

Putting these together, and rescaling the time horizon to $\alpha t/T_x$ and $\alpha t/T_y$ respectively, we obtain the continuous-time dynamics:

$$\dot{x}_i = x_i\left[(\boldsymbol{Ay})_i - \boldsymbol{x}^T\boldsymbol{Ay} + T_x\sum_j x_j\ln(x_j/x_i)\right] \tag{A2}$$

$$\dot{y}_i = y_i\left[(\boldsymbol{Bx})_i - \boldsymbol{y}^T\boldsymbol{Bx} + T_y\sum_j y_j\ln(y_j/y_i)\right]. \tag{A3}$$

Appendix A.4. The Exploration Term Increases Entropy

Now, we show that the exploration term in the Q-learning dynamics results in the increase of the entropy:

Lemma A1. *Suppose $A = \mathbf{0}$ and $B = \mathbf{0}$. The system entropy*

$$H(\boldsymbol{x},\boldsymbol{y}) = H(\boldsymbol{x}) + H(\boldsymbol{y}) = -\sum_i x_i\ln x_i - \sum_i y_i\ln y_i$$

for the dynamics (2) increases with time, i.e.,

$$\dot{H}(\boldsymbol{x},\boldsymbol{y}) > 0$$

if $\boldsymbol{x}$ and $\boldsymbol{y}$ are not uniformly distributed.

Proof of Lemma A1. It is equivalent that we consider the single agent dynamics:

$$\dot{x}_i = x_iT_x\left[-\ln x_i + \sum_j x_j\ln x_j\right].$$

Taking the derivative of the entropy $H(\boldsymbol{x})$, we have

$$\dot{H}(\boldsymbol{x}) = \sum_i(-\ln x_i - 1)\dot{x}_i = -T_x\left[-\sum_i x_i(\ln x_i)^2 + \left(\sum_j x_i\ln x_i\right)^2\right],$$

and since we have $\sum_i x_i = 1$, by Jensen's inequality, we can find that

$$\left(\sum_j x_i\ln x_i\right)^2 \le \sum_i x_i(\ln x_i)^2$$

where equality holds if and only if $\boldsymbol{x}$ is a uniform distribution. Consequently, if we have $x_i \in (0,1)$, and $\boldsymbol{x}$ is not a uniform distribution, $\dot{H}(\boldsymbol{x})$ is strictly positive, which means that the system entropy increases with time. □

Appendix B. Convergence of Dissipative Learning Dynamics in 2×2 Games

Appendix B.1. Liouville's Formula

Liouville's formula can be applied to any system of autonomous differential equations with a continuously differentiable vector field V on an open domain of $\mathcal{S}\subset\mathbb{R}^k$. The divergence of V at $x\in\mathcal{S}$ is defined as the trace of the corresponding Jacobian at x, i.e., $\mathrm{div}[V(x)] \equiv \sum_{i=1}^k \frac{\partial V_i}{\partial x_i}(x) = tr(DV(x))$. Since divergence is a continuous function we can compute its integral over measurable sets $A\subset\mathcal{S}$ (with respect to Lebesgue measure μ on $\mathbb{R}^n$). Given any such set A, let $\phi_t(A) = \{\phi(x_0,t) : x_0\in A\}$ be the image of A under map Φ at time t. $\phi_t(A)$ is measurable and its measure is $\mu(\phi_t(A))) = \int_{\phi_t(A)} dx$.

Liouville's formula states that the time derivative of the volume $\phi_t(A)$ exists and is equal to the integral of the divergence over $\phi_t(A)$: $\frac{d}{dt}[A(t)] = \int_{\phi_t(A)} \text{div}[V(x)]dx$. Equivalently,

Theorem A1 ([46], p. 356). $\frac{d}{dt}\mu(\phi_t(A)) = \int_{\phi_t(A)} tr(DV(x))d\mu(x)$.

A vector field is called divergence free if its divergence is zero everywhere. Liouville's formula trivially implies that volume is preserved in such flows.

This theorem extends in a straightforward manner to systems where the vector field $V : X \to TX$ is defined on an affine set $X \subset \mathbb{R}^n$ with tangent space TX. In this case, μ represents the Lebesgue measure on the (affine hull) of X. Note that the derivative of V at a state $x \in X$ must be represented using the derivate matrix $DV(x) \in \mathbb{R}^{n\times n}$, which by definitions has rows in TX. If $\hat{V} : \mathbb{R}^n \to R^n$ is a C^1 extension of V, then $DV(x) = D\hat{V}(x)P_{TX}$, where $P_{TX} \in \mathbb{R}^{n\times n}$ is the orthogonal projection[2] of $\mathbb{R}^n$ onto the subspace TX.

Appendix B.2. Poincaré–Bendixson Theorem

The Poincaré–Bendixson theorem is a powerful theorem that implies that two-dimensional systems cannot effectively exhibit chaos. Effectively, the limit behavior is either going to be an equilibrium, a periodic orbit, or a closed loop, punctuated by one (or more) fixed points. Formally, we have

Theorem A2 ([47,48]). *Given a differentiable real dynamical system defined on an open subset of the plane, then every non-empty compact ω-limit set of an orbit, which contains only finitely many fixed points, is either a fixed point, a periodic orbit, or a connected set composed of a finite number of fixed points together with homoclinic and heteroclinic orbits connecting these.*

Appendix B.3. Bendixson–Dulac Theorem

By excluding the possibility of closed loops (i.e., periodic orbits, homoclinic cycles, and heteronclinic cycles) we can effectively establish global convergence to equilibrium. The following criterion, which was first established by Bendixson in 1901 and further refined by French mathematician Dulac in 1933, allows us to do that. It is typically referred to as the Bendixson–Dulac negative criterion. It focuses exactly on the planar system where the measure of initial conditions always shrinks (or always increases) with time, i.e., dynamical systems with vector fields whose divergence is always negative (or always positive).

Theorem A3 ([49], p. 210). *Let $D \subset \mathbb{R}^2$ be a simply connected region and (f, g) in $C^1(D, \mathbb{R})$ with $div(f, g) = \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y}$ being not identically zero and without a change of sign in D. Then the system*

$$\frac{dx}{dt} = f(x, y)$$

$$\frac{dy}{dt} = g(x, y)$$

has no loops lying entirely in D.

The function $\varphi(x, y)$ is typically called the Dulac function.

---

[2] To find the matrix of the orthogonal projection onto TX (or any subspace Y of $\mathbb{R}^n$) it suffices to find a basis $(\vec{v_1}, \vec{v_2}, \ldots, \vec{v_m})$. Let B be the matrix with columns $\vec{v_i}$; then $P = B(B^TB)^{-1}B^T$.

Remark A1. *This criterion can also be generalized. Specifically, it holds for the system:*

$$\frac{dx}{dt} = \rho(x,y)f(x,y)$$

$$\frac{dy}{dt} = \rho(x,y)g(x,y)$$

if $\rho(x,y) > 0$ is continuously differentiable. Effectively, we are allowed to rescale the vector field by a scalar function (as long as this function does not have any zeros), before we prove that the divergence is positive (or negative). That is, it suffices to find $\rho(x,y) > 0$ continuously differentiable, such that $(\rho(x,y)f(x,y))_x + (\rho(x,y)g(x,y))_y$ possesses a fixed sign.

