ESSAYS ON POPULATION LEARNING DYNAMICS AND BOUNDEDLY RATIONAL BEHAVIOR

by

Russell Golman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2009

Doctoral Committee:

> Professor Andreas R. Blass, Co-Chair
> Professor Scott E. Page, Co-Chair
> Professor Carl P. Simon
> Professor Charles R. Doering
> Assistant Professor Stephan Lauermann

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

Introduction

Perfectly rational behavior in game theory is characterized by accurate expectations of other players' actions and choices that maximize individual utility. This is not what we typically observe when real people play games. This dissertation contains four essays investigating the aggregate behavior of agents who are boundedly rational or who learn about which actions to take. Bounded rationality means that agents make mistakes when trying to choose actions that maximize utility. Agents who follow a learning rule do not accurately predict other players' behavior and in this respect also fall short of the standard of perfect rationality.

Theories of bounded rationality seek to explain divergences from utility maximization that may arise in complicated strategic environments or in one-shot games before learning can occur. In Chapter II, we take up the quantal response model of bounded rationality. In a quantal response equilibrium (McKelvey and Palfrey, 1995), mistakes are caused by noise in observed payoffs, and players react to everybody else's noisy best responses with their own. The mistakes one player makes thus affect the strategic context that another player faces. This framework allows for error cascades in which small decision errors induce a large change in equilibrium play.

1

Theories of learning in games do not aim to replace standard equilibrium analysis, but rather to complement it. Equilibrium can be thought of as a long-term outcome of a learning process or an evolutionary dynamic. Learning models are concerned with behavior during this period of equilibration. Population learning rules provide a mechanism through which a rational Nash equilibrium may arise without assuming that all players are perfectly rational and have common knowledge of each other's rational decision making rule. In a population learning rule, agents choose pure strategies according to a given algorithm, and the resulting configuration of actions determines a population mixed strategy. It is this population mixed strategy we are interested in. A given learning rule may or may not lead to a steady state of the population mixed strategy in a particular game. In fact, Hofbauer and Swinkels (1996) and Hart and MasColell (2003) show that no reasonable dynamic converges to Nash equilibrium in all games.

In Chapters III, IV, and V, we compare various deterministic learning models in games that have multiple equilibria to see if differences in learning styles lead to significant differences in outcomes. Different learning dynamics, even when they produce the same sets of stable equilibria, may attain different equilibrium points in that set. The basin of attraction of each equilibrium depends on how players learn. A recurring device in our analysis is our finding that the overlap in the basins of attraction under different learning dynamics can be arbitrarily small. Chapter III develops a necessary condition and sufficient conditions for there to be vanishing overlap in the basins of attraction under two common learning rules – best response dynamics and replicator dynamics – and then extends this result to broader classes of dynamics. Chapter IV applies the best response dynamics and the replicator dynamics to a class of generalized stag hunt games, finding that in this context, learning style can

determine whether an efficient equilibrium is reached. In Chapter V, we consider compositions of best response and replicator dynamics and show that outcomes can be highly sensitive to the precise specification of the learning dynamic.

A central theme in Chapters II and V is analysis of heterogeneous populations. In Chapter II, agents have heterogeneous quantal response functions – that is, they have different probabilities of making errors. In Chapter V, the population admits heterogeneous learning styles. In both cases, the presence of heterogeneity allows for behavior that could not occur in homogeneous populations. These results remind us to take seriously the fact that people err and learn at different rates and in different ways.

BIBLIOGRAPHY

[1] Hart, S., and A. Mas-Colell (2003). Uncoupled Dynamics Do Not Lead to Nash Equilibrium. *American Economic Review* 93, 1830-1836.

[2] Hofbauer, J., and J. M. Swinkels (1996). A Universal Shapley Example. Unpublished manuscript, University of Vienna and Northwestern University.

[3] McKelvey, R., and T. Palfrey (1995). Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10, 6-38.

CHAPTER II

Quantal Response Equilibria with Heterogeneous Agents

2.1 Introduction

Quantal response equilibrium extends the Nash Equilibrium notion to allow bounded rationality. Players can be seen as making errors while trying to choose optimal strategies, or equivalently, as observing payoffs disturbed by idiosyncratic noise. The result is that players may select any action with positive probability assigned by their quantal response functions.

This chapter introduces a general model of quantal response equilibrium with heterogeneous agents. We show that the aggregate behavior of a population of heterogeneous agents can be captured by a representative agent. But, the representative agent may be very different than the actual agents in the population. This illustrates the need to consider heterogeneity and offers insight for how to work around that heterogeneity with representative-agent models. After presenting the representative-agent picture, which allows for arbitrary distributions of payoff noise and applies for all normal form games, we then consider logit responses in the context of a single choice between two pure strategies that is part of a fixed game. We find that in a heterogeneous population of agents, all having their own logit rationality parameters, a mis-specified homogeneous logit parameter will always exhibit a downward

bias making the population appear to be less rational.

We consider structural quantal response equilibria (QRE) [22, 11] in the context of a population game. In a large population of agents, we should expect heterogeneity of behavior [19, 21]. A population of quantal responders should consist of agents who may have different error rates, or different distributions of payoff noise. In fact, McKelvey, et. al. [23] find experimental evidence for heterogeneous error distributions in trying to fit logit QRE to data on two-by-two asymmetric games. [1]

Prior research into quantal response equilibria with heterogeneous agents has considered a distribution of parameters which parametrize the distributions of payoff noise [25], with particular interest in distributions of logit responders [4]. Here, we model heterogeneous distributions of payoff noise with a functional defined over distribution functions. As we do not assume that distributions of payoff noise take any particular functional forms, this approach allows for more distribution functions than can be described with finitely many parameters.

Our interest is in the behavior of an entire population, and we seek a representative agent whose mixed strategy quantal response always matches the population aggregate. We need representative-agent models because while we believe people really are heterogeneous, we cannot determine each person's quantal response function individually when we fit data. The representative agent is what we can estimate in an experiment.

With weak assumptions on the agents' distributions of payoff noise we prove existence of a representative agent. However, the distribution of payoff disturbances necessary to produce representative choices is not representative of the noise the actual agents observe in their payoffs. We show that in games with enough pure

---

[1]Further motivation to consider heterogeneity in a population of quantal responders comes from recent findings that models of heterogeneous learners often cannot be adequately approximated by representative-agent models with common parameter values for all [27, 15, 12].

strategies, a representative agent could not have payoff disturbances independent and identically distributed across actions even if the actual agents did. On the other hand, we find that if agents all use regular quantal response functions (as defined by Goeree, et. al. [11]), then the representative agent's quantal response must also be regular. Different roles in asymmetric games will in general have different representative agents.

Much of the QRE literature looks to the logit equilibrium in particular to explain experimental data [5, 8, 10, 6, 1]. Because of the prominence of this logit response specification, we consider a population of heterogeneous logit responders as a special case. Our interest here is how a mis-specified homogeneous logit model misrepresents the heterogeneous agents. Because the representative agent for the population is not itself a logit responder, the homogeneous model cannot explain equilibrium choice probabilities and payoffs in a choice between more than two actions. When the population has just two pure strategies, we find that the homogeneous logit parameter is systematically biased below the average value of the heterogeneous logit parameters. We describe the extent of this bias as it varies with the difference in the two strategies' equilibrium payoffs.

Nash proposed a population game interpretation of equilibrium in his unpublished PhD dissertation [26]. Following his lead, we assume that there is a population of agents for each role in a game. A generic n-player game involves n populations of agents, but if multiple players have identical roles and we adopt the restriction that players in identical roles should play identical population mixed strategies, then these players may be selected from the same population. So, in a totally symmetric game, we may have only a single population of agents. We assume the populations are large, and we are interested in the fraction of a population playing a given strategy.

An agent's payoff is the average of his payoffs against all other combinations of agents (or equivalently his expected payoff given random matching).

Population games provide a framework for the use of evolutionary learning dynamics. Learning rules that assume that players noisily best respond often converge to QRE [7, 17, 20, 16, 2, 18]. This chapter focuses on the QRE itself and not on any particular learning rule that might lead to it. Population games also describe experimental settings well, as data is accumulated through the randomly matched interactions of many subjects.

This chapter is organized as follows. Section 2.2 introduces the notation in the context of a single population and provides definitions of a QRE and a representative agent. Section 2.3 contains our general results describing a representative agent. In Section 2.4, we extend our framework and our results to n-player asymmetric games. Section 2.5 focuses on logit responders, and section 2.6 concludes. The Appendix contains proofs omitted from the text.

2.2 A Single Population

To simplify the presentation, we begin with a single population of agents. The context can be thought of as a symmetric game or alternatively a single player decision subject to incomplete information. In Section 2.4, we show how to apply these results to general n-player asymmetric games.

Let $S = \{s_1, \ldots, s_J\}$ be the set of pure strategies available to the agents. The collective play of all the agents defines the population mixed strategy x. Formally, $x \in \triangle^{J-1}$, the $(J-1)$-dimensional simplex where $x_j \geq 0$ for all j and $\sum_j x_j = 1$.

A structural QRE arises when agents' utility functions are modified by noise terms, privately observed stochastic payoff disturbances. Denote by π_j the payoff

from taking pure strategy s_j. Of course, payoffs are a function of the strategies used by all the players, $\pi_j = \pi_j(x)$, but we omit the function's argument for ease of notation. We denote the vector of payoffs to each pure strategy by $\pi = \pi_1, \ldots, \pi_J$. Formally, $\pi : \triangle^{J-1} \to \Re^J$. For each pure strategy s_j, agent μ observes a payoff disturbance ϵ_j^μ, making agent μ's disturbed payoff $\pi_j^\mu = \pi_j + \epsilon_j^\mu$. This is the function agents maximize with their choice of strategy in a QRE.

The distribution of payoff disturbances is assumed to be *admissible*, meaning that:

(a1) the disturbances are independent across agents;

(a2) each agent has an absolutely continuous joint distribution of $(\epsilon_1^\mu, \ldots, \epsilon_J^\mu)$ that is independent of the population mixed strategy x, i.e., all marginal densities exist;

(a3) disturbances are unbiased in the sense that they all have mean zero.

Allowing only admissible distributions guarantees the existence of a QRE. Here, we make the additional assumption that for each agent, disturbances are independent and identically distributed (iid) across the set of actions. This assumption could be relaxed, but some such restriction is necessary for the QRE notion to produce falsifiable predictions [14].

When the setup for QRE does not explicitly involve populations of agents, it is assumed that each player has a distribution of payoff disturbances. In the context of a population game, this corresponds to each agent within the population having an identical distribution of disturbances. That is, the convention is to assume homogeneous populations. Here, we specifically want to leave open the possibility that agents in the same population have different distributions of payoff shocks. So, we do *not* assume identical distributions of ϵ_j^μ for all μ.

To model heterogeneity in the distributions of payoff disturbances, consider a functional defined over such distributions. Let $P_\epsilon(\cdot)$ be a distribution function for the payoff disturbance to a particular action. Each agent has a distinct P_ϵ, which then applies to ϵ_j^μ for all $1 \leq j \leq J$, i.e., is the same for all actions in that agent's strategy space. Define a functional $F_\epsilon[P_\epsilon]$ that associates to each distribution function P_ϵ a probability mass or density describing the fraction of the population with payoff disturbances distributed by P_ϵ. Technically, we make use of a second functional $I_\epsilon[P_\epsilon]$ that equals 1 to indicate a mass point on P_ϵ and 0 to indicate that $F_\epsilon[P_\epsilon]$ represents a probability density. For this to make sense we require $I_\epsilon[P_\epsilon] = 1$ for only countably many P_ϵ and

$$\sum_{P_\epsilon : I_\epsilon[P_\epsilon]=1} F_\epsilon[P_\epsilon] + \int_{P_\epsilon : I_\epsilon[P_\epsilon]=0} F_\epsilon[P_\epsilon]\, dP_\epsilon = 1.$$

The appropriate measure dP_ϵ depends on the particular form of the heterogeneity.

In this approach, the functional captures a distribution of distributions of payoff shocks in the population. It thus provides a general way to think about heterogeneity of quantal responses. The conventional assumption of a homogeneous population can be recaptured, for example, by taking $F_\epsilon[P_\epsilon] = 1$ for a particular P_ϵ and 0 everywhere else.

The quantal response function for each agent returns the agent's likelihood of choosing each strategy given the agent's undisturbed payoffs. Let $Q_j^\mu(\pi)$ be the probability that agent μ selects strategy s_j given the payoffs to each strategy. Formally, for any vector $\pi' = (\pi_1', \ldots, \pi_J') \in \Re^J$, define

$$R_j^\mu(\pi') = \{(\varepsilon_1^\mu, \ldots, \varepsilon_J^\mu) \in \Re^J : \pi_j' + \varepsilon_j^\mu \geq \pi_{j'}' + \varepsilon_{j'}^\mu \text{ for all } j' = 1, \ldots, J\}$$

to be the set of realizations of agent μ's joint set of payoff disturbances that would lead to choosing action s_j. Then $Q_j^\mu(\pi) = \text{Prob}\left\{(\epsilon_1^\mu, \ldots, \epsilon_J^\mu) \in R_j^\mu(\pi)\right\}$.