By [16], after a change of variables, $u_k = \frac{\ln(x_{k+1})}{\ln x_1}$, $v_k = \frac{\ln(y_{k+1})}{\ln y_1}$ for $k = 1, \ldots, n-1$, the replicator system transforms to the following system:

$$\dot{u}_k = \frac{\sum_j \hat{a}_k j e^{v_j}}{1 + \sum_j e^{v_j}} - T_x u_k, \dot{v}_k = \frac{\sum_j \hat{a}_k j e^{u_j}}{1 + \sum_j e^{u_j}} - T_x v_k, \text{(II)}$$

where $\hat{a}_{kj} = a_{k+1,j+1} - a_{1,j+1}$, $\hat{b}_{kj} = b_{k+1,j+1} - a_{1,j+1}$.

In the case of 2×2 games, we can apply both the Poincaré–Bendixson theorem as well as the Bendixson–Dulac theorem, since the resulting dynamical system is planar and $\frac{\partial \dot{u}_1}{\partial u_1} + \frac{\partial \dot{v}_1}{\partial v_1} = -(T_x + T_y) < 0$. Hence, for any initial condition system, (II) converges to equilibria. The flow of the original replicator system in the 2×2 game is *diffeomorhpic*[3] to the flow of System (II); thus, the replicator dynamics with positive temperatures T_x, T_y converges to equilibria for all initial conditions as well.

Appendix C. Bifurcation Analysis for Games with Only One Nash Equilibrium

In this section, we present the results for the class of games with only one Nash equilibrium, where it can be either a pure one or a mixed one, where the mixed Nash equilibrium is defined as follows.

Definition A1 (Mixed Nash equilibrium)**.** *A strategy profile (x_{NE}, y_{NE}) is a mixed Nash equilibrium if*

$$x_{NE} \in \arg\max_{x \in [0,1]} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{y}_{NE} \qquad\qquad y_{NE} \in \arg\max_{y \in [0,1]} \boldsymbol{y}^T \boldsymbol{B} \boldsymbol{x}_{NE}.$$

This corresponds to the case where b_X, a_Y, or b_Y is negative. Similarly, our analysis is based on the second form representation described in Equations (6) and (7), which demonstrates insights from the first player's perspective.

Appendix C.1. No Dominating Strategy for the First Player

More specifically, this is the case when there is no dominating strategy for the first player, i.e., both a_X and b_X are positive. From Equation (7), we can presume that the characteristics of the bifurcation diagrams depend on the value of $a_Y + b_Y$ since it affects whether y^{II} is increasing with x or not. Additionally, we can find some interesting phenomenon from the discussion below.

First, we consider the case when $a_Y + b_Y > 0$. This can be considered as a more general case as we have discussed in Section 4.3. In fact, the statements we have made in Theorems 1–3 applies to this case. However, there are some subtle difference that should be noticed. If $a_Y > b_Y$, where we can

---

[3] A function f between two topological spaces is called a *diffeomorphism* if it has the following properties: f is a bijection, f is continuously differentiable, and f has a continuously differentiable inverse. Two flows $\Phi^t : A \to A$ and $\Psi^t : B \to B$ are *diffeomorhpic* if there exists a diffeomorphism $g : A \to B$ such that for each $x \in A$ and $t \in \mathbb{R}$ $g(\Phi^t(x)) = \Psi^t(g(x))$. If two flows are diffeomorphic, then their vector fields are related by the derivative of the conjugacy. That is, we get precisely the same result that we would have obtained if we simply transformed the coordinates in their differential equations [50].

assume $b_Y < 0$, then by the second part of Theorem 2, there are no QREs in $x \in (0, 0.5)$, since T_B is now a negative number. This means that we always only have the principal branch. On the other hand, if $a_Y < b_Y$, where we can assume $a_Y < 0$, then, similar to the example in Figures 4 and 5, there could still be two branches. However, we can presume that the second branch vanishes *before* T_y actually goes to zero, as the state $(1, 1)$ is not a Nash equilibrium.

Theorem A4. *Given a* 2×2 *game in which the diagonal form has* $a_X, b_X > 0$, $a_Y + b_Y > 0$, *and* $a_Y < b_Y$, *and given* T_y, *if* $T_y < T_A$, *where* $T_A = \frac{-a_Y}{\ln(a_Y/b_Y)}$, *then there is no QRE correspondence in* $x \in (0.5, 1)$.

The proof of the above theorem directly follows from Proposition A4 in the appendix. An interesting observation here is that we can still make the first player achieve his/her desired state by changing T_y to some value that is greater than T_A.

Next, we consider $a_Y + b_Y \leq 0$. The bifurcation diagram is illustrated in Figures A1 and A2. We can find that in this case the principal branch directly goes toward its unique Nash equilibrium. We present the results formally in the following theorem, where the proof follows from Appendix D.1.2 in the appendix.

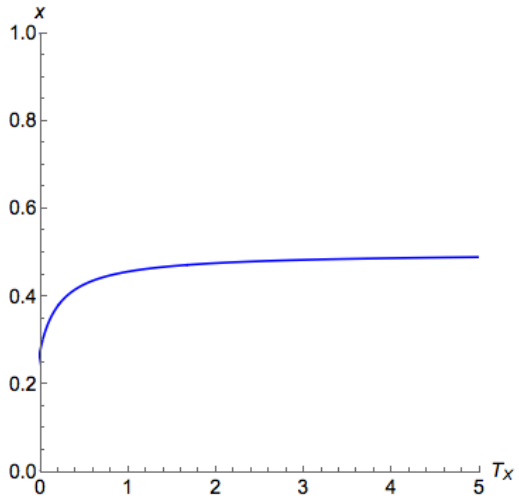

Figure A1. Bifurcation diagram for a game with no dominating strategy for the first player, $a_Y + b_Y < 0$, and a low T_Y.

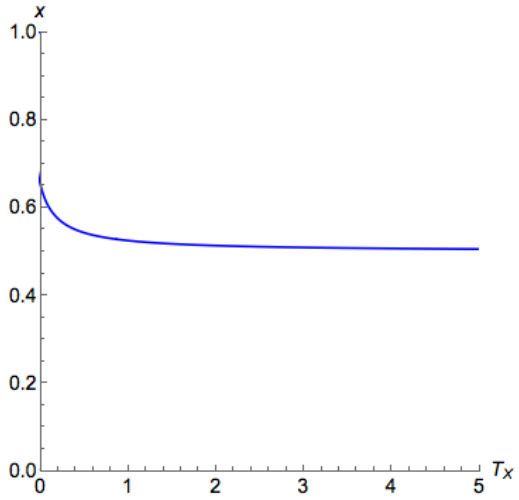

Figure A2. Bifurcation diagram for a game with no dominating strategy for the first player, $a_Y + b_Y < 0$, and a high T_Y.

Theorem A5. *Given a* 2×2 *game in which the diagonal form has* $a_X, b_X > 0$, $a_Y + b_Y \leq 0$, *QRE is unique given* T_x *and* T_y.

Appendix C.2. Dominating Strategy for the First Player

Finally, we consider the case when there is a dominating strategy for the first player, i.e., $b_X < 0$. According to Figures A3 and A4, the principal branch seems always goes towards $x = 1$. This means that the first player always prefers his/her dominating strategy. We formalize this observation, as well as some important characteristics for this case in the theorem below, where the proof can be found in Appendix D.2 in the appendix.

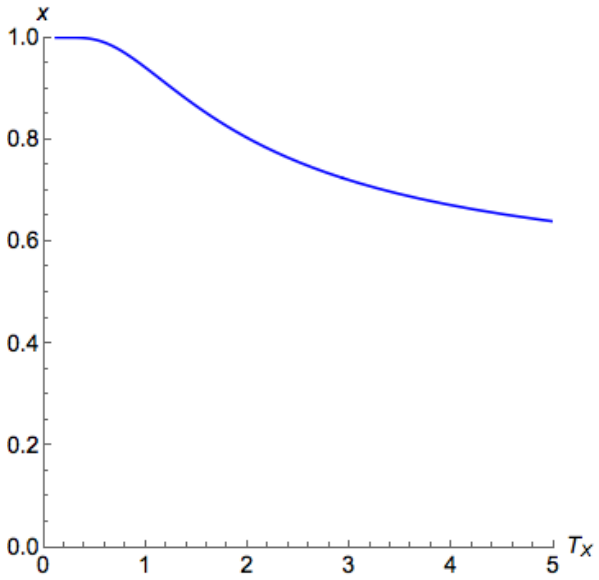

Figure A3. Bifurcation diagram for a game with one dominating strategy for the first player and $a_Y + b_Y < 0$.

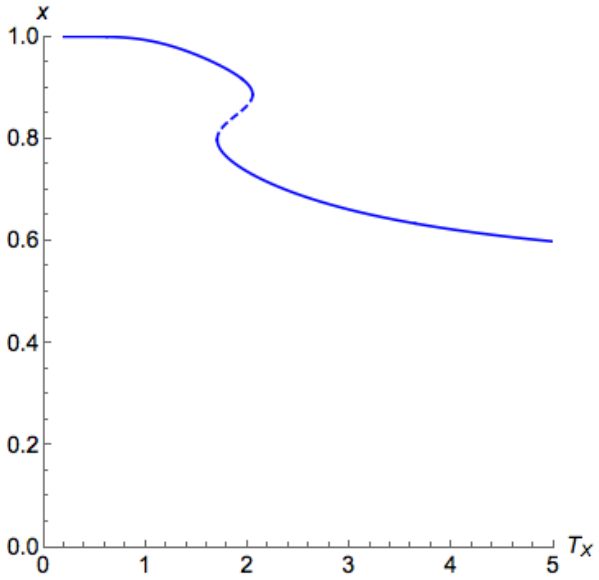

Figure A4. Bifurcation diagram for a game with one dominating strategy for the first player, $a_Y + b_Y > 0$, and $a_Y < b_Y$.

Theorem A6. *Given a* 2×2 *game in which the diagonal form has* $a_X > 0$, $b_X < 0$, $a_X + b_X > 0$, *and, given* T_y, *the following statements are true:*

1. *The region* $(0, 0.5)$ *contains the principal branch.*
2. *There is no QRE correspondence for* $x \in (0.5, 1)$.
3. *If* $a_Y + b_Y < 0$ *or* $a_Y > b_Y$, *then the principal branch is continuous.*
4. *If* $a_Y + b_Y > 0$ *and* $b_Y > a_Y$, *then the principal branch may not be continuous.*

As we can see from Theorem A6, for most cases, the principal branch is continuous. One special case is when $a_Y + b_Y > 0$ with $b_Y > a_Y$. In fact, this can be seen as a duality, i.e., flipping the role of two players, of the case we have discussed in Part 3 of Theorem A4, where, if T_y is within T_A and T_I, there can be three QRE correspondences.

Appendix D. Detailed Bifurcation Analysis for General 2×2 Game

In this section, we provide technical details for the results we stated in Section 4.3 and Appendix C. Before we get into details, we state some results that will be useful throughout the analysis in the following lemma. The proof of this lemma is straightforward and we omit it in this paper.

Lemma A2. *The following statements are true.*

1. *The derivative of* T_X^{II} *is given as*

$$\frac{\partial T_X^{II}}{\partial x}(x, T_y) = \frac{-(a_X + b_X)L(x, T_y) + b_X}{x(1-x)[\ln(1/x - 1)]^2} \tag{A4}$$

where

$$L(x, T_y) = y^{II} + x(1-x)\ln\left(\frac{1}{x} - 1\right)\frac{\partial y^{II}}{\partial x}. \tag{A5}$$

2. *The derivative of* y^{II} *is given as*

$$\frac{\partial y^{II}}{\partial x} = y^{II}(1 - y^{II})\frac{a_Y + b_Y}{T_y}.$$

3. *For* $x \in (0, 1/2) \cup (1/2, 1)$, $\frac{\partial T_X^{II}}{\partial x} > 0$ *if and only if* $L(x, T_y) < \frac{b_X}{a_X + b_X}$; *on the other hand,* $\frac{\partial T_X^{II}}{\partial x} < 0$ *if and only if* $L(x, T_y) > \frac{b_X}{a_X + b_X}$.

Appendix D.1. Case 1: $b_X \geq 0$

First, we consider the case $b_X \geq 0$. As we are going to show in Proposition A1, the direction of the principal branch relies on $y^{II}(0.5, T_y)$, which is the strategy the second player is performing, assuming the first player is indifferent to his/her payoff. The idea is that if $y^{II}(0.5, T_y)$ is large, then it means that the second player pays more attention to the action that the first player thinks is better. This is more likely to happen when the second player has less rationality, i.e., high temperature T_y. On the other hand, if the second player pays more attention to the other action, the first player is forced to choose that as it gets more expected payoff.

We show that, for $T_y > T_I$, the principal branch lies on $x \in \left(\frac{1}{2}, 1\right)$; otherwise, the principal branch lies on $x \in \left(0, \frac{1}{2}\right)$. This result follows from the following proposition:

Proposition A1. *For Case 1, if* $T_y > T_I$, *then* $y^{II}(1/2, T_y) > \frac{b_X}{a_X + b_X}$; *hence,*

$$\lim_{x \to \frac{1}{2}^+} T_X^{II}(x, T_y) = +\infty \quad \textit{and} \quad \lim_{x \to \frac{1}{2}^-} T_X^{II}(x, T_y) = -\infty.$$

On the other hand, if $T_y < T_I$, *then* $y^{II}(1/2, T_y) < \frac{b_X}{a_X+b_X}$; *hence,*

$$\lim_{x\to\frac{1}{2}^+} T_X^{II}(x, T_y) = -\infty \quad \textit{and} \quad \lim_{x\to\frac{1}{2}^-} T_X^{II}(x, T_y) = +\infty.$$

Proof. First, consider the case where $b_Y > a_Y$, then, for $T_y > T_I = \frac{b_Y - a_Y}{2\ln(a_X/b_X)}$,

$$y^{II}\left(\frac{1}{2}, T_y\right) = \left(1 + e^{\frac{b_Y-a_Y}{2T_y}}\right)^{-1} > \left(1 + e^{\frac{b_Y-a_Y}{2T_I}}\right)^{-1} = \left(1 + \frac{a_X}{b_X}\right)^{-1} = \frac{b_X}{a_X + b_X}.$$

Then, for the case where $a_Y > b_Y$,

$$y^{II}\left(\frac{1}{2}, T_y\right) = \left(1 + e^{\frac{b_Y-a_Y}{2T_y}}\right)^{-1} > \left(1 + e^{0}\right)^{-1} = \frac{1}{2} \geq \frac{b_X}{a_X + b_X}.$$

For the case where $a_Y = b_Y$, since we assumed $a_X \neq b_X$,

$$y^{II}\left(\frac{1}{2}, T_y\right) = \left(1 + e^{\frac{b_Y-a_Y}{2T_y}}\right)^{-1} = \left(1 + e^{0}\right)^{-1} = \frac{1}{2} > \frac{b_X}{a_X + b_X}.$$

As a result, the numerator of Equation (6) at $x = \frac{1}{2}$ is negative for $T_y > T_I$, which proves the first two limits.