The quantal response functions for all the agents can be aggregated across the population to give the population mixed strategy response to any given population state. In a finite population of m agents, the population aggregate quantal response is $Q_j = \frac{1}{m} \sum_{\mu=1}^{m} Q_j^\mu$ for all j. More generally, the aggregate quantal response in an infinite population is

$$(2.1) \qquad Q_j = \sum_{P_\epsilon : I_\epsilon[P_\epsilon]=1} F_\epsilon[P_\epsilon] \, Q_j^\mu + \int_{P_\epsilon : I_\epsilon[P_\epsilon]=0} F_\epsilon[P_\epsilon] \, Q_j^\mu \, dP_\epsilon$$

where we abuse notation by letting $\mu = \mu(P_\epsilon)$ be an agent with payoff disturbances iid from P_ϵ. This is just the expectation of agents' quantal response functions with respect to the probability mass / density functional F_ϵ. It can be taken pointwise, i.e., independently for every value of the payoff vector π.

We can now define a quantal response equilibrium and then formally describe a representative agent for this heterogeneous population.

Definition II.1. A *quantal response equilibrium (QRE)* is defined by the fixed point equation $x_j = Q_j \left(\pi(x) \right)$ for all j.

Whereas a Nash Equilibrium is a state of play with everybody simultaneously playing a best response, a QRE is a state with everybody simultaneously playing according to their quantal response functions.

Definition II.2. A *representative agent* would have a quantal response function $\hat{Q}(\pi)$ equal to the population aggregate quantal response function:

$$(2.2) \qquad \hat{Q} = (Q_1, \ldots, Q_J).$$

For all games, the population as a whole behaves exactly as if it were homogeneously composed of representative agents.

The property of having the population aggregate quantal response function is the most transparent characterization of a representative agent, but other equivalent characterizations are more useful. We now develop these alternative characterizations.

Strategy choices are determined by comparing the disturbed payoffs, so the most relevant variables are the differences between payoff shocks, $\delta_{jj'}^{\mu} = \epsilon_j^{\mu} - \epsilon_{j'}^{\mu}$. These $\delta_{jj'}^{\mu}$ are identically distributed across all j and $j' \neq j$ because the ϵ_j^{μ} are iid across all j. By absolute continuity, the marginal densities exist, and they are even functions because the $\delta_{jj'}^{\mu}$ are antisymmetric in the indices j and j'. There is obviously dependence among these random variables across j and j'. We will consider the $(J-1)$-dimensional random vector $\delta_j^{\mu} = \left(\delta_{1j}^{\mu}, \ldots, \widehat{\delta_{jj}^{\mu}}, \ldots, \delta_{Jj}^{\mu} \right)$ for a particular j, which then determines the value of $\delta_{j'}^{\mu}$ for all other j'. Note that admissibility of the payoff disturbances implies δ_j^{μ} has zero mean because all the ϵ_j^{μ} have zero mean. Let $P_{\delta_j^{\mu}} : \Re^{J-1} \to [0, 1]$ be the joint distribution function of δ_j^{μ}. Then

$$(2.3) \qquad Q_j^{\mu}(\pi) = P_{\delta_j^{\mu}}(\pi_j - \pi_1, \pi_j - \pi_2, \ldots, \pi_j - \pi_J),$$

naturally omitting $\pi_j - \pi_j$ just as we did δ_{jj}^{μ}. Thus, an agent's quantal response function is determined by the joint distribution of differences between payoff shocks.

Heterogeneity in the distributions of payoff shocks leads to heterogeneity in the distributions of the differences between payoff shocks. That is, the functional over P_ϵ induces a functional defined on the joint distributions for δ_j^{μ}. Let $P(\cdot)$ be a joint distribution function of δ_j^{μ} for some μ and any j in $1, \ldots, J$. We construct the functional $F[P]$ as follows. Let $\mathcal{P}$ be the set of $P_\epsilon(\cdot)$, distribution functions of ϵ_j^{μ}, which give rise to $P(\cdot)$. If there exists a $P_\epsilon \in \mathcal{P}$ such that $I_\epsilon[P_\epsilon] = 1$, then define

$I[P] = 1$ and

$$F[P] = \sum_{\{P_\epsilon : I_\epsilon[P_\epsilon]=1, P_\epsilon \in \mathcal{P}\}} F_\epsilon[P_\epsilon].$$

Otherwise, define $I[P] = 0$ and

$$F[P] = \int_{P_\epsilon \in \mathcal{P}} F_\epsilon[P_\epsilon] \, dP_\epsilon.$$

Note that there are joint distribution functions that could not apply to any δ_j^μ because they do not describe differences between iid random variables, and our definition implies $F = 0$ for these functions.

In an abuse of notation, we will use $\int dP$ as a shorthand for

$$\sum_{P:I[P]=1} + \int_{P:I[P]=0} dP.$$

This notation will be used even when there are mass points and is not meant to suggest their exclusion. It merely reflects our desire not to worry about the particular form the heterogeneity takes.[2]

Our definition of a representative agent can now be translated into a statement about the representative joint distribution of differences between payoff shocks. It means the representative agent would have δ_j distributed according to a joint distribution function $\hat{P}(\cdot)$ such that

$$(2.4) \qquad \hat{P} = \int F[P] \, P \, dP.$$

We can think of $\hat{P}$ as the population's expected joint distribution function for differences between payoff shocks, with respect to the induced probability mass / density functional F. The representative agent's quantal response function can in turn be

---

[2] The Lebesgue theory of integration allows a combination of a sum and an integral to be written simply as an integral with respect to a suitable measure. That is, the functional integral approach we take here corresponds to representing heterogeneity with a probability measure f_ϵ on the space of admissible distribution functions; f_ϵ would have discrete mass points for the support of I_ϵ and would be absolutely continuous where F_ϵ represents a density. Then, in place of $\sum_{P_\epsilon : I_\epsilon[P_\epsilon]=1} F_\epsilon[P_\epsilon] \cdot G[P_\epsilon] + \int_{P_\epsilon : I_\epsilon[P_\epsilon]=0} F_\epsilon[P_\epsilon] \cdot G[P_\epsilon] \, dP_\epsilon$, the standard notation for the expectation of G would be $\int G[P_\epsilon] f_\epsilon(dP_\epsilon)$.

found by using $\hat{P}$ in equation (2.3). This provides a working definition of a representative agent that is more useful than equation (2.2).

Given a functional that describes the heterogeneity of the population, we can use characteristic functions to identify a representative agent. This approach is effective because there is a bijection between distribution functions and characteristic functions. Let $\theta : \Re \rightarrow \mathbb{C}$ be the characteristic function of a payoff disturbance ϵ_j with distribution function $P_\epsilon(\cdot)$,

$$\theta(t) = E(e^{it\epsilon_j}).$$

Note that θ is a complex-valued function of a single real variable and $\theta(-t) = \bar{\theta}(t)$. It must be uniformly continuous, satisfy $\theta(0) = 1$ and $|\theta(t)| \leq 1$, and the quadratic form in u and v with $\theta(u-v)$ as its kernel must be non-negative definite. (These properties can be used to define an arbitrary characteristic function.) Take $\phi : \Re^{J-1} \rightarrow \mathbb{C}$ to be the characteristic function associated with the joint distribution $P(\cdot)$ of δ_j. We still write $\phi(t)$, now assuming $t = (t_1, \ldots, t_{J-1})$ to be a vector in $\Re^{J-1}$. We can express ϕ in terms of θ,

$$\begin{aligned}
\phi(t) &= E(e^{it \cdot \delta_j}) \\
&= E(e^{it_1 \delta_{1j}} \cdots e^{it_{J-1} \delta_{Jj}}) \\
&= E(e^{it_1(\epsilon_1 - \epsilon_j)} \cdots e^{it_{J-1}(\epsilon_J - \epsilon_j)}) \\
&= E(e^{it_1 \epsilon_1}) \cdots E(e^{it_{J-1} \epsilon_J}) \cdot E(e^{-i(\sum_{l=1}^{J-1} t_l)\epsilon_j}) \\
&= \theta(t_1) \cdots \theta(t_{J-1}) \cdot \theta(-\sum_{l=1}^{J-1} t_l).
\end{aligned}$$

(2.5)

In addition to the properties just mentioned, we also know that if $\sum_{l=1}^{J} r_l = 0$, then $\phi(r_1, \ldots, \widehat{r_j}, \ldots, r_J)$ is independent of j, because by equation (2.5) it has the same expansion in terms of θ for all j. If there are only two actions, $J = 2$, then ϕ is real

and positive because P is symmetric (see equation (2.5), where $\phi(t)$ would be the product of $\theta(t)$ and its complex conjugate). The functional F_ϵ induces a distribution over characteristic functions $\Psi_\epsilon[\theta] = F_\epsilon[P_\epsilon]$, with $\Upsilon_\epsilon[\theta] = I_\epsilon[P_\epsilon]$. Similarly, define $\Psi[\phi] = F[P]$ along with $\Upsilon[\phi] = I[P]$.

Let

$$(2.6) \qquad\qquad \hat{\phi}(t) = \int \Psi[\phi]\, \phi(t)\, d\phi$$

be the expectation of characteristic functions for δ_j^μ in the population. This representative characteristic function can be constructed by taking the integral pointwise, i.e., independently for every value of t. Fixing the input point t, we know that the functional integral $\int \Psi[\phi]\, \phi(t)\, d\phi$ always converges because $|\phi(t)| \leq 1$. (The abuse of notation in this context is the same as for the functional over distribution functions.)

2.3 A Representative Agent

2.3.1 Existence of a Representative Agent

The first issue to address is whether a representative agent exists. Theorem II.3 tells us that there is only one pathological type of heterogeneity for which the population does not have a representative agent. The joint distribution function $\hat{P}(\cdot)$ can always be constructed given the functional $F[P]$ describing the heterogeneity in the population, but there is a danger that it is not an admissible distribution function. Specifically, it may fail to have finite mean. A particular consequence of the theorem is the fact that a representative agent is sure to exist whenever only finitely many different distribution functions are in use in the population. Alternatively, relaxing the requirement that distributions of disturbances must have zero mean also ensures the existence of a representative agent.

Theorem II.3. *Define $\hat{P}(\cdot)$ as in equation (2.4). If $\hat{P}(\cdot)$ has finite mean, then*

a representative agent exists with δ_j distributed by $\hat{P}(\cdot)$ and having characteristic function $\hat{\phi}(t)$.

Proof. It is well known that the maps between distribution functions and characteristic functions are linear. Apply the Levy continuity theorem to equation (2.6). This requires $\hat{\phi}(t)$ to be continuous at $t = 0$, which we establish with Lemma II.15 in the Appendix. Lemma II.16 in the Appendix establishes that the mean of $\hat{P}(\cdot)$ is 0 if it exists, and thus $\hat{P}(\cdot)$ is admissible when this is the case. $\qquad\square$

Corollary II.4. *If $F[P] > 0$ for only finitely many joint distribution functions $P(\cdot)$, then a representative agent exists.*

Proof. The only source for divergence of the mean of $\hat{P}(\cdot)$ is the limit that results from $F[P] > 0$ for infinitely many P. All the joint distribution functions in the support of F have zero mean, so a finite linear combination of them also describes a random vector with zero mean. Then Theorem II.3 applies. $\qquad\square$

Taking a closer look at an example $\hat{P}(\cdot)$ that has divergent mean and thus fails to be an admissible joint distribution function offers insight into how such cases arise.

Example II.5. For simplicity, assume $J = 2$. The example works just as well with more pure strategies, but the notation becomes cluttered. Partition the set of joint distribution functions $P(\cdot)$ into $\mathcal{P}_y$ such that $P(\cdot) \in \mathcal{P}_y$ implies $P(e^y) \leq 1 - \alpha$ for some fixed positive $\alpha < \frac{1}{2}$. This partition is not uniquely determined, but as long as the $\mathcal{P}_y$ are non-empty, it will do. Consider the functional $F[P]$ where

$$\int_{\mathcal{P}_y} F[P]\, dP = \begin{cases} e^{-y} \text{ for } y \geq 0 \\ \\ 0 \text{ for } y < 0. \end{cases}$$

Then the mean of $\hat{P}(\cdot)$ is divergent because

$$\int_0^\infty \delta \, d\hat{P}(\delta) = \int_0^\infty \delta \int_0^\infty \int_{\mathcal{P}_y} F[P] P'(\delta) \, dP \, dy \, d\delta$$

$$\geq \int_0^\infty \int_{\mathcal{P}_y} F[P] \int_{e^y}^\infty \delta \, P'(\delta) \, d\delta \, dP \, dy$$

$$\geq \int_0^\infty \int_{\mathcal{P}_y} F[P] \alpha e^y \, dP \, dy$$

$$= \int_0^\infty \alpha \, dy.$$

Admissibility requires $\hat{P}(\cdot)$ to have zero mean, but when this fails, we shouldn't conclude that a representative quantal response function does not exist. Instead, we can relax the requirements of admissibility to guarantee that a representative agent always exists. The restriction to zero mean payoff disturbances is not necessary for the existence of a QRE, as fixed point theorems can be applied without it. The desire for unbiased disturbances appears to be aesthetic, and the possible inadmissibility of representative agents is an artifact of the way it is implemented. Consider replacing the zero mean assumption (a3) with the following alternative:

(a3') the Cauchy principal value of the mean of each payoff disturbance is zero[3], and

$$\lim_{\gamma \to \infty} \gamma \operatorname{Prob} \left\{ |\epsilon_j^\mu| \geq \gamma \right\} = 0 \text{ for each } \epsilon_j^\mu.$$

Assumption (a3') holds whenever assumption (a3) is satisfied, so this is a weaker condition to impose on the payoff disturbances. Even though the mean of ϵ_j^μ may blow up under assumption (a3'), these disturbances are still unbiased, and their likelihood still decays sufficiently quickly as they get large.