For the remaining two limits, we only need to consider the case $b_Y > a_Y$; otherwise, $T_I = 0$, which is meaningless. For $b_Y > a_Y$ and $T_y < T_I$,

$$y^{II}\left(\frac{1}{2}, T_y\right) = \left(1 + e^{\frac{b_Y-a_Y}{2T_y}}\right)^{-1} < \left(1 + e^{\frac{b_Y-a_Y}{2T_I}}\right)^{-1} = \left(1 + \frac{a_X}{b_X}\right)^{-1} = \frac{b_X}{a_X + b_X}.$$

This makes the numerator of Equation (6) at $x = \frac{1}{2}$ positive and proves the last two limits. □

Appendix D.1.1. Case 1a: $b_X \geq 0$, $a_Y + b_Y > 0$

In this section, we consider a relaxed version of the class of coordination game as in Section 4.3. We prove theorems presented in Section 4.3, showing that these results can in fact be extended to the case where $a_Y + b_Y > 0$, instead of requiring $a_Y > 0$ and $b_Y > 0$.

First, $a_Y + b_Y > 0$, y^{II} is an increasing function of x, meaning

$$\frac{\partial y^{II}}{\partial x} = y^{II}(1 - y^{II})\frac{a_Y + b_Y}{T_y} > 0.$$

This implies that both players tend to agree to each other. Intuitively, if $a_Y \geq b_Y$, then both players agree that the first action is the better one. For this case, we can show that, no matter what T_y is, the principal branch lies on $x \in \left(\frac{1}{2}, 1\right)$. In fact, this can be extended to the case whenever $T_y > T_I$, which is the first part of Theorem 1.

Proof of Part 1 of Theorem 1. We can find that, for $T_y > T_I$, $y^{II}(1/2, T_Y) > \frac{b_X}{a_X+b_X}$ for any T_y according to Proposition A1. Since y^{II} is monotonically increasing with x, $y^{II} > \frac{b_X}{a_X+b_X}$ for $x > 1/2$. This means that $T_X^{II} > 0$ for any $x \in (1/2, 1)$. Additionally, it is easy to see that $\lim_{x\to 1^-} T_X^{II} = 0$. As a result, $(0.5, 1)$ contains the principal branch. □

For Case 1a with $a_Y \geq b_Y$, on the principal branch, the lower the T_x, the closer x is to 1. We are able to show these monotonicity characteristics in Proposition A2, and they can be used to justify the stability owing to Lemma 1.

Proposition A2. *In Case 1a, if* $a_Y \geq b_Y$, *then* $\frac{\partial T_X^{II}}{\partial x} < 0$ *for* $x \in \left(\frac{1}{2}, 1\right)$.

Proof. It suffices to show that $L(x, T_y) > \frac{b_X}{a_X + b_X}$ for $x \in \left(\frac{1}{2}, 1\right)$. Note that, according to Proposition A1, if $a_Y \geq b_Y$,

$$L(1/2, T_y) = y^{II}(1/2, T_y) \geq \frac{1}{2} \geq \frac{b_X}{a_X + b_X}. \tag{A6}$$

Since $y^{II}(x, T_y)$ is monotonically increasing when $a_Y + b_Y > 0$, $y^{II}(x, T_y) > \frac{1}{2}$ for $x \in \left(\frac{1}{2}, 1\right)$. As a result, $1 - 2y^{II} < 0$; hence, we can see that, for $x \in \left(\frac{1}{2}, 1\right)$,

$$\frac{\partial L}{\partial x} = \left[(1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y}\right] \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}}{\partial x} > 0.$$

Consequently, for $x \in \left(\frac{1}{2}, 1\right)$, $L(x, T_y) > \frac{b_X}{a_X + b_X}$; hence, $\frac{\partial T_X^{II}}{\partial x} < 0$ according to Lemma A2. □

Proof of Part 1 of Theorem 3. According to Lemma 1, Proposition A2 implies that all $x \in (0.5, 1)$ is on the principal branch. This directly leads us to Part 1 of Theorem 3. □

Next, if we look into the region $x \in (0, 1/2)$, we can find that, in this region, QREs appears only when T_x and T_y are low. This observation can be formalized in the proposition below. We can see that this proposition directly proves Parts 2 and 3 of Theorem 2, as well as Part 2 of Theorem 3.

Proposition A3. *Consider Case 1a. Let* $x_1 = \min\left\{\frac{1}{2}, \frac{-T_y \ln\left(\frac{a_X}{b_X}\right) + b_Y}{a_Y + b_Y}\right\}$ *and* $T_B = \frac{b_Y}{\ln(a_X/b_X)}$. *The following statements are true for* $x \in (0, 1/2)$:

1. *If* $T_y > T_B$, *then* $T_X^{II} < 0$.
2. *If* $T_y < T_B$, *then* $T_X^{II} > 0$ *if and only if* $x \in (0, x_1)$.
3. $\frac{\partial L}{\partial x} > 0$ *for* $x \in (0, x_1)$.
4. *If* $T_y < T_I$, *then* $\frac{\partial T_X^{II}}{\partial x} > 0$.
5. *If* $T_y > T_I$, *then there is a nonnegative* critical temperature $T_C(T_y)$ *such that* $T_X^{II}(x, T_Y) \leq T_C(T_y)$ *for* $x \in (0, 1/2)$. *If* $T_Y < T_B$, *then* $T_C(T_y)$ *is given as* $T_X^{II}(x_L)$, *where* $x_L \in (0, x_1)$ *is the unique solution to* $L(x, T_y) = \frac{b_X}{a_X + b_X}$.

Proof. For the first and second part, consider any $x \in (0, 1/2)$ and we can see that

$$\begin{aligned} T_X^{II} > 0 &\Leftrightarrow y^{II} < \frac{b_X}{a_X + b_X} \\ &\Leftrightarrow \left(1 + e^{\frac{1}{T_y}(-(a_Y + b_Y)x + b_Y)}\right)^{-1} < \frac{b_X}{a_X + b_X} \\ &\Leftrightarrow x < \min\left\{\frac{1}{2}, \frac{-T_y \ln\left(\frac{a_X}{b_X}\right) + b_Y}{a_Y + b_Y}\right\}. \end{aligned}$$

Note that for $T_y > \frac{b_Y}{\ln(a_X/b_X)} = T_B$, we have $x_1 < 0$; hence, $T_X < 0$.

From the above derivation, for all $x \in (0, 1/2)$ such that $T_X^{II}(x, T_y) > 0$, $y^{II} < 1/2$ since $\frac{b_X}{a_X + b_X} < 1/2$. Then

$$\frac{\partial L}{\partial x} = \left[(1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y}\right] \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}}{\partial x} > 0.$$

Further, when $T_y < T_I$, $y^{II}(1/2, T_y) < \frac{b_X}{a_X+b_X}$. This implies that, for $x \in (0, 1/2)$, $y^{II}(x, T_y) < \frac{b_X}{a_X+b_X}$. Since $\frac{\partial L}{\partial x} > 0$, and L is continuous, $L(x, T_y) < \frac{b_X}{a_X+b_X}$ for $x \in (0, 1/2)$. This implies the fourth part of the proposition.