Definition II.6. We say payoff disturbances are *weakly admissible* if assumptions (a1), (a2) and (a3') hold.

---

[3]The Cauchy principal value of an improper integral $\int_{-\infty}^\infty f(t) \, dt$ is defined as $\lim_{T \to \infty} \int_{-T}^T f(t) \, dt$.

With just this slight relaxation of admissibility, we always get a representative agent.

Corollary II.7. *Allow weakly admissible payoff disturbances. A representative agent exists with δ_j distributed by $\hat{P}(\cdot)$ and having characteristic function $\hat{\phi}(t)$.*

Proof. Lemma II.16 shows that $\hat{P}(\cdot)$ always satisfies the weak admissibility assumption (a3'). In turn, there exists a joint distribution of $(\epsilon_1, \ldots, \epsilon_J)$ that satisfies (a3') and is consistent with δ_j being distributed by $\hat{P}(\cdot)$. □

2.3.2 Payoff Disturbances for the Representative Agent

We have defined a representative agent with the property that the agent's choice of strategy is representative of the population as a whole. We now show that this is not equivalent to having representative noise in the underlying payoffs. We say $\hat{P}_\epsilon(\cdot)$ is a *representative distribution of payoff shocks* if it is a (weakly) admissible distribution function and

$$(2.7) \qquad \hat{P}_\epsilon = \int F_\epsilon[P_\epsilon] \, P_\epsilon \, dP_\epsilon.$$

By applying the Levy continuity theorem here too, we find that a representative distribution of payoff shocks has characteristic function $\hat{\theta}(t) = \int \Psi_\epsilon[\theta] \, \theta(t) \, d\theta$. With this groundwork in place, we are ready for Theorem II.8, which says that a representative quantal response function does not arise from a representative distribution of payoff shocks.

Theorem II.8. *A representative agent has a representative distribution of payoff shocks if and only if the population is homogeneous.*

Proof. Let Θ be the set of characteristic functions of ϵ_j that give rise to a given $\phi(\cdot)$. Using equation (2.5), $\Theta = \{\theta : \phi(t) = \left(\prod_{l=1}^{J-1} \theta(t_l)\right) \cdot \theta(-\sum_{l=1}^{J-1} t_l)\}$. From the

relationships between the functionals, we have

$$\Psi[\phi] = \int_{\Theta} \Psi_{\epsilon}[\theta]\, d\theta \qquad\qquad \text{if } \Upsilon[\phi] = 0$$

$$\sum_{\{\theta \in \Theta : \Upsilon_{\epsilon}[\theta]=1\}} \Psi_{\epsilon}[\theta] \qquad\qquad \text{if } \Upsilon[\phi] = 1.$$

We can then express a representative agent's characteristic function for δ_j as

$$\hat{\phi}(t) = \int \Psi_{\epsilon}[\theta] \left(\prod_{l=1}^{J-1} \theta(t_l) \right) \theta(-\sum_{l=1}^{J-1} t_l)\, d\theta.$$

But

(2.8)
$$\int \Psi_{\epsilon}[\theta] \left(\prod_{l=1}^{J-1} \theta(t_l) \right) \theta(-\sum_{l=1}^{J-1} t_l)\, d\theta \neq \left(\prod_{l=1}^{J-1} \int \Psi_{\epsilon}[\theta]\, \theta(t_l)\, d\theta \right) \cdot \int \Psi_{\epsilon}[\theta]\, \theta(-\sum_{l=1}^{J-1} t_l)\, d\theta$$

unless for each t_l, $\theta(t_l)$ is the same for all θ in the support of Ψ_{ϵ}. Since t_l is an arbitrary variable, this would mean there could only be one function in the support of Ψ_{ϵ}, i.e., no heterogeneity of distributions of payoff shocks in the population. $\qquad\square$

In light of the fact that a representative agent for a heterogeneous population does not have a representative distribution of payoff shocks, the question arises as to what distribution of payoff shocks could actually produce a representative agent. According to the next result, if there are enough actions and there is heterogeneity of the $\delta_{jj'}$, then the representative agent cannot arise from any distribution of payoff shocks that is iid across the set of actions. Theorem II.9 says that if there are just two actions, there is an iid distribution of payoff shocks (possibly many such distributions) that generates the representative agent. But, if there are at least four actions, assuming heterogeneity of the $\delta_{jj'}$, it is impossible for an iid distribution of payoff shocks to generate the representative agent.[4]

---

[4]Examples indicate that when there are three actions, the representative agent usually cannot arise from iid shocks, but we cannot rule out special cases of heterogeneity for which the representative agent is compatible with iid disturbances.

Theorem II.9. *Given a representative agent, if $J = 2$, there exists a distribution of payoff shocks iid across all actions and each with characteristic function $\hat{\vartheta}(\cdot)$ such that*

$$(2.9) \qquad \hat{\phi}(t) = \left(\prod_{l=1}^{J-1} \hat{\vartheta}(t_l) \right) \cdot \hat{\vartheta}\left(-\sum_{l=1}^{J-1} t_l\right).$$

But, when $J \geq 4$, there is no $\hat{\vartheta}(\cdot)$ that satisfies equation (2.9) unless every $P_\epsilon(\cdot)$ in the support of F_ϵ gives the same distribution of the $\delta_{jj'}$.

Proof. When $J = 2$, we must find a $\hat{\vartheta}(\cdot)$ such that $\hat{\phi}(t_1) = \hat{\vartheta}(t_1) \cdot \hat{\vartheta}(-t_1)$. Recall $J = 2$ implies that all $\phi(\cdot)$ are real and positive, and hence so is $\hat{\phi}$. It suffices to take $\hat{\vartheta}(t_1) = \hat{\vartheta}(-t_1) = \sqrt{\hat{\phi}(t_1)}$.

Now consider $J \geq 4$. Given that individual agents do have payoff shocks that are iid across all actions, any $\phi(\cdot)$ in the population can be expressed in terms of $\theta(\cdot)$ with equation (2.5). Specifically, $\phi(a, -a, a, 0, \cdots, 0) = (\theta(a)\theta(-a))^2$. Similarly, $\phi(a, 0, \cdots, 0) = \theta(a)\theta(-a)$. Thus,

$$\phi(a, -a, a, 0, \cdots, 0) = (\phi(a, 0, \cdots, 0))^2.$$

But

$$\int \Psi[\phi]\,\phi(a, -a, a, 0, \cdots, 0)\,d\phi \neq \left(\int \Psi[\phi]\,\phi(a, 0, \cdots, 0)\,d\phi \right)^2$$

unless there is no variance of $\theta(a)\theta(-a)$ in the population. Note that $\delta_{jj'}$ has characteristic function $\theta(t)\theta(-t)$. Thus, if there are two distribution functions in the support of $F_\epsilon[P_\epsilon]$ that give different distributions of $\delta_{jj'}$, then for some a,

$$\hat{\phi}(a, -a, a, 0, \cdots, 0) \neq \left(\hat{\phi}(a, 0, \cdots, 0) \right)^2.$$

This would mean $\hat{\phi}(\cdot)$ could not be expressed as $\left(\prod_{l=1}^{J-1} \hat{\vartheta}(t_l) \right) \cdot \hat{\vartheta}(-\sum_{l=1}^{J-1} t_l)$ for any $\hat{\vartheta}(\cdot)$. □

Theorem II.9 sounds a cautionary note that even if we believe all agents have noise in their payoffs that is iid across their actions, heterogeneity of the agents leads the population as a whole to behave as if payoff disturbances were not iid across actions.

We desired agents with payoff noise iid across actions because this assumption imposes restrictions on behavior that can be tested empirically. Although it turns out the representative agent may not have payoff noise iid across actions, the representative agent notion still has empirical content because some properties are inherited from the underlying agents.

2.3.3 Regularity of a Representative Agent

Goeree, et. al. [11] introduce four axioms which define a *regular quantal response function* $Q^\mu : \Re^J \to \triangle^{J-1}$ without reference to payoff noise:

(A1) *Interiority*: $Q_j^\mu(\pi) > 0$ for all $j = 1, \ldots, J$ and for all $\pi \in \Re^J$.

(A2) *Continuity*: $Q_j^\mu(\pi)$ is a continuous and differentiable function for all $\pi \in \Re^J$.

(A3) *Responsiveness*: $\frac{\partial Q_j^\mu(\pi)}{\partial \pi_j} > 0$ for all $j = 1, \ldots, J$ and for all $\pi \in \Re^J$.

(A4) *Monotonicity*: $\pi_j > \pi_{j'}$ implies $Q_j^\mu(\pi) > Q_{j'}^\mu(\pi)$, for all $j, j' = 1, \ldots, J$.

They argue that all quantal response functions obey Continuity and weakly obey Responsiveness. If the density of payoff disturbances has full support, then Interiority and Responsiveness are strictly satisfied. When payoff disturbances are iid across actions, then the quantal response function obeys Monotonicity as well.

We now show that any regularity property that holds for the underlying agents in the population also holds for the representative agent.

Theorem II.10. *If a regularity axiom $\{(A1), (A2), (A3), or (A4)\}$ applies to Q^μ*

for all μ (i.e., for $\mu = \mu(P)$ whenever $P(\cdot)$ is in the support of F), then that axiom applies to the representative agent's quantal response function $\hat{Q}$.

Proof. Continuity holds for all quantal response functions as a result of the admissibility assumption (a2) that distributions of payoff noise must be absolutely continuous [11]. Interiority, Responsiveness, and Monotonicity each follow from equations (2.1) and (2.2), which define a representative agent. Essentially, we just use the fact that an integral (or sum) must be positive if the integrand (summand) is always positive. For Responsiveness, we pass the partial derivative inside the integral and sum in equation (2.1). For Monotonicity, we express $\hat{Q}_j(\pi) - \hat{Q}_{j'}(\pi)$ using equation (2.1) and then pair up terms to form a single sum and integral. $\qquad\square$

Theorem II.10 tells us that in our framework, the representative agent's quantal response function always satisfies Monotonicity. It is this Monotonicity property that carries empirical content. In principle, subjects in an experiment could violate Monotonicity and choose actions with lower payoffs more often than actions with higher payoffs. This would be inconsistent with the predicted behavior of the representative agent.

2.4 Asymmetric Games

All of these results, initially presented in the context of a single population, apply to general asymmetric games. Consider a normal form game with n populations of agents. The strategy sets may differ across players, so we let $S_i = \{s_{i1}, \ldots, s_{iJ_i}\}$ be the set of pure strategies available to agents in population i. Now $x = x_1 \times \cdots \times x_n$ denotes the mixed strategy profile across all n populations in the game, with each mixed strategy vector $x_i = (x_{i1}, \ldots, x_{iJ_i}) \in \triangle^{J_i - 1}$.

The vector $\pi_i = \pi_{i1}, \ldots, \pi_{iJ_i}$ denotes the payoff to agents in population i from

their available pure strategies. Now, $\pi_i : \triangle^{J_1-1} \times \cdots \times \triangle^{J_n-1} \to \Re^{J_i}$. Agent μ observes payoff disturbances ϵ^μ_{ij} for each strategy. The variables $\delta^\mu_{ijj'} = \epsilon^\mu_{ij} - \epsilon^\mu_{ij'}$ and $\delta^\mu_{ij} = \left(\delta^\mu_{i1j}, \ldots, \widehat{\delta^\mu_{ijj}}, \ldots, \delta^\mu_{iJ_ij} \right)$ are similarly defined for each population. The quantal response functions $Q^\mu_{ij}(\pi_i)$ depend on the payoffs in population i, which in turn depend on the entire n-population mixed strategy profile x. Thus, equation (2.3) applies for each i. The functionals also have to be indexed for the population so that $F^i_\epsilon[P_\epsilon]$ describes the fraction of population i with payoff disturbances distributed by P_ϵ and each F^i_ϵ induces a $F^i[P]$, $\Psi^i_\epsilon[\theta]$, and $\Psi^i[\phi]$. Equation (2.1) now applies for each i, making $(Q_{i1}, \ldots, Q_{iJ_i})$ the ith population aggregate quantal response. The fixed point equation defining a quantal response equilibrium becomes $x_{ij} = Q_{ij}(\pi_i(x))$ for all $i \in 1, \ldots, n$ and all $j \in 1, \ldots, J_i$.

Theorem II.3 now describes the existence of a representative agent for population i with δ_{ij} distributed by the joint distribution function $\hat{P}^i(\cdot)$ and having characteristic function $\hat{\phi}^i(t)$. For each i, these representative functions are given by equations (2.4) and (2.6), just as before. And Theorems II.8, II.9, and II.10 apply to the representative agent from any given population.

However, while obtaining representative agents for each role in the game, we caution that there is no reason to assume the existence of a single representative agent the same for all players of an asymmetric game. Such an assumption would deny heterogeneity across the different roles of the game. And given the fact that a representative agent does not have a representative distribution of payoff shocks and that different players may have different strategy sets, it's not clear exactly what is meant by a single representative agent for all players of an asymmetric game. The problem is that we want to have a representative quantal response function $\hat{Q}_{ij}(\cdot)$ that is independent of i, for each fixed j, but this does not make sense when the set

of actions j depends on i.

Here, we say that there may be a single representative agent for all players if there exist $\hat{P}^i$ such that $J_{i_1} \leq J_{i_2}$ implies that for all $(\varpi_1, \ldots, \varpi_{J_{i_1}})$ and any $j \leq J_{i_1}$,

$$\hat{Q}_{i_1 j}(\varpi_1, \ldots, \varpi_{J_{i_1}}) = \lim_{\varpi_{J_{i_1}+1} \to -\infty} \cdots \lim_{\varpi_{J_{i_2}} \to -\infty} \hat{Q}_{i_2 j}(\varpi_1, \ldots, \varpi_{J_{i_2}}).$$

With this definition, we single out the representative agent from a population i that maximizes J_i. We can think of this agent with δ_{ij} distributed by $\hat{P}^i(\cdot)$ as representative of all players in all roles by assuming that when playing a role with too few possible actions, the agent imagines there are additional actions with infinitely negative payoffs. In the particular case that all players have the same number of actions, $J_i = J$ for all i, a single representative agent for all players would have differences in payoff shocks jointly distributed by a $\hat{P}$ that satisfies $\hat{P} = \hat{P}^i$ for all i. The representative agent for each population would have to be the same. There are plenty of QRE which are incompatible with identical representative agents for all populations, as the following game illustrates.

Example II.11.

<div align="center">

Asymmetric Matching Pennies

</div>

	Left	Right
Up	9, -1	-1, 1
Down	-1, 1	1, -1

Goeree, et. al. [11] analyze the asymmetric matching pennies game shown above and find the set of possible QRE is the rectangle $\frac{1}{6} < p < \frac{1}{2}$, $\frac{1}{2} < q < 1$, where p is the probability the column player chooses left and q is the probability the row player chooses up. However, given the restriction that representative row and column players have the same quantal response function, a QRE must satisfy the additional
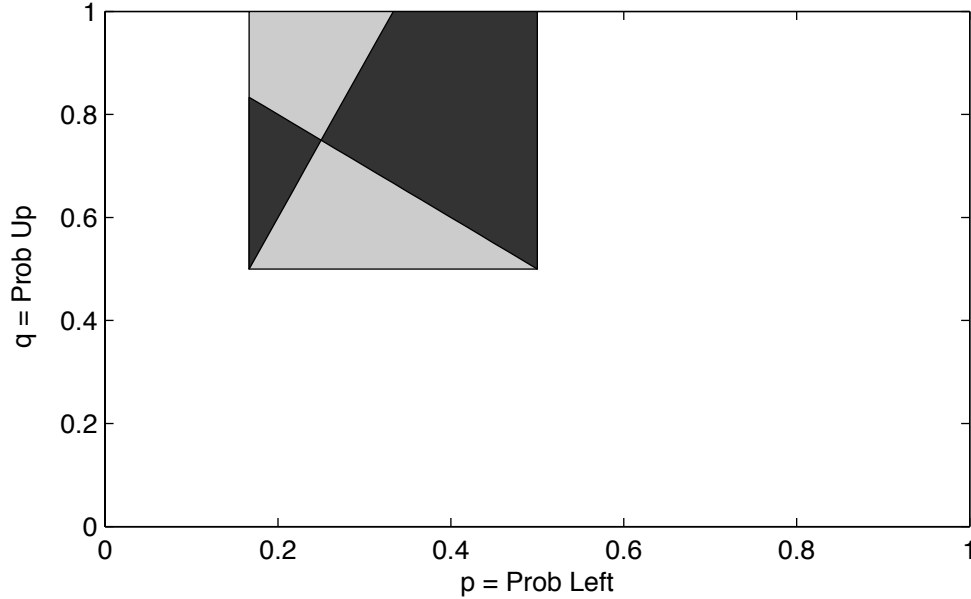
Figure 2.1: On the horizontal axis is the probability the column player chooses left and on the vertical axis is the probability the row player chooses up. The entire shaded area represents the set of possible QRE in the given asymmetric matching pennies game when row and column players may differ. The darkened area represents the subset of these QRE consistent with the assumption that representative row and column players have the same quantal response function.

constraint $q < 3p$ if and only if $q > 1 - p$. (See Figure 2.1.) This is because $q < 3p$ means that $\pi_{2R} - \pi_{2L} < \pi_{1U} - \pi_{1D}$, i.e., the cost of an error is higher for the row player, and must lead to relatively fewer errors by the row player, $q > 1 - p$. The converse holds equivalently. Note also that if both row and column players use identical logit responses, the set of possible QRE is reduced to a curve extending from the center of the strategy space to the Nash Equilibrium $(p = \frac{1}{6}, q = \frac{1}{2})$ [22]. In summary, unlike the existence of a representative agent for each population, we do not necessarily have a single representative agent for all players.

2.5 Logit Responders

Most of the literature on QRE has assumed that the payoff disturbances are all independently drawn from an extreme value distribution, which generates tractable

logit response functions:

$$(2.10) \qquad x_{ij}(\pi_i) = \frac{e^{\lambda \pi_{ij}}}{\sum_{l=1}^{J_i} e^{\lambda \pi_{il}}}$$

where the parameter λ can be interpreted as the agents' level of rationality. As λ goes to infinity, agents best respond perfectly, producing a Nash Equilibrium. Conversely, as λ tends to zero, agents play the uniform mixed strategy, choosing each action with the same probability without regard to payoffs.

We now assume agents' quantal response functions take this logit form, but we preserve heterogeneity in the populations by allowing the agents to have their own individual rationality parameters. Thus,

$$(2.11) \qquad Q_{ij}^{\mu}(\pi_i) = \frac{e^{\lambda_{\mu} \pi_{ij}}}{\sum_{l=1}^{J_i} e^{\lambda_{\mu} \pi_{il}}}.$$

For the purposes of this section, it suffices to consider finite populations of agents, so

$$(2.12) \qquad x_{ij} = \frac{1}{m_i} \sum_{\mu=1}^{m_i} Q_{ij}^{\mu}$$

for all i and j.

It would be straightforward to apply the results from Section 2.3 and identify representative agents. In a truly heterogeneous population, i.e., with logit parameters not all degenerate, the representative agent will not be a logit responder. In this section, we see what happens when a theorist tries to force a homogeneous logit equilibrium model on a population that is actually heterogeneous. Because the homogeneous logit equilibrium is a mis-specified model of the populations we're assuming, the value of the single logit parameter will vary with the game being considered. But, a single logit parameter value can explain any particular choice probabilities between two actions if payoff monotonicity is preserved (i.e., if choice probabilities

are increasing in the payoffs). For this reason, we restrict attention to a population with two pure strategies taking part in a (possibly larger) fixed game.

We identify a downward bias in the single logit parameter determined by the mis-specified homogeneous model as compared with the average of the true logit parameters in use. Thus, the population seems to behave less rationally if the modeler believes the agents are all alike when in fact they each have their own levels of rationality. This bias is exacerbated as the magnitude of the difference in payoffs between the two actions grows.

First, we present a formula relating the logit parameter of the homogeneous model to the true logit parameters and the equilibrium payoffs in the heterogeneous model. Let $J_i = 2$ in a particular population i. Fix equilibrium choice probabilities and payoffs in accordance with equations (2.11) and (2.12), and denote them x_i^* and π_i^* respectively in population i. Assume a game in which the equilibrium payoffs to the two actions in population i are not equal, $\pi_{i1}^* \neq \pi_{i2}^*$. Denote by λ the logit parameter of the homogeneous model describing behavior in population i. We can use equation (2.10) to express λ in terms of the choice probabilities and payoffs in population i. The ratio of the choice probabilities is $\frac{x_{i1}^*}{x_{i2}^*} = e^{\lambda(\pi_{i1}^* - \pi_{i2}^*)}$. Thus,

$$(2.13) \qquad \lambda = \frac{1}{\pi_{i1}^* - \pi_{i2}^*} \ln \left(\frac{x_{i1}^*}{x_{i2}^*} \right).$$

Equation (2.13) could also be derived as the maximum likelihood estimate of the homogeneous logit parameter given data on the equilibrium choice probabilities and payoffs in population i. With sufficiently many observations of choice probabilities and payoffs, these data should accurately reflect the equilibrium satisfying equations (2.11) and (2.12). These equations, (2.11) and (2.12), give us the actual ratio

of choice probabilities in population i. We plug into equation (2.11) to get

$$Q_{i1}^{\mu}(\pi_i^*) = \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi} + 1}$$

and

$$Q_{i2}^{\mu}(\pi_i^*) = \frac{1}{e^{\lambda_\mu \Delta\pi} + 1},$$

where we let $\Delta\pi = \pi_{i1}^* - \pi_{i2}^*$ to simplify the notation. Then, by equation (2.12),

$$\frac{x_{i1}^*}{x_{i2}^*} = \frac{\frac{1}{m_i} \sum_{\mu=1}^{m_i} \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi}+1}}{\frac{1}{m_i} \sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1}}.$$

Finally, we obtain our desired formula:

(2.14)
$$\lambda = \frac{1}{\Delta\pi} \ln\left(\frac{\sum_{\mu=1}^{m_i} \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi}+1}}{\sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1}} \right).$$

Observe that λ depends both on the heterogeneous logit parameters $\{\lambda_\mu\}$ and the equilibrium payoff difference $\Delta\pi$. We sometimes refer to the function given by equation (2.14) as $\lambda(\{\lambda_\mu\}, \Delta\pi)$.

Our next result helps us interpret this formula. Theorem II.12 says that this homogeneous logit parameter is always less than the average of the heterogeneous logit parameters actually used by the agents. Moreover, the size of this bias in the homogeneous model depends on the equilibrium payoffs. When the magnitude of the difference in payoffs between the two actions gets large, the homogeneous logit parameter approaches the smallest of the heterogeneous logit parameters in the population. In this limit, the population behaves like its single most irrational agent. On the other hand, when the magnitude of the payoff difference gets small, the homogeneous logit parameter approaches the average of the agents' true logit parameters.

Theorem II.12. *Consider a quantal response equilibrium in accordance with equations (2.11) and (2.12) such that population i has two actions with different equilib-*

rium payoffs, i.e., $J_i = 2$, and $\Delta\pi \neq 0$. Let $\bar{\lambda} = \frac{1}{m_i}\sum_{\mu=1}^{m_i}\lambda_\mu$ be the average of the heterogeneous logit parameters used by the agents in population i, and let λ be the homogeneous logit parameter that explains the population's choice probabilities for these particular payoffs. Then

(2.15)
$$\lambda \leq \bar{\lambda}$$

with equality if and only if $\lambda_1 = \lambda_2 = \ldots = \lambda_{m_i}$.[5] Additionally,

(2.16)
$$\lim_{\Delta\pi \to \pm\infty} \lambda = \min\{\lambda_\mu\}$$

and

(2.17)
$$\lim_{\Delta\pi \to 0} \lambda = \bar{\lambda}.$$

Proof. Equation (2.14) gives the exact value of λ. We twice apply Jensen's Inequality to pieces of this expression in order to derive (2.15).

Without loss of generality, assume action 1 has the higher equilibrium payoff so that $\Delta\pi > 0$. Then $\frac{1}{e^{\xi\Delta\pi}+1}$ is a concave up function of ξ. Applying Jensen's Inequality to this function,

(2.18)
$$\frac{1}{m_i}\sum_{\mu=1}^{m_i}\frac{1}{e^{\lambda_\mu\Delta\pi}+1} \geq \frac{1}{e^{\bar{\lambda}\Delta\pi}+1}$$

with equality if and only if $\lambda_1 = \lambda_2 = \ldots = \lambda_{m_i}$. Similarly, $\frac{e^{\xi\Delta\pi}}{e^{\xi\Delta\pi}+1}$, being equivalent to $1 - \frac{1}{e^{\xi\Delta\pi}+1}$, is a concave down function of ξ. So Jensen's Inequality implies

(2.19)
$$\frac{1}{m_i}\sum_{\mu=1}^{m_i}\frac{e^{\lambda_\mu\Delta\pi}}{e^{\lambda_\mu\Delta\pi}+1} \leq \frac{e^{\bar{\lambda}\Delta\pi}}{e^{\bar{\lambda}\Delta\pi}+1}$$

---

[5]The convention of using the parameter λ to represent a player's rationality is somewhat arbitrary in the sense that a modeler could just as well have defined $\kappa = e^\lambda$ to be the rationality parameter. Proposition II.17 in the Appendix establishes an inequality analogous to (2.15), showing downward bias for such an alternative rationality parameter.

with equality if and only if $\lambda_1 = \lambda_2 = \ldots = \lambda_{m_i}$. When we plug into equation (2.14), the denominators on the right-hand sides of (2.18) and (2.19) cancel, giving us

$$\lambda \leq \frac{1}{\Delta \pi} \ln \left(e^{\bar{\lambda} \Delta \pi} \right)$$
$$= \bar{\lambda}.$$

Here again, equality holds exactly when $\lambda_1 = \lambda_2 = \ldots = \lambda_{m_i}$.