Next, if we look at the derivative of T_X^{II},

$$\frac{\partial T_X^{II}}{\partial x}(x, T_y) = \frac{-(a_X + b_X)L(x, T_y) + b_X}{x(1-x)[\ln(1/x - 1)]^2}.$$

We can see that any critical point in $x \in (0, 1/2)$ must satisfy $L(x, T_y) = \frac{b_X}{a_X+b_X}$. When $T_y > T_I$, $x_1 < 1/2$, and $L(x_1, T_y) > y^{II}(x_1, T_y) = \frac{b_X}{a_X+b_X}$. If $T_y < \frac{b_Y}{\ln(a_X/b_X)}$, then $\lim_{x\to 0+} T_X = y^{II}(0, T_Y) < \frac{b_X}{a_X+b_X}$. Hence, there is exactly one critical point for T_X for $x \in (0, x_1)$, which is a local maximum for T_X. If $T_y > \frac{b_Y}{\ln(a_X/b_X)}$, then we can see that T_X is always negative, in which case the critical temperature is zero. □

The results in Proposition A3 not only apply for the case $a_Y \geq b_Y$ but also general cases about the characteristics on $(0, 1/2)$. According to this proposition, we can conclude the following for the case $a_Y \geq b_Y$, as well as the case $a_Y < b_Y$ when $T_y > T_I$:

1. The temperature $T_B = \frac{b_Y}{\ln(a_X/b_X)}$ determines whether there is a branch appears in $x \in (0, 1/2)$.
2. There is some critical temperature T_C. If we raise T_x above T_C, then the system is always on the principal branch.
3. The critical temperature T_C is given as the solution to the equality $L(x, T_Y) = \frac{b_X}{a_X+b_X}$.

When there is a positive critical temperature, though it has no closed form solution, we can perform a binary search to look for $x \in (0, x_1)$ that satisfies $L(x, T_y) = \frac{b_X}{a_X+b_X}$.

Another result we are able to obtain from Proposition A3 is that the principal branch for Case 1a when $T_y < T_I$ lies on $(0, 1/2)$.

Proof of Part 2 of Theorem 1. First, we note that $T_y < T_I$ is meaningful only when $b_Y > a_Y$, for which case we always have $T_I < T_B$. From Proposition A3, we can see that for $T_Y^{II} < T_I$, we have $x_1 = 1/2$; hence, $T_X^{II} > 0$ for $x \in (0, 1/2)$. From Proposition A1, we already have $\lim_{x\to \frac{1}{2}^-} T_X^{II} = \infty$. Additionally, it is easy to see that $\lim_{x\to 0^+} T_X^{II} = 0$. As a result, since T_X^{II} is continuously differentiable over $(0, 0.5)$, for any $T_x > 0$, there exists $x \in (0, 0.5)$ such that $T_X^{II}(x, T_y) = T_x$. □

What remains to be shown is the characteristics on the side $(1/2, 1)$ when $b_Y > a_Y$. In Figures 4 and 5, for low T_y, the branch on the side $(1/2, 1)$ demonstrated a similar behavior as what we have shown in Proposition A3 for the side $(0, 1/2)$. However, for a high T_y, while we still can find that $(0, 1/2)$ contains the principal branch, the principal branch is not continuous. These observations are formalized in the following proposition. From this proposition, the proof of Part 4 of Theorem 2 directly follows.

Proposition A4. *Consider Case 1a with $b_Y > a_Y$. Let $x_2 = \max\left\{\frac{1}{2}, \frac{-T_Y \ln\left(\frac{a_X}{b_X}\right)+b_Y}{a_Y+b_Y}\right\}$ and $T_A = \max\left\{0, \frac{-a_Y}{\ln(a_X/b_X)}\right\}$. The following statements are true for $x \in (1/2, 1)$.*

1. *If $T_y < T_A$, then $T_X^{II} < 0$.*
2. *If $T_y > T_A$, then $T_X^{II} > 0$ if and only if $x \in (x_2, 1)$.*
3. *For $x \in \left[\frac{b_Y}{a_Y+b_Y}, 1\right)$, we have $\frac{\partial L}{\partial x} > 0$.*
4. *If $T_y \in (T_A, T_I)$, then there is a positive critical temperature $T_C(T_y)$ such that $T_X^{II}(x, T_y) \leq T_C(T_y)$ for $x \in (1/2, 1)$, given as $T_C(T_y) = T_X^{II}(x_L)$, where $x_L \in (1/2, 1)$ is the unique solution of $L(x, T_y) = \frac{b_X}{a_X+b_X}$.*

Proof. For the first part and the second part, consider $x \in (1/2, 1)$, and we can find that

$$\begin{aligned} T_X^{II} > 0 &\Leftrightarrow y^{II} > \frac{b_X}{a_X + b_X} \\ &\Leftrightarrow \left(1 + e^{\frac{1}{T_y}(-(a_Y+b_Y)x + b_Y)}\right)^{-1} > \frac{b_X}{a_X + b_X} \\ &\Leftrightarrow x > \max\left\{\frac{1}{2}, \frac{-T_y \ln\left(\frac{a_X}{b_X}\right) + b_Y}{a_Y + b_Y}\right\} = x_2. \end{aligned}$$

Note that, for $T_y > T_I$, $x_2 = 1/2$. Additionally, if $T_y < T_A$, then $T_X^{II} < 0$ for all $x \in (1/2, 1)$. For the third part, $y^{II} \geq \frac{1}{2}$ for all $x \geq \frac{b_Y}{a_Y + b_Y}$ and $\frac{b_Y}{a_Y+b_Y} > \frac{1}{2}$. Thus,

$$\frac{\partial L}{\partial x} = \left[(1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y}\right] \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}}{\partial x} > 0.$$

For the fourth part, we can find that any critical point of $L(x, T_Y)$ in $(0, 1)$ must be either $x = \frac{1}{2}$ or satisfies the following equation:

$$(1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y} = 0. \tag{A7}$$

Consider $G(x, T_y) = (1-2x) + x(1-x)(1-2y^{II})\frac{a_Y+b_Y}{T_y}$. For $b_Y > a_Y$, $y^{II}(1/2, T_y)$ is strictly less than $1/2$. Additionally, $\frac{b_Y}{a_Y+b_Y} > 1/2$. Now, $G(1/2, T_y) > 0$ and $G(\frac{b_Y}{a_Y+b_Y}, T_y) < 0$. Next, we can see that $G(x, T_y)$ is monotonically decreasing with respect to x for $x \in \left(\frac{1}{2}, \frac{b_Y}{a_Y+b_Y}\right)$ by looking at its derivative:

$$\frac{\partial G(x, T_y)}{\partial x} = -2 + \frac{a_Y + b_Y}{T_y}\left[(1-2x)(1-2y^{II}) - 2x(1-x)\frac{\partial y^{II}}{\partial x}\right] < 0.$$

As a result, there is some $x^* \in \left(\frac{1}{2}, \frac{b_Y}{a_Y+b_Y}\right)$ such that $G(x^*, T_y) = 0$. This implies that $L(x, T_y)$ has exactly one critical point for $x \in \left(\frac{1}{2}, \frac{b_Y}{a_Y+b_Y}\right)$. Additionally, if $G(x, T_y) > 0$, $\frac{\partial L}{\partial x} < 0$; if $G(x, T_y) < 0$, then $\frac{\partial L}{\partial x} > 0$. Therefore, x^* is a local minimum for L.