We prove the limits in (2.16) and (2.17) in the Appendix. $\qquad\square$

Theorem II.12 describes a downward bias in the determination of a homogeneous logit parameter when agents are actually heterogeneous. The less rational agents seem to leave a larger mark on the aggregate population behavior. This bias gets worse when one action's equilibrium payoff gets much larger than the other's. Conversely, the bias disappears as the payoff difference tends to zero.

Our formula for λ, equation (2.14), also allows us to ask whether a determination of the homogeneous logit parameter from data on a choice between two actions restricts the set of possible logit parameters for members of the population. The next result says it very well may. A large value of the homogeneous logit parameter imposes a minimum possible value on the set of heterogeneous parameters. Conversely, a small homogeneous logit parameter precludes any individual agent from having too large a value. For intermediate homogeneous logit parameters, however, we cannot rule out any parameters for a single agent. Naturally, these bounds depend on the population size and are much less restrictive for a large population.

Theorem II.13. *Retain the context of Theorem II.12. If $e^{\lambda |\Delta \pi|} > 2m_i - 1$, then*

$$\min\{\lambda_\mu\} \geq \frac{1}{|\Delta \pi|} \ln \left(\frac{1}{m_i} \left(e^{\lambda |\Delta \pi|} - (m_i - 1) \right) \right).$$

If $e^{\lambda|\Delta\pi|} < \frac{m_i+1}{m_i-1}$, *then*

$$\max\{\lambda_\mu\} \leq \frac{1}{|\Delta\pi|}\ln\left(\frac{(m_i+1)e^{\lambda|\Delta\pi|}-(m_i-1)}{m_i+1-(m_i-1)e^{\lambda|\Delta\pi|}}\right).$$

Proof. See Appendix. □

Homogeneous logit parameters are estimated in much of the experimental literature on two-by-two games, although often with data pooled across many populations and many games. Theorem II.13 applies to a homogeneous logit parameter calculated for a single population in a particular game. If we believe that agents use logit responses, but are heterogeneous in their levels of rationality, this theorem translates a mis-specified homogeneous logit parameter into restrictions on the set of possible logit parameters in a finite population.

To illustrate these results, we can compare a homogeneous logit model fit to data in a two-by-two symmetric game to compatible heterogeneous logit models featuring two types of responders – one with a high rationality parameter and the other with a low one. To make the example as simple as possible, we assume exactly half the agents are of each type (though with the data coming from an experiment on 214 subjects, we have no reason to actually believe there are just two types). We consider Guyer and Rapoport's "No Conflict" game, Game #6 in their series of experiments [13]. The payoff matrix is:

<div align="center">

No Conflict

	A_2	B_2
A_1	4, 4	2, 3
B_1	3, 2	1, 1

</div>

The players have a dominant strategy choosing action A. Guyer and Rapoport observe action A played 90% of the time. Choosing to model this as a homogeneous logit equilibrium, we have an equilibrium payoff difference $\Delta\pi = 1$ (as the payoff to A happens to always exceed the payoff to B by one), and thus $\lambda = \ln(9)$ in accordance with equation 2.13.[6]

Plugging $\Delta\pi = 1$ and $\lambda = \ln(9)$ into equation 2.14 produces an equation implicitly relating λ_1 and λ_2. Figure 2.2 shows possible values of these heterogeneous logit parameters. Pairs of λ_1 and λ_2 values are determined by fixed x-values in the graph. Larger x-values correspond to greater dispersion in the heterogeneous logit parameter values, but the scaling along this axis is arbitrary. We can see that the average of λ_1 and λ_2 always exceeds $\ln(9)$, and the lower value is bounded below by $\ln(4)$ while the higher value may be arbitrarily large. Guyer and Rapoport's data thus puts a bound on how irrational the low-type agents can be, and they only approach this bound if the other agents are hyper-rational.

Because a homogeneous logit model is mis-specified in the presence of heterogeneity, estimates of a single rationality parameter do not translate across different game environments. Theorems II.12 and II.13 imply the following result, which tells us that an estimate of a homogeneous rationality parameter in a particular game environment places no restriction on such an estimate in an alternative game environment, even with a working assumption that agents' rationality levels are fixed across games. Theorem II.14 states that the set of heterogeneous logit parameters that is consistent with a given logit equilibrium in any one game could in some other game give rise to behavior consistent with any other homogeneous logit parameter.

---

[6]Goeree and Holt [9] estimate a homogeneous logit parameter from data pooled across 37 games, including this one, from Guyer and Rapoport's study. We obtain a different value of the homogeneous logit parameter because we use data from just this one game.
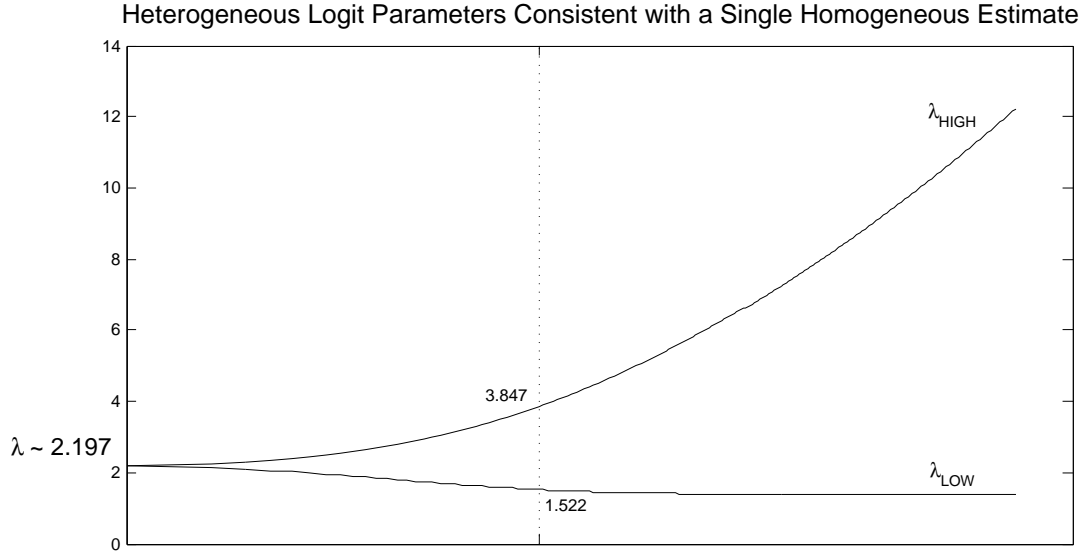
Figure 2.2: Possible values of a pair of logit parameters (determined at any fixed x-value) that would be consistent with a homogeneous $\lambda = \ln(9)$, when $\Delta\pi = 1$. These values fit data from Guyer and Rapoport's (1972) "No Conflict" game.

Theorem II.14. *Consider normal form games for which population i has two actions, $J_i = 2$. For any logit equilibrium with population i having payoff difference $\Delta\pi^* \neq 0$ and rationality parameter $\lambda^* > 0$ in such a game Γ, and any alternative value $\lambda' > 0$, there exists a set of heterogeneous logit parameters $\{\lambda'_\mu\}$ that are consistent with the homogeneous logit model applied to population i in Γ,*

$$\text{(2.20)} \qquad\qquad \lambda\left(\{\lambda'_\mu\}, \Delta\pi^*\right) = \lambda^*,$$

and there exists a game Γ' with a heterogeneous logit equilibrium in which population i has payoff difference $\Delta\pi' \neq 0$, such that

$$\text{(2.21)} \qquad\qquad \lambda\left(\{\lambda'_\mu\}, \Delta\pi'\right) = \lambda'.$$

Proof. See Appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Recall that $\lambda\left(\{\lambda_\mu\}, \Delta\pi\right)$ gives the homogeneous logit parameter that produces the same equilibrium choice probabilities as the heterogeneous logit parameters $\{\lambda_\mu\}$ when the equilibrium payoff difference is $\Delta\pi$. Thus, equation (2.20) means that any estimate of a homogeneous rationality parameter in a given game environment can be explained by some set of heterogeneous logit parameters, and equation (2.21) means that these heterogeneous logit parameters could be consistent with any other homogeneous parameter in an alternative game environment. We should not expect mis-specified parameter estimates to accurately describe behavior across all games.

2.6   Discussion

We have proposed a model of heterogeneous populations playing quantal response equilibria. The chapter contributes general results that apply to quantal response equilibria without specification of their functional form as well as particular results that are specific to the logit response model.

We have paid extra attention to the logit specification because it is so commonly employed in practice. The representative agent for a population of heterogeneous logit responders is not another logit responder. In the case of heterogeneous logit responders choosing between two pure strategies, we have obtained a formula (equation 2.14) relating a mis-specified homogeneous logit parameter to the actual heterogeneous parameters in the population. Maximum likelihood estimation could be used to fit a homogeneous logit parameter to the behavior of heterogeneous agents choosing between any number of pure strategies, but a closed form solution is not generally possible. Our formula provides insights in two directions. It tells us that the homogeneous model is biased towards less rationality, as the homogeneous logit parameter is always less than the average of the heterogeneous ones. It also allows us

to bound the possible values of the true logit parameters if we have a mis-specified homogeneous model already in place.

These results are applicable to experimental work in which a homogeneous logit model has been fit to data. One particular extension is to explicitly model the existence of clueless players by giving some fraction of the agents a logit parameter of zero. This would address the common problem of some subjects not understanding the game they are playing [3].

Working with a general model that does not assume that quantal responses take any particular functional forms, we have found that representative agents exist for heterogeneous populations if we allow weakly admissible payoff disturbances. A representative agent chooses strategies in the same proportions as the entire population, but does not have payoff disturbances distributed in the same proportions as the population. In games with many pure strategies, representative behavior cannot arise from any iid distribution of disturbances.

This impossibility of having a representative agent with disturbances iid across actions stems from the fact that averaging probability distributions almost never preserves independence. Thus, if we believe populations of agents are heterogeneous, but desire representative-agent models, we must be willing to consider noise terms that are jointly dependent across actions. Our findings support the use of regular quantal response functions. Regular quantal response equilibrium does generate falsifiable predictions and is consistent with the representative-agent framework.

2.7   Appendix

Lemma II.15. $\hat{\phi}(t)$ *is continuous at* $t = 0$.

Proof. Recall that $|\phi(t)| \leq 1$ and $\phi(0) = 1$ for all ϕ and thus for $\hat{\phi}$ as well. We will

show for all $h > 0$, there exists $k > 0$ such that $\|t\| < k$ implies $\operatorname{Re}\left\{\hat{\phi}(t)\right\} > 1 - h$.

Let $\mathcal{K}$ be a compact subset of characteristic functions ϕ such that $\int_{\mathcal{K}} \Psi[\phi]\, d\phi > 1 - \frac{h}{4}$. Because all the ϕ are continuous at $t = 0$, we can choose $k[\phi] > 0$ such that $\operatorname{Re}\left\{\phi(t)\right\} > 1 - \frac{h}{2}$ for all $\|t\| < k[\phi]$ and $\phi \mapsto k[\phi]$ is continuous. Then take $k = \min_{\phi \in \mathcal{K}} k[\phi]$, and $k > 0$ because the minimum is taken over a compact space and the extreme value theorem applies. We then obtain for all $\|t\| < k$,

$$
\begin{aligned}
\operatorname{Re}\left\{\hat{\phi}(t)\right\} &= \int_{\phi \in \mathcal{K}} \operatorname{Re}\left\{\phi(t)\right\} \Psi[\phi]\, d\phi + \int_{\phi \notin \mathcal{K}} \operatorname{Re}\left\{\phi(t)\right\} \Psi[\phi]\, d\phi \\
&> \left(1 - \frac{h}{2}\right)\left(1 - \frac{h}{4}\right) + (-1)\left(\frac{h}{4}\right) \\
&= 1 - h + \frac{h^2}{8} > 1 - h.
\end{aligned}
$$

$\square$

Lemma II.16. *The Cauchy principal value of the mean of $\hat{P}(\cdot)$ is 0. Additionally, if the random vector $(\bar{\delta}_{1j}, \ldots, \widehat{\bar{\delta}_{jj}}, \ldots, \bar{\delta}_{Jj})$ is distributed according to $\hat{P}(\cdot)$, then*

$$
\lim_{\gamma \to \infty} \gamma \operatorname{Prob}\left\{|\bar{\delta}_{j'j}| \geq \gamma\right\} = 0 \text{ for all } j'.
$$

Proof. A property of characteristic functions is that $\left[\frac{\partial}{\partial t_{j'}} \phi(t)\right]_{t=0}$ exists if and only if:

(i) PV $\langle \delta_{j'j} \rangle$ exists and

(ii) $\lim_{\gamma \to \infty} \gamma \operatorname{Prob}\left\{|\delta_{j'j}| \geq \gamma\right\} = 0$,

and when these conditions are satisfied, $\left[\frac{\partial}{\partial t_{j'}} \phi(t)\right]_{t=0} = i\, \text{PV}\, \langle \delta_{j'j} \rangle$ [28, 24]. So, it suffices to show $\left[\frac{\partial}{\partial t_{j'}} \hat{\phi}(t)\right]_{t=0} = 0$ for all j'. Differentiability of $\hat{\phi}(t)$ follows from the differentiability of all ϕ in the support of Ψ, using an argument completely

analogous to the proof of continuity of $\hat{\phi}(t)$, Lemma II.15. Thus, $\left[\frac{\partial}{\partial t_{j'}}\hat{\phi}(t)\right]_{t=0} =$ $\int \Psi[\phi]\left[\frac{\partial \phi}{\partial t_{j'}}\right]_{t=0} d\phi$. For all ϕ in the support of Ψ, all j', $\left[\frac{\partial \phi}{\partial t_{j'}}\right]_{t=0} = 0$ because PV $\langle \delta_{j'j}\rangle = 0$ and $\lim_{\gamma \to \infty} \gamma \operatorname{Prob}\{|\delta_{j'j}| \geq \gamma\} = 0$. Each $\delta_{j'j}$ must satisfy these two conditions because the underlying ϵ_j and $\epsilon_{j'}$ are required to obey them by assumption (a3) or (a3'). $\square$

Proposition II.17. *Retain the context of Theorem II.12. Let $f : \Re_+ \to \Re$ be a monotonically increasing function, and $g : \Re \to \Re_+$ be its inverse, $g = f^{-1}$. Denote $\kappa = f(\lambda)$ and $\kappa_\mu = f(\lambda_\mu)$ for all μ. Let $\bar{\kappa} = \frac{1}{m_i}\sum_{\mu=1}^{m_i} \kappa_\mu$. If*

(2.22) $$\frac{g''(\xi)}{(g'(\xi))^2} < \Delta\pi\left(\frac{e^{g(\xi)\Delta\pi}-1}{e^{g(\xi)\Delta\pi}+1}\right) \text{ for all } \xi \in [\min\{\kappa_\mu\}, \max\{\kappa_\mu\}],$$

then $\kappa \leq \bar{\kappa}$ with equality if and only if $\kappa_1 = \kappa_2 = \ldots = \kappa_{m_i}$.[7]

Proof. It is straightforward, albeit tedious, to take a second derivative of $\frac{1}{e^{g(\xi)\Delta\pi}+1}$ and obtain inequality (2.22) as the condition implying that this function is concave up (assuming once again $\Delta\pi > 0$ without loss of generality). By the logic used to prove Theorem II.12, $\lambda \leq g(\bar{\kappa})$ with equality if and only if $\kappa_1 = \kappa_2 = \ldots = \kappa_{m_i}$. Because f is monotonically increasing, we can apply it to both sides of this inequality to obtain $\kappa \leq \bar{\kappa}$. $\square$

Completing the Proof of Theorem II.12.

To obtain

$$\lim_{\Delta\pi \to \pm\infty} \lambda = \min\{\lambda_\mu\},$$

we take the limit as $\Delta\pi$ goes to ∞. By symmetry, the result then holds when $\Delta\pi$ goes to $-\infty$ as well. First, we use algebra in equation (2.14) to come up with a new

---

[7]Note that the hypothesis is satisfied when f is an exponential function, i.e., for $\kappa = e^\lambda$, because in this case $g(\xi) = \ln(\xi)$ and g'' is always negative whereas the right-hand side of inequality (2.22) is always positive.

expression for λ:

$$\lambda = \frac{1}{\Delta\pi} \ln \left(\frac{\sum_{\mu=1}^{m_i} e^{\lambda_\mu \Delta\pi} \prod_{\omega \neq \mu} \left(e^{\lambda_\omega \Delta\pi} + 1 \right)}{\sum_{\mu=1}^{m_i} \prod_{\omega \neq \mu} \left(e^{\lambda_\omega \Delta\pi} + 1 \right)} \right).$$

In the limit of $\Delta\pi$ going to ∞,

$$e^{\lambda_\mu \Delta\pi} \prod_{\omega \neq \mu} \left(e^{\lambda_\omega \Delta\pi} + 1 \right) \rightarrow e^{\lambda_\mu \Delta\pi} \prod_{\omega \neq \mu} e^{\lambda_\omega \Delta\pi}$$

$$= \prod_\omega e^{\lambda_\omega \Delta\pi}$$

and

$$\sum_{\mu=1}^{m_i} \prod_{\omega \neq \mu} \left(e^{\lambda_\omega \Delta\pi} + 1 \right) \rightarrow \prod_{\omega \neq \arg\min\{\lambda_\mu\}} e^{\lambda_\omega \Delta\pi}.$$

Thus,

$$\lim_{\Delta\pi \to \infty} \lambda = \lim_{\Delta\pi \to \infty} \frac{1}{\Delta\pi} \ln \left(\frac{\sum_{\mu=1}^{m_i} \prod_\omega e^{\lambda_\omega \Delta\pi}}{\prod_{\omega \neq \arg\min\{\lambda_\mu\}} e^{\lambda_\omega \Delta\pi}} \right)$$

$$= \lim_{\Delta\pi \to \infty} \frac{1}{\Delta\pi} \ln \left(m_i \, e^{\min\{\lambda_\mu\}\Delta\pi} \right)$$

$$= \lim_{\Delta\pi \to \infty} \frac{\min\{\lambda_\mu\}\Delta\pi + \ln(m_i)}{\Delta\pi}$$

$$= \min\{\lambda_\mu\}.$$

To obtain

$$\lim_{\Delta\pi \to 0} \lambda = \bar{\lambda},$$

we apply l'Hospital's Rule to the expression for λ given in equation (2.14). We have

$$\frac{d}{d\Delta\pi} \left[\ln \left(\frac{\sum_{\mu=1}^{m_i} \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi}+1}}{\sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1}} \right) \right] =$$

$$\left(\frac{\sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1}}{\sum_{\mu=1}^{m_i} \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi}+1}} \right) \frac{\left(\sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1} \right) \left(\sum_{\mu=1}^{m_i} \frac{\lambda_\mu e^{\lambda_\mu \Delta\pi}}{\left(e^{\lambda_\mu \Delta\pi}+1\right)^2} \right) - \left(\sum_{\mu=1}^{m_i} \frac{e^{\lambda_\mu \Delta\pi}}{e^{\lambda_\mu \Delta\pi}+1} \right) \left(\sum_{\mu=1}^{m_i} \frac{-\lambda_\mu e^{\lambda_\mu \Delta\pi}}{\left(e^{\lambda_\mu \Delta\pi}+1\right)^2} \right)}{\left(\sum_{\mu=1}^{m_i} \frac{1}{e^{\lambda_\mu \Delta\pi}+1} \right)^2}.$$

So

$$\frac{d}{d\Delta\pi}\left[\ln\left(\frac{\sum_{\mu=1}^{m_i}\frac{e^{\lambda_\mu\Delta\pi}}{e^{\lambda_\mu\Delta\pi}+1}}{\sum_{\mu=1}^{m_i}\frac{1}{e^{\lambda_\mu\Delta\pi}+1}}\right)\right]_{\Delta\pi=0} = (1)\frac{\left(\frac{m_i}{2}\right)\left(\sum_{\mu=1}^{m_i}\frac{\lambda_\mu}{4}\right) - \left(\frac{m_i}{2}\right)\left(\sum_{\mu=1}^{m_i}\frac{-\lambda_\mu}{4}\right)}{\left(\frac{m_i}{2}\right)^2}$$

$$= \frac{\sum_{\mu=1}^{m_i}\lambda_\mu}{m_i}$$

$$= \bar{\lambda}.$$

The denominator in (2.14) is $\Delta\pi$, so its derivative is 1. Thus,

$$\lim_{\Delta\pi\to 0}\lambda = \bar{\lambda}. \quad \square$$

Proof of Theorem II.13.

Without loss of generality, assume $\Delta\pi > 0$.

To obtain

$$\min\{\lambda_\mu\} \geq \frac{1}{\Delta\pi}\ln\left(\frac{1}{m_i}\left(e^{\lambda\Delta\pi} - (m_i - 1)\right)\right),$$

we make use of the following inequalities: $\frac{e^{\lambda_\mu\Delta\pi}}{e^{\lambda_\mu\Delta\pi}+1} \leq 1$ and $\frac{1}{e^{\lambda_\mu\Delta\pi}+1} \geq 0$ for all $\mu \neq \arg\min\{\lambda_\omega\}$. Applying these inequalities to our formula for λ in equation (2.14), we get

$$\lambda \leq \frac{1}{\Delta\pi}\ln\left(\frac{\frac{e^{\min\{\lambda_\mu\}\Delta\pi}}{e^{\min\{\lambda_\mu\}\Delta\pi}+1} + m_i - 1}{\frac{1}{e^{\min\{\lambda_\mu\}\Delta\pi}+1}}\right)$$

$$= \frac{1}{\Delta\pi}\ln\left(e^{\min\{\lambda_\mu\}\Delta\pi} + (m_i - 1)(e^{\min\{\lambda_\mu\}\Delta\pi} + 1)\right)$$

$$= \frac{1}{\Delta\pi}\ln\left(m_i\, e^{\min\{\lambda_\mu\}\Delta\pi} + m_i - 1\right).$$

So

$$e^{\lambda\Delta\pi} \leq m_i\, e^{\min\{\lambda_\mu\}\Delta\pi} + m_i - 1,$$

and thus,

$$\frac{1}{\Delta\pi}\ln\left(\frac{1}{m_i}\left(e^{\lambda\Delta\pi} - (m_i - 1)\right)\right) \leq \min\{\lambda_\mu\}.$$

Note that this bound is meaningful only if $e^{\lambda \Delta \pi} > 2m_i - 1$.

To obtain

$$\max\{\lambda_\mu\} \leq \frac{1}{\Delta \pi} \ln \left(\frac{(m_i + 1)e^{\lambda \Delta \pi} - (m_i - 1)}{m_i + 1 - (m_i - 1)e^{\lambda \Delta \pi}} \right),$$

we follow a similar approach using the fact that $\frac{e^{\lambda_\mu \Delta \pi}}{e^{\lambda_\mu \Delta \pi}+1} \geq \frac{1}{2}$ and $\frac{1}{e^{\lambda_\mu \Delta \pi}+1} \leq \frac{1}{2}$ for all $\mu \neq \arg\max\{\lambda_\omega\}$. Putting these inequalities into equation (2.14) produces

$$
\begin{aligned}
\lambda &\geq \frac{1}{\Delta \pi} \ln \left(\frac{\frac{e^{\max\{\lambda_\mu\}\Delta \pi}}{e^{\max\{\lambda_\mu\}\Delta \pi}+1} + (m_i - 1)\frac{1}{2}}{\frac{1}{e^{\max\{\lambda_\mu\}\Delta \pi}+1} + (m_i - 1)\frac{1}{2}} \right) \\
&= \frac{1}{\Delta \pi} \ln \left(\frac{2e^{\max\{\lambda_\mu\}\Delta \pi} + (m_i - 1)(e^{\max\{\lambda_\mu\}\Delta \pi} + 1)}{2 + (m_i - 1)(e^{\max\{\lambda_\mu\}\Delta \pi} + 1)} \right) \\
&= \frac{1}{\Delta \pi} \ln \left(\frac{(m_i + 1)e^{\max\{\lambda_\mu\}\Delta \pi} + m_i - 1}{(m_i - 1)e^{\max\{\lambda_\mu\}\Delta \pi} + m_i + 1} \right).
\end{aligned}
$$

So

$$e^{\lambda \Delta \pi} \geq \frac{(m_i + 1)e^{\max\{\lambda_\mu\}\Delta \pi} + m_i - 1}{(m_i - 1)e^{\max\{\lambda_\mu\}\Delta \pi} + m_i + 1},$$

and thus,

$$(m_i + 1)e^{\lambda \Delta \pi} - (m_i - 1) \geq (m_i + 1)e^{\max\{\lambda_\mu\}\Delta \pi} - (m_i - 1)e^{\lambda \Delta \pi} e^{\max\{\lambda_\mu\}\Delta \pi},$$

and finally,

$$\frac{1}{\Delta \pi} \ln \left(\frac{(m_i + 1)e^{\lambda \Delta \pi} - (m_i - 1)}{m_i + 1 - (m_i - 1)e^{\lambda \Delta \pi}} \right) \geq \max\{\lambda_\mu\}.$$

This bound is meaningful only if $e^{\lambda \Delta \pi} < \frac{m_i+1}{m_i-1}$. $\qquad \square$

Proof of Theorem II.14.