From the above arguments, we can conclude that the shape of $L(x, T_y)$ for $T_y < T_I$ is as follows:

1. There is a local maximum at $x = 1/2$, where $L(1/2, T_y) = y(1/2, T_y) < \frac{b_X}{a_X+b_X}$.
2. L is decreasing on the interval $\left(\frac{1}{2}, x^*\right)$, where x^* is the unique solution to Equation (A7).
3. L is increasing on the interval $(x^*, 1)$. If $T_y > T_A$, then $\lim_{x\to 1^-} L(x, T_y) = y(1, T_y) > \frac{b_X}{a_X+b_X}$.

Finally, we can claim that there is a unique solution to $L(x, T_Y) = \frac{b_X}{a_X+b_X}$, and such a point gives a local maximum to T_X^{II}. □

The above proposition suggests that, for $T_y \in (T_A, T_I)$, we are able to use binary search to find the critical temperature. For $T_y > T_I$, unfortunately, with a similar argument of Proposition A4, we can find that there are potentially at most two critical points for T_X^{II} on $(1/2, 1)$, as shown in Figure 5, which may induce an unstable segment between two stable segments. This also proves Part 3 of Theorem 3.

Now, we have enough materials to prove the remaining statements in Section 4.3.

Proof of Parts 1, 5, and 6 of Theorem 2, Part 4 of Theorem 3. For $T_y > T_I$, by Proposition A3, we can conclude that, for $x \in (0, x_L)$, we have $\frac{\partial T_X^{II}}{\partial x} > 0$, for which the QREs are stable by Lemma 1. With similar arguments, we can conclude that the QREs on $x \in (x_L, x_1)$ are unstable. Additionally, given T_x, the stable QRE $x_a \in (0, x_L)$ and the unstable $x_b \in (x_L, x_1)$ that satisfies $T_X^{II}(x_a, T_y) = T_X^{II}(x_b, T_y) = T_x$

appear in pairs. For $T_y < T_I$, with the same technique and by Proposition A4, we can claim that the QREs in $x \in (x_2, x_L)$ are unstable, while the QREs in $x \in (x_L, 1)$ are stable. This proves the first part of Theorem 2 and Part 4 of Theorem 3.

Parts 5 and 6 of Theorem 2 are corollaries of Part 5 of Proposition A3 and Part 4 of Proposition A4. □

Appendix D.1.2. Case 1b: $b_X > 0$, $a_Y + b_Y < 0$

In this case, both players have different preferences. For the game within this class, there is only one Nash equilibrium (either pure or mixed). We presented examples in Figures A1 and A2. We can see that, in these figures, there is only one QRE given T_x and T_y. We show in the following two propositions that this observation is true for all instances.

Proposition A5. *Consider Case 1b. Let* $x_3 = \max\left\{0, \frac{-T_y \ln(a_X/b_X)+b_Y}{a_Y+b_Y}\right\}$. *If* $T_y < T_I$, *then the following statements are true:*

1. $T_X^{II}(x, T_y) < 0$ *for* $x \in (1/2, 1)$.
2. $T_X^{II}(x, T_y) > 0$ *for* $x \in \left(x_3, \frac{1}{2}\right)$.
3. $\frac{\partial T_X^{II}(x,T_y)}{\partial x} > 0$ *for* $x \in \left(x_3, \frac{1}{2}\right)$.
4. $\left(x_3, \frac{1}{2}\right)$ *contains the principal branch.*

Proof. Note that, if $T_y < T_I$, $x_3 < 1/2$. Additionally, according to Proposition A2, $y^{II}(1/2, T_y) < \frac{b_X}{a_X+b_X}$. Since y^{II} is continuous and monotonically decreasing with x, $y^{II} < \frac{b_X}{a_X+b_X}$ for $x > 1/2$. Therefore, the numerator of Equation (6) is always positive for $x \in (1/2, 1)$, which makes T_X^{II} negative. This proves the first part of the proposition.

For the second part, observe that, for $x \in (0, 1/2)$, $T_X^{II} > 0$ if and only if $y^{II} < \frac{b_X}{a_X+b_X}$. This is equivalent to $x > \frac{-T_y \ln(a_X/b_X)+b_Y}{a_Y+b_Y}$.

For the third part, note that, for $x \in (0, 1/2)$, $x(1-x)\ln(1/x-1)\frac{\partial y^{II}}{\partial x} < 0$. This implies $L(x, T_y) < y^{II}(x, T_y) < \frac{b_X}{a_X+b_X}$ for $x \in (x_3, 1/2)$, from which we can conclude that $\frac{\partial T_X^{II}(x,T_y)}{\partial x} > 0$.

Finally, we note that if $x_3 > 0$, then $T_X^{II}(x_3, T_y) = 0$. If $x_3 = 0$, we have $\lim_{x\to 0^+} T_X^{II} = 0$. As a result, we can conclude that $(x_3, 1/2)$ contains the principal branch. □

With the similar arguments, we are able to show the following proposition for $T_y > T_I$:

Proposition A6. *Consider Case 1b. Let* $x_3 = \min\left\{1, \frac{-T_y \ln(a_X/b_X)+b_Y}{a_Y+b_Y}\right\}$. *If* $T_y > T_I$, *then the following statements are true:*

1. $T_X^{II}(x, T_y) < 0$ *for* $x \in (0, 1/2)$.
2. $T_X^{II}(x, T_y) > 0$ *for* $x \in \left(\frac{1}{2}, x_3\right)$.
3. $\frac{\partial T_X^{II}(x,T_y)}{\partial x} < 0$ *for* $x \in \left(\frac{1}{2}, x_3\right)$.
4. $\left(\frac{1}{2}, x_3\right)$ *contains the principal branch.*

Appendix D.1.3. Case 1c: $a_Y + b + Y = 0$

In this case, we have $T_I = \frac{b_Y}{\ln(a_X/b_X)}$, and y^{II} is a constant with respect to x. The proof of Theorem A5 for $a_Y + b_Y = 0$ directly follows from the following proposition.

Proposition A7. *Consider Case 1c. The following statements are true:*

1. *If* $T_y < T_I$, *then* $T_X^{II}(x, T_y) < 0$ *for* $x \in (0.5, 1)$, *and* $T_X^{II}(x, T_y) > 0$ *for* $x \in (0, 0.5)$.

2. *If* $T_y > T_I$, *then* $T_X^{II}(x, T_y) < 0$ *for* $x \in (0, 0.5)$, *and* $T_X^{II}(x, T_y) > 0$ *for* $x \in (0.5, 1)$.
3. *If* $T_y < T_I$, *then* $\frac{\partial T_X^{II}(x,T_y)}{\partial x} > 0$ *for* $x \in (0, 0.5)$.
4. *If* $T_y > T_I$, *then* $\frac{\partial T_X^{II}(x,T_y)}{\partial x} < 0$ *for* $x \in (0.5, 1)$.

Proof. Note that $y^{II} = \left(1 + e^{b_Y/T_y}\right)^{-1}$.

First consider the case when $a_Y > b_Y$. In this case, $T_I = 0$ and $b_Y < 0$. Therefore, $y^{II} > \frac{b_X}{a_X + b_X}$, from which we can conclude that $T_X^{II} > 0$ for $x \in (0.5, 1)$ and $T_X^{II} < 0$ for $x \in (0, 0.5)$, for any positive T_y.

Now consider the case where $a_Y < b_Y$. If $T_y < T_I$, $y^{II} < \frac{b_X}{a_X + b_X}$; hence, we get $T_X^{II}(x, T_y) < 0$ for $x \in (0.5, 1)$ and $T_X^{II}(x, T_y) > 0$ for $x \in (0, 0.5)$, which is the first part of the proposition statement. Similarly, if $T_y > T_I$, $y^{II} > \frac{b_X}{a_X + b_X}$, from which the second part of the proposition follows.