Choose $m_i \in \mathbb{N}$ such that $m_i > \frac{e^{\lambda^* \Delta \pi^*}+1}{2}$ and $m_i > \frac{e^{\lambda^* \Delta \pi^*}+1}{e^{\lambda^* \Delta \pi^*}-1}$. This ensures that neither of the bounds in Theorem II.13 apply. Thus, we can take $\lambda'_1 = 0$ and $\lambda'_{m_i} > m_i \lambda'$ and still be able to choose the remaining $\{\lambda'_\mu\}$, for $\mu = 2 \ldots m_i - 1$, such that equation (2.20) holds. That means these heterogeneous logit parameters will be consistent with the homogeneous logit model with rationality parameter λ^* and

equilibrium payoff difference $\Delta \pi^*$. We have specifically chosen λ'_1 and λ'_{m_i} so that $\lambda'_1 < \lambda' < \overline{\lambda'}$. Thus, noting the limits we take in Theorem II.12, we establish that $\lambda\left(\{\lambda'_\mu\}, \Delta\pi\right)$ is above λ' when $\Delta\pi \approx 0$ and is below λ' when $\Delta\pi$ is large. Because $\lambda\left(\{\lambda'_\mu\}, \Delta\pi\right)$ is continuous in $\Delta\pi$, there is some $\Delta\pi'$ for which $\lambda\left(\{\lambda'_\mu\}, \Delta\pi'\right) = \lambda'$. $\square$

BIBLIOGRAPHY

[1] Anderson, S., Goeree, J., and Holt, C. (2002) The Logit Equilibrium: A Perspective on Intuitive Behavioral Anomalies. *Southern Economic Journal* 69(1), 21-47.

[2] Anderson, S., Goeree, J., and Holt, C. (2004) Noisy Directional Learning and the Logit Equilibrium. *Scandinavian Journal of Economics* 106(3), 581-602.

[3] Camerer, C. (2003) *Behavioral Game Theory*. Princeton University Press, Princeton, NJ.

[4] Camerer, C., Palfrey, T., and Rogers, B. (2008) Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies. *Journal of Economic Theory* forthcoming.

[5] Capra, M., Goeree, J., Gomez, R., and Holt, C. (1999) Anomalous Behavior in a Traveler's Dilemma? *American Economic Review* 89(3), 678-690.

[6] Fey, M., McKelvey, R., and Palfrey, T. (1996) An Experimental Study of Constant-Sum Centipede Games. *International Journal of Game Theory* 25(3), 269-287.

[7] Fudenberg, D., and Levine, D. (1998) *The Theory of Learning in Games*. MIT Press, Cambridge, MA.

[8] Goeree, J., and Holt, C. (1999) Stochastic Game Theory: For Playing Games, Not Just for Doing Theory. *Proceedings of the National Academy of Sciences* 96(19), 10564-10567.

[9] Goeree, J., and Holt, C. (2004) A Model of Noisy Introspection. *Games and Economic Behavior* 46(2), 365-382.

[10] Goeree, J., Holt, C., and Palfrey, T. (2002) Quantal Response Equilibrium and Overbidding in Private-Value Auctions. *Journal of Economic Theory* 104(1), 247-272.

[11] Goeree, J., Holt, C., and Palfrey, T. (2005) Regular Quantal Response Equilibrium. *Experimental Economics* 8, 347-367.

[12] Golman, R. (2008) Why Learning Doesn't Add Up: Equilibrium Selection and Compositions of Learning Rules. Working Paper, University of Michigan.

[13] Guyer, M., and Rapoport, A. (1972) 2x2 Games Played Once. *The Journal of Conflict Resolution* 16(3), 409-431.

[14] Haile, P., Hortacsu, A., and Kosenok, G. (2008) On the Empirical Content of Quantal Response Equilibrium. *American Economic Review* 98(1), 180-200.

[15] Ho, T., Wang, X., and Camerer, C. (2008) Individual Differences in EWA Learning with Partial Payoff Information. *The Economic Journal* 118, 37-59.

[16] Hofbauer, J., and Hopkins, E. (2005) Learning in Perturbed Asymmetric Games. *Games and Economic Behavior* 52(1), 133-152.

[17] Hofbauer, J., and Sandholm, W.H. (2002) On the Global Convergence of Stochastic Fictitious Play. *Econometrica* 70, 2265-2294.

[18] Hofbauer, J., and Sandholm, W.H. (2007) Evolution in Games with Randomly Disturbed Payoffs. *Journal of Economic Theory* 132, 47-69.

[19] Hommes, C. (2006) Heterogeneous Agent Models in Economics and Finance. In: Judd, K., Tesfatsion, L. (eds) Handbook of Computational Economics, vol 2, Agent-Based Computational Economics. North Holland, Amsterdam.

[20] Hopkins, E. (1999) A Note on Best Response Dynamics. *Games and Economic Behavior* 29, 138-150.

[21] Kirman, A. (2006) Heterogeneity in Economics. *Journal of Economic Interaction and Coordination* 1, 89-117.

[22] McKelvey, R., and Palfrey, T. (1995) Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior* 10(1), 6-38.

[23] McKelvey, R., Palfrey, T., and Weber, R. (2000) The effects of payoff magnitude and heterogeneity on behavior in 2x2 games with unique mixed strategy equilibria. *Journal of Economic Behavior & Organization* 42, 523-548.

[24] Pitman, E. J. G. (1956) On the Derivatives of a Characteristic Function at the Origin. *The Annals of Mathematical Statistics* 27(4), 1156-1160.

[25] Tsakas, E. (2005) Mixed Quantal Response Equilibria for Normal Form Games. Paper presented at the International Conference on Game Theory.

[26] Weibull, J. (1994) The Mass-Action interpretation of Nash equilibrium. Paper presented at Nobel symposium in honor of John Nash.

[27] Wilcox, N. (2006) Theories of Learning in Games and Heterogeneity Bias. *Econometrica* 74, 1271-1292.

[28] Wolfe, S. (1973) On the Local Behavior of Characteristic Functions. *The Annals of Probability* 1(5), 862-866.

CHAPTER III

Basins of Attraction and Equilibrium Selection Under Different Learning Rules

3.1 Background

The existence of an equilibrium in a game is insufficient proof of its plausibility as an outcome. We must also describe a process through which players can achieve it. The distinction between the existence of an equilibrium and its attainability, and the necessity of the latter, rests at the foundations of game theory. In Nash's 1951 thesis, he proposed an adjustment dynamic built on a mass action model to support the convergence to an equilibrium (Weibull, 1996). The Nash adjustment dynamic relies on self interested behavior to move a population of players toward equilibrium. Unfortunately, it fails to achieve equilibria for many games. For this reason, game theorists building on Nash's original work focused instead on fictitious play, a learning rule in which players successively choose a pure strategy which is optimal against the cumulated history of the opponent's plays (Brown, 1951). More recent research by economists, psychologists, and theoretical biologists has produced a variety of adjustment dynamics, many of which fall into two broad categories: *belief based learning* and *reinforcement based learning*.[1] In the former, players take actions based on their beliefs of the actions of others. In the latter, players mimic actions

[1]These categories also go by the terms *epistemic learning* and *behavioral learning* respectively (Walliser, 1998).

44

that have been successful in the past (see Fudenberg and Levine, 1998; Camerer, 2003; Swinkels, 1993).

In this chapter, we focus on two learning dynamics /adjustment processes: *continuous time best response dynamics* (Gilboa and Matsui, 1991) and *replicator dynamics* (Taylor and Jonker, 1978) and explore the extent to which they can differ in their basins of attraction for symmetric games with strict equilibria. For any two-by-two symmetric game, these two learning rules produce identical basins of attraction. We show that by adding a single action, we can produce a game in which these two learning rules create basins of attraction that have arbitrarily small overlap. In other words, best response dynamics lead to a different equilibrium than replicator dynamics almost always. Within our class of three-by-three games, the equilibrium found by the replicator dynamics is a *uniformly evolutionarily stable strategy*, but it is almost never the initial best response. The condition that pure, uniformly evolutionarily stable strategies satisfy this *never an initial best response property* proves to be a necessary requirement for the two learning rules to share vanishing overlap in their basins of attraction for strict equilibria. These results extend to classes of dynamics that generalize the best response and replicator rules.

To show how these rules can produce such different outcomes, we must first describe how best response and replicator dynamics model a population of adapting agents in the aggregate. In general, we assume players are randomly matched from large population pools. Best response dynamics are a form of belief-based learning – players' action choices depend on their beliefs about the actions of other players. In continuous time best response dynamics, a population of players moves toward a best response to the current state of the opposing population. Fictitious play relies on belief-based learning in discrete time. In each period, the rule assigns new beliefs

based on the average play of the opponent. Actions are chosen rationally given those beliefs. The best response dynamics can be thought of as the extension of fictitious play to continuous time (Hofbauer and Sigmund, 2003).[2]

In contrast to best response dynamics, replicator dynamics are a form of reinforcement learning – actions spread based on their past success (Erev and Roth, 1998).[3] Replicator dynamics have ecological foundations: payoffs are analogous to fitness, and fitter actions are more apt to survive and grow. Note that actions initially not present in the population can never be tried with replicator dynamics.

In this chapter, we consider symmetric matrix games. The learning dynamics thus operate in a single, large, well-mixed population. In this setting, the continuous time best response dynamics and replicator dynamics can be derived as the expected behavior of agents with stochastic protocols for switching their actions (Sandholm, 2009). In the best response dynamics, some infinitesimal proportion of the agents are always switching their action to match the current best response. The resulting flows are piecewise linear. In the replicator dynamics, agents copy better performing members of the population (Schlag, 1998). Players do not rely on beliefs about the actions of others. They need only know the payoffs of actions they encounter. Learning by imitation at the agent level thus leads to reinforcement learning at the population level.

Belief-based learning rules, such as best response, and reinforcement learning rules, such as replicator dynamics can be combined in a single learning rule called experience-weighted attraction (EWA) learning (Camerer and Ho, 1999). EWA can be made to fit either model exactly or to create a hybrid model that balances beliefs

---

[2]The connection between fictitious play and best response dynamics requires the view that in fictitious play, a new agent enters the population each round with an action that is fixed forever. The state variable must then take on an interpretation as the opponent's population mixed strategy.

[3]The aforementioned Nash learning rule, or what is now called the Brown - von Neumann - Nash (BNN) dynamics also can be interpreted as a form of reinforcement learning (Brown and von Neumann, 1950; Skyrms, 1990).

about future plays against past history of success. In experimental tests across a variety of games, belief-based learning, reinforcement learning, and EWA learning all predict behavior with reasonable accuracy. EWA outperforms the two pure models, though this is partly due to the fact that it has more free parameters.

The extant theoretical and empirical literature suggests that often these distinct learning rules make similar predictions about rates of change of actions and that for many games, they select identical equilibria. We know, for example, that any strict pure Nash Equilibrium will be dynamically stable under nearly all learning dynamics and that interior evolutionarily stable strategies are globally stable for both replicator dynamics (Hofbauer et al., 1979) and best response dynamics (Hofbauer, 1995; Hofbauer, 2000). Hopkins (1999) shows that stability properties of equilibria are robust across many learning dynamics, and, most relevant for our purposes, that best response dynamics and replicator dynamics usually have the same asymptotic properties. Best response dynamics and replicator dynamics are both myopic adjustment dynamics – they both flow towards higher immediate payoffs (Swinkels, 1993). Feltovich (2000) finds that belief-based learning and reinforcement learning generate qualitatively similar patterns of behavior, as does Salmon (2001), whose analytic survey concludes that only subtle differences exist across the various learning rules in extant experiments. Thus, advocates of each learning rule can point to substantial empirical support.

Our finding, that the choice of learning rule has an enormous effect on the choice of equilibrium, points to the importance of determining how people actually learn. And while the experimental work just mentioned has found this a difficult prospect, our class of games offers an opportunity to distinguish between different types of learning. Experiments on our games would have to find one, or possibly both, of the

learning rules to be inconsistent with observed behavior.

Our results may at first seem to contradict the existing current literature. We want to make clear that they do not. First, many experiments consider two-by-two games. And as we review here, the two learning rules generate identical basins of attraction for two-by-two symmetric matrix games. The learning rules differ only in the time that they take to reach those equilibria. Second, our analysis focuses on *basins of attraction*, i.e. we ask which equilibrium is reached given an initial point. Most of the existing theorems consider *stability*, i.e. whether an equilibrium is stable to perturbations. Proving that an equilibrium is stable does not imply anything about the size of its basin of attraction. An equilibrium with a basin of attraction of measure epsilon can be stable. Thus, results that strict equilibria are stable for both replicator dynamics and best response dynamics do not imply that the two dynamics generate similar basins of attraction.

Conditions on payoff matrices that imply that best response dynamics, replicator dynamics, and Nash dynamics all produce similar stability properties need not place much restriction on basins of attraction, unless the stability is global. Conditions for global stability of each dynamic, for example if the mean payoff function is strictly concave (Hofbauer and Sigmund, 2003), imply identical basins of attraction. However, such conditions also imply a unique stable equilibrium.[4] One branch of the learning literature does consider games in which stability depends on the learning dynamic (Kojima, 2006) as well as games with distinct basins of attraction for different learning rules (Hauert et al., 2004). Those models rely on nonlinear payoff structures. Here, we consider matrix games with linear payoffs.