For the third part and the fourth part, note that $L(x, T_y) = y^{II}$ in this case, as $\frac{\partial y^{II}}{\partial x} = 0$ as per Equation (A5), and the sign of the derivative of T_X^{II} can be seen from Lemma A2. □

Appendix D.2. Case 2: $b_X < 0$

In this case, the first action is a dominating strategy for the first player. Note that both $-(a_X + b_X)$ and b_X are not positive, which means that the numerator of Equation (6) is always smaller than or equal to zero. This implies that all QRE correspondences appear on $x \in \left(\frac{1}{2}, 1\right)$. In fact, since $y^{II} > 0$ for $x \in (1/2, 1)$, the numerator of Equation (6) is always negative, we have $T_X^{II} > 0$ for $x \in (1/2, 1)$. Additionally, we can easily see that

$$\lim_{x \to \frac{1}{2}^+} T_X^{II}(x, T_y) = +\infty.$$

This implies that $(1/2, 1)$ contains the principal branch. First, we show the result when $a_Y + b_Y < 0$ in the following proposition. Additionally, the bifurcation diagram is presented in Figure A3.

Proposition A8. *For Case 2, if* $a_Y + b_Y < 0$, *then for* $x \in (1/2, 1)$, $\frac{\partial T_X^{II}}{\partial x} < 0$.

Proof. In this case, y^{II} is monotonically decreasing with x. We can see that

$$L(x, T_Y) = y^{II} + x(1 - x) \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}}{\partial x} > y^{II} > 0$$

since $x(1 - x) \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}}{\partial x}$ is positive for $x \in (1/2, 1)$. Bringing this back to Equation (A4), we have $\frac{\partial T_X^{II}}{\partial x} < 0$. □

For $a_Y + b_Y > 0$, if $a_Y > b_Y$, the bifurcation diagram has the similar trend as in Figure A3; while, if $a_Y < b_Y$, we lose the continuity on the principal branch.

Proposition A9. *For Case 2, if* $a_Y + b_Y > 0$, *then for* $x \in (1/2, 1)$, *we have*

1. *if* $a_Y > b_Y$, *then* $\frac{\partial T_X^{II}}{\partial x} < 0$.
2. *if* $a_Y < b_Y$, *then* T_X *has at most two local extrema.*

Proof. In this case, y^{II} is monotonically increasing with x. For $a_Y > b_Y$, we can find that $y^{II}(1/2, T_y) > 0$ and $L(1/2, T_y) = y^{II}(1/2, T_y) > 0$. Additionally, we can obtain that L is monotonically increasing for $x \in (1/2, 1)$ by inspecting

$$\frac{\partial L(x, T_y)}{\partial x} = \left[(1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y}\right] \ln\left(\frac{1}{x} - 1\right) \frac{\partial y^{II}(x, T_y)}{\partial x} > 0.$$

Hence, for $x \in (1/2, 1)$, $L(x, T_y) > 0$. This implies $\frac{\partial T_X^{II}}{\partial x} < 0$ for $x \in (1/2, 1)$.

For the second part, we can find that, for $a_Y < b_Y$, $y^{II}(1/2) < 1/2$. Let $x_2 = \min\left\{1, \frac{b_Y}{a_Y + b_Y}\right\}$. First note that, if $x_2 < 1$, then, for $x > x_2$, we have $y > 1/2$, and further we can get $\frac{\partial L(x, T_y)}{\partial x} > 0$ for $x \in (x_2, 1)$. We use the same technique as in the proof of Proposition A4. Let $G(x, T_y) = (1-2x) + x(1-x)(1-2y^{II})\frac{a_Y + b_Y}{T_y}$. Note that $G(1/2, T_y) > 0$ and $G(x_2, T_y) < 0$. Next, observe that $G(x, T_y)$ is monotonically decreasing for $x \in \left(\frac{1}{2}, x_2\right)$. Hence, there is an $x^* \in (1/2, x_2)$ such that $G(x^*, T_y) = 0$. This x^* is a local minimum for L. We can conclude that, for $x \in (1/2, 1)$, L has the following shape:

1. There is a local maximum at $x = 1/2$, where $L(1/2, T_y) = y(1/2, T_y) > 0$.
2. L is decreasing on the interval $x \in (1/2, x^*)$, where x^* is the solution to $G(x^*, T_y) = 0$.
3. L is increasing on the interval $x \in (x^*, x_2)$. Note that $\lim_{x \to 1^-} L(x, T_y) = y^{II}(1, T_y) > 0$.

As a result, if $L(x^*, T_y) > \frac{b_X}{a_X + b_X}$, then T_X^{II} is monotonically decreasing; otherwise, T_X^{II} has a local minimum and a local maximum on $(1/2, 1)$. □

References

1. Devaney, R.L. *A First Course in Chaotic Dynamical Systems*; Westview Press: Boulder, CO, USA, 1992.
2. Roughgarden, T. Intrinsic robustness of the price of anarchy. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC 2009), Bethesda, MD, USA, 31 May–2 June 2009; pp. 513–522.
3. Palaiopanos, G.; Panageas, I.; Piliouras, G. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5874–5884.
4. Watkins, C.J.C.H. Learning from Delayed Rewards. Ph.D Thesis, University of Cambridge, Cambridge, UK, 1989.
5. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
6. Tan, M. Multi-agent reinforcement learning: Independent vs. cooperative agents. In Proceedings of the tenth international conference on machine learning, Amherst, MA, USA, 27–29 June 1993; pp. 330–337.
7. McKelvey, R.D.; Palfrey, T.R. Quantal response equilibria for normal form games. *Games Econ. Behav.* **1995**, *10*, 6–38. [CrossRef]
8. Nash, J. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 48–49. [CrossRef] [PubMed]
9. Wolpert, D.H.; Harré, M.; Olbrich, E.; Bertschinger, N.; Jost, J. Hysteresis effects of changing the parameters of noncooperative games. *Phys. Rev. E* **2012**, *85*, 036102. [CrossRef] [PubMed]
10. Kianercy, A.; Veltri, R.; Pienta, K.J. Critical transitions in a game theoretic model of tumour metabolism. *Interface Focus* **2014**, *4*, 20140014. [CrossRef] [PubMed]
11. Cominetti, R.; Melo, E.; Sorin, S. A payoff-based learning procedure and its application to traffic games. *Games Econ. Behav.* **2010**, *70*, 71–83. [CrossRef]
12. Coucheney, P.; Gaujal, B.; Mertikopoulos, P. Entropy-Driven Dynamics and Robust Learning Procedures in Games. Available online: https://hal.inria.fr/hal-00790815/document (accessed on 25 April 2018).
13. Sato, Y.; Crutchfield, J.P. Coupled replicator equations for the dynamics of learning in multiagent systems. *Phys. Rev. E* **2003**, *67*, 015206. [CrossRef] [PubMed]

14. Tuyls, K.; Verbeeck, K.; Lenaerts, T. A selection-mutation model for q-learning in multi-agent systems. In Proceedings of the 2nd international joint conference on Autonomous agents and multiagent systems, Melbourne, Australia, 14–18 July 2003; pp. 693–700.
15. Sandholm, W.H. Evolutionary game theory. In *Encyclopedia of Complexity and Systems Science*; Springer: Berlin, Germany, 2009; pp. 3176–3205.
16. Kianercy, A.; Galstyan, A. Dynamics of Boltzmann q learning in two-player two-action games. *Phys. Rev. E* **2012**, *85*, 041145. [CrossRef] [PubMed]
17. Hofbauer, J.; Hopkins, E. Learning in perturbed asymmetric games. *Games Econ. Behav.* **2005**, *52*, 133–152. [CrossRef]
18. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
19. Perko, L. *Differential Equations and Dynamical Systems*, 3rd ed.; Springer: Berlin, Germany, 1991.
20. Tomlinson, I.; Bodmer, W. Modelling the consequences of interactions between tumour cells. *Br. J. Cancer* **1997**, *75*, 157–160. [CrossRef] [PubMed]
21. Kaznatcheev, A.; Scott, J.G.; Basanta, D. Edge effects in game-theoretic dynamics of spatially structured tumours. *J. R. Soc. Interface* **2015**, *12*, 20150154. [CrossRef] [PubMed]
22. Basanta, D.; Scott, J.G.; Fishman, M.N.; Ayala, G.; Hayward, S.W.; Anderson, A.R. Investigating prostate cancer tumour–stroma interactions: Clinical and biological insights from an evolutionary game. *Br. J. Cancer* **2012**, *106*, 174–181. [CrossRef] [PubMed]
23. Kaznatcheev, A.; Velde, R.V.; Scott, J.G.; Basanta, D. Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature. *arXiv* **2016**, arXiv:1608.00985. [PubMed]
24. Basanta, D.; Simon, M.; Hatzikirou, H.; Deutsch, A. Evolutionary game theory elucidates the role of glycolysis in glioma progression and invasion. *Cell Prolif.* **2008**, *41*, 980–987. [CrossRef] [PubMed]
25. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70. [CrossRef]
26. Axelrod, R.; Axelrod, D.E.; Pienta, K.J. Evolution of cooperation among tumor cells. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13474–13479. [CrossRef] [PubMed]
27. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [CrossRef] [PubMed]
28. Ribeiro, M.; Silva, A.S.; Bailey, K.M.; Kumar, N.B.; Sellers, T.A.; Gatenby, R.A.; Ibrahim-Hashim, A.; Gillies, R.J. Buffer Therapy for Cancer. *J. nutr. Food Sci.* **2012**, *2*, 6. [CrossRef] [PubMed]
29. Piliouras, G.; Nieto-Granda, C.; Christensen, H.I.; Shamma, J.S. Persistent Patterns: Multi-agent Learning Beyond Equilibrium and Utility. In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (AAMAS), Paris, France, 5–9 May 2014; pp. 181–188.
30. Papadimitriou, C.; Piliouras, G. From Nash Equilibria to Chain Recurrent Sets: Solution Concepts and Topology. In Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, 14–16 January 2016; pp. 227–235.
31. Panageas, I.; Piliouras, G. 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, Maastricht, The Netherlands, 24–28 July 2016; pp. 703–720.
32. Bloembergen, D.; Tuyls, K.; Hennes, D.; Kaisers, M. Evolutionary dynamics of multi-agent learning: A survey. *J. Artif. Intell. Res.* **2015**, *53*, 659–697.
33. Romero, J. The effect of hysteresis on equilibrium selection in coordination games. *J. Econ. Behav. Organ.* **2015**, *111*, 88–105. [CrossRef]
34. Kleinberg, R.; Ligett, K.; Piliouras, G.; Tardos, É. Beyond the Nash equilibrium barrier. In Proceedings of the Symposium on Innovations in Computer Science (ICS), Beijing, China, 7–9 January 2011.
35. Piliouras, G.; Shamma, J.S. Optimization Despite Chaos: Convex Relaxations to Complex Limit Sets via Poincaré Recurrence. In Proceedings of the Symposium of Discrete Algorithms (SODA), Portland, OR, USA, 5–7 January 2014.
36. Kleinberg, R.; Piliouras, G.; Tardos, É. Multiplicative Updates Outperform Generic No-Regret Learning in Congestion Games. In Proceedings of the ACM Symposium on Theory of Computing (STOC), Bethesda, MD, USA, 31 May–2 June 2009.

37. Mehta, R.; Panageas, I.; Piliouras, G.; Yazdanbod, S. The Computational Complexity of Genetic Diversity. In Proceedings of the 24th Annual European Symposium on Algorithms (ESA 2016), Aarhus, Denmark, 22–24 August 2016; Sankowski, P., Zaroliagis, C., Eds.; *Leibniz International Proceedings in Informatics (LIPIcs)*; Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik: Dagstuhl, Germany, 2016; Volume 57, p. 65.
38. Livnat, A.; Papadimitriou, C.; Dushoff, J.; Feldman, M.W. A mixability theory for the role of sex in evolution. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 19803–19808. Available online: http://www.pnas.org/content/105/50/19803.full.pdf+html (accessed on 20 April 2018). [CrossRef] [PubMed]
39. Chastain, E.; Livnat, A.; Papadimitriou, C.H.; Vazirani, U.V. Multiplicative updates in coordination games and the theory of evolution. In Proceedings of the 4th Innovations in Theoretical Computer Science (ITCS) conference, Berkeley, CA, USA, 10–12 January 2013; pp. 57–58.
40. Chastain, E.; Livnat, A.; Papadimitriou, C.; Vazirani, U. Algorithms, games, and evolution. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10620–10623. Available online: http://www.pnas.org/content/early/2014/06/11/1406556111.full.pdf+html (accessed on 20 April 2018). [CrossRef] [PubMed]
41. Livnat, A.; Papadimitriou, C.; Rubinstein, A.; Valiant, G.; Wan, A. Satisfiability and evolution. In Proceedings of the 2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS), Philadelphia, PA, USA, 18–21 October 2014; pp. 524–530.
42. Meir, R.; Parkes, D. A Note on Sex, Evolution, and the Multiplicative Updates Algorithm. In Proceedings of the 12th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 15), Istanbul, Turkey, 4–8 May 2015.
43. Mehta, R.; Panageas, I.; Piliouras, G. Natural Selection as an Inhibitor of Genetic Diversity: Multiplicative Weights Updates Algorithm and a Conjecture of Haploid Genetics. In Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, 11–13 January 2015.
44. Mehta, R.; Panageas, I.; Piliouras, G.; Tetali, P.; Vazirani, V.V. Mutation, Sexual Reproduction and Survival in Dynamic Environments. In Proceedings of the 2017 Conference on Innovations in Theoretical Computer Science (To Appear), ITCS' 17, Berkeley, CA, USA, 9–11 January 2017.
45. Livnat, A.; Papadimitriou, C. Sex as an algorithm: The theory of evolution under the lens of computation. *Commun. ACM (CACM)* **2016**, *59*, 84–93. [CrossRef]
46. Sandholm, W.H. *Population Games and Evolutionary Dynamics*; MIT Press: Cambridge, MA, USA, 2010.
47. Bendixson, I. Sur les courbes définies par des équations différentielles. *Acta Math.* **1901**, *24*, 1–88. [CrossRef]
48. Teschl, G. *Ordinary Differential Equations and Dynamical Systems*; American Mathematical Soc.: Providence, RI, USA, 2012; Volume 140.
49. Müller, J.; Kuttler, C. *Methods and Models in Mathematical Biology*; Springer: Berlin, Germany, 2015.
50. Meiss, J. *Differential Dynamical Systems*; SIAM: Philadelphia, PA, USA, 2007.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).