Of course, in a symmetric rock-paper-scissors game or an asymmetric matching

---

[4]Similar logic applies to repelling equilibria: if the mean payoff function is strictly convex, then a possible interior Nash Equilibrium must be repelling for each dynamic. Hofbauer and Sigmund's theorem (2003) follows from earlier work with each dynamic (Hofbauer and Sigmund, 1998; Hofbauer, 2000; Hopkins, 1999).

pennies game, best response dynamics converges to the mixed equilibrium while replicator dynamics cycles. In these games, the mixed equilibrium is attracting under best response dynamics, but is only neutrally stable under replicator dynamics. Rock-paper-scissors is a knife edge case, where a slight change in payoffs could make the equilibrium stable under replicator dynamics, but matching pennies illustrates the inability of replicator dynamics to attain a mixed equilibrium in any asymmetric game. Our focus here is different. We analyze symmetric games with strict equilibria. The equilibria are asymptotically stable under both dynamics. We identify divergent behavior of the learning rules, not because one rule fails to attain an equilibrium, but because the two rules select different equilibria.

To prove our results, we consider each possible initial distribution over actions and then characterize how the various learning rules specify the path of future distributions. In the games we consider, these continuous flows attain equilibria. Thus, the equilibrium selected can be thought of as a function of the initial population distribution of actions and the learning rule.

Our result that the choice of learning rule can determine the equilibrium selected can be interpreted through the lens of the equilibrium refinement literature (Harsanyi and Selten, 1988; Govindan and Wilson, 2005; Samuelson, 1997; Basov, 2004). In games with multiple strict Nash Equilibria, dynamical models with persistent randomness select long run, stochastically stable equilibria, which generalize the notion of risk dominance from two-by-two games (Foster and Young, 1990; Kandori et al., 1993). The stochastically stable equilibrium in a 3-by-3 game can vary with the learning dynamic (Ellison, 2000). Long run stochastic stability depends on the relative sizes of basins of attraction, given the underlying deterministic dynamic. Thus, even though we deal with deterministic dynamics only, our result complements the

literature on stochastic stability by further supporting the conclusion that long run equilibria can be sensitive to how players learn. Our findings establish that the importance of learning style in equilibrium selection does not strictly rely on the presence of shocks that shift the population from one equilibrium to another.

The remainder of this chapter is organized as follows. In the next section, we define the learning rules and show how they generate similar behavior in a simple three-by-three coordination game. Then, we present our main results, which show that belief-based learning and reinforcement learning can be very different. In Section 3.4, we introduce generalizations of best response and replicator dynamics and extend our results to these classes of dynamics. We conclude with a discussion of the relevance of the attainability of equilibria.

3.2 The Learning Rules

In a population game, the state space for a given population X is the unit simplex $\triangle$. A point $\mathbf{x} \in \triangle$ denotes the fraction of the population playing each action and is thus called a population mixed strategy. A learning rule for population X operates on the state space $\triangle$ by specifying for any given payoff structure a dynamical system $\dot{\mathbf{x}} = \mathbf{V}_\pi(\mathbf{x}, t)$ such that $\triangle$ is forward invariant, i.e., trajectories stay within the simplex. We interpret the learning dynamic as tracking the changes in the proportions of agents choosing the various actions.

We first introduce our learning rules in the context of a two-player game with large populations X and Y of randomly matched agents with n and m actions respectively. Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_m)$ be the population mixed strategy vectors. The component x_i (or y_i) is the fraction of population X (or Y) choosing action i. We will refer to the fraction of population X (or Y) choosing an action other than i

as x_{-i} (or y_{-i}). Denote by π_i^μ the payoff a player in population μ gets from action i. Of course, payoffs are a function of the opposing population mixed strategy, but we omit the function's argument for ease of notation, writing π_i^X in place of $\pi_i^X(\mathbf{y})$. Denote the vector of these payoffs by $\vec{\pi}^\mu = (\pi_1^\mu, \ldots, \pi_n^\mu)$.

The continuous time replicator dynamics can be written as

$$\dot{x}_i = x_i(\pi_i^X - \bar{\pi}^X)$$

$$\dot{y}_i = y_i(\pi_i^Y - \bar{\pi}^Y)$$

where $\bar{\pi}^\mu$ is the average payoff in population μ. Specifically, $\bar{\pi}^X = \mathbf{x} \cdot \vec{\pi}^X$ and $\bar{\pi}^Y = \mathbf{x} \cdot \vec{\pi}^Y$.

Let $\mathrm{BR}(\mathbf{y})$ be the set of best replies to $\mathbf{y}$ (for a player in population X),

$$\mathrm{BR}(\mathbf{y}) = \arg\max_{\mathbf{v} \in \triangle^{n-1}} \mathbf{v} \cdot \vec{\pi}^X.$$

Similarly, the set of best replies to $\mathbf{x}$ is:

$$\mathrm{BR}(\mathbf{x}) = \arg\max_{\mathbf{v} \in \triangle^{m-1}} \mathbf{v} \cdot \vec{\pi}^Y.$$

Continuous time best response dynamics can be written as

$$\dot{\mathbf{x}} \in \mathrm{BR}(\mathbf{y}) - \mathbf{x} \qquad \dot{\mathbf{y}} \in \mathrm{BR}(\mathbf{x}) - \mathbf{y}.$$

The discrete fictitious play learning rule can be written as

$$\mathbf{x}(t+1) = \frac{t\mathbf{x}(t) + \mathbf{b}(t)}{t+1}$$

where $\mathbf{x}(t)$ is the vector of frequencies each action has been played through period t and $\mathbf{b}(t)$ is a best response to the opponent's history at this point. Fictitious play closely approximates continuous time best response dynamics. To avoid repetition, we focus on the best response dynamics. Results for best response hold for fictitious play as well.

This chapter focuses on symmetric matrix games. In these games, both players have the same set of available actions and payoffs are linear. We can define the learning rules in the context of a single, well-mixed population, suitable for a symmetric game. The replicator dynamics are

$$\dot{x}_i = x_i(\pi_i - \bar{\pi}).$$

The superscripts can be dropped because in the single population setting there is no ambiguity in referring to the payoff to the average payoff $\bar{\pi}$ or the payoff to action i, π_i. The best response dynamics are

$$\dot{\mathbf{x}} \in \mathrm{BR}(\mathbf{x}) - \mathbf{x}.$$

3.2.1 An Example

To show how to apply these learning rules, we begin with an example of a simple three-by-three coordination game. In this game, the various learning rules generate similar basins of attraction. We borrow this game from Haruvy and Stahl (1999; 2000) who used it to study learning dynamics and equilibrium selection in experiments with human subjects. The payoff matrix for the Haruvy-Stahl game is written as follows:

$$\begin{pmatrix} 60 & 60 & 30 \\ 30 & 70 & 20 \\ 70 & 25 & 35 \end{pmatrix}.$$

The entry in row i and column j gives the payoff to a player who chooses action i and whose opponent chooses action j. This game has two strict pure Nash Equilibria: $(0, 1, 0)$ and $(0, 0, 1)$ as well as a mixed Nash Equilibrium at $(0, \frac{1}{4}, \frac{3}{4})$. It can be shown
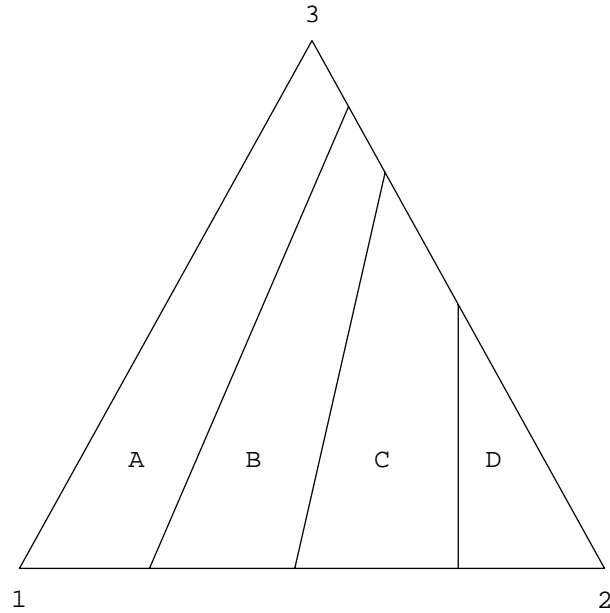
Figure 3.1: Best response regions. In region A, action 3 is the best response. In regions B and C, action 1 is the best response, but in B $\pi_3 > \pi_2$, while in C the opposite is true. In region D, action 2 is the best response.

for both best response dynamics and replicator dynamics that the two pure equilibria are stable and that the mixed equilibrium is unstable.

Given that this game has three possible actions, we can write any distribution of actions in the two dimensional simplex $\triangle^2$. To locate the basins of attraction of each equilibrium, we must first identify those regions of the simplex $\triangle^2$ in which each action is a best response. This is accomplished by finding the lines where each pair of actions performs equally well. Let π_i be the payoff from action i. We find $\pi_1 = \pi_2$ when $4x_2 + 2x_3 = 3$, $\pi_2 = \pi_3$ when $17x_2 + 5x_3 = 8$, and $\pi_1 = \pi_3$ when $9x_2 + x_3 = 2$. These three lines determine the best response regions shown in Figure 3.1.

We can use Figure 3.1 to describe the equilibrium chosen under best response dynamics. Regions A, B, and C all lie the basin of attraction of action 3, while region D is in the basin of action 2. Note that the boundary of the basins of attraction under best response dynamics is a straight line.
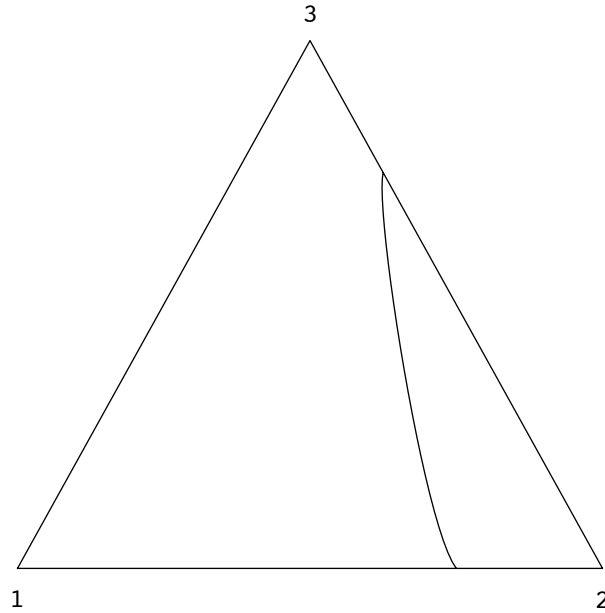
Figure 3.2: Basins of attraction under replicator dynamics.

In Figure 3.2, we characterize the basins of attraction for replicator dynamics. The boundary separating the basins of attraction here becomes a curve from the point $(\frac{1}{4}, \frac{3}{4}, 0)$ to $(0, \frac{1}{4}, \frac{3}{4})$ entirely within region C of Figure 3.1. Notice that the basins of attraction under best response dynamics and replicator dynamics differ. Best response dynamics creates basins with straight edges. Replicator dynamics creates basins with curved edges. This curvature arises because the second best action can also grow in the population under replicator dynamics. As it grows in proportion, it can become the best response. As a result, the population can slip from one best response basin into another one. Even so, notice that the difference in the two basins of attraction comprises a small sliver of the action space. We show this in Figure 3.3.

In games such as this, the two dynamics not only select the same equilibrium almost all of the time, but also generate qualitatively similar behavior. If the initial distribution of actions is close to $(0, 1, 0)$, the dynamics flow to that equilibrium point. If not, they flow to $(0, 0, 1)$.
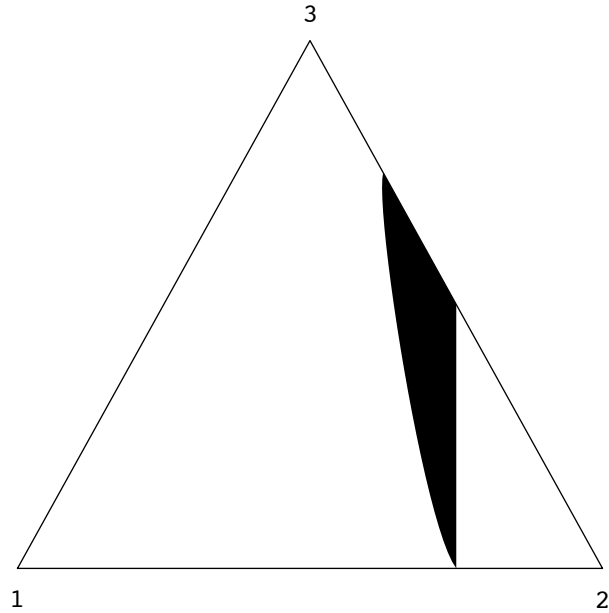
Figure 3.3: The small difference between best response and replicator dynamics. The shaded area flows to action 2 under replicator dynamics, to action 3 with best response dynamics.

In this game, the two learning rules create similar basins of attraction. Intuitively, we might expect only these small differences for all games with three actions, given the similarities of the learning rules. However, as we show in the next section, even with three-by-three games, the sliver can become almost the entire simplex.

3.3 Results

We now turn to our main results. We first present the well known fact that best response dynamics and replicator dynamics are identical for games with two possible actions. We consider learning dynamics to be identical if the direction of their flows is the same. This allows for differences in the speed of the flow. We then define a class of games with three actions in which the two learning rules generate basins of attraction with vanishing overlap. Within that class of games, an equilibrium action is almost never the initial best response. We show that to be a necessary condition for any symmetric game for which the two learning rules almost always

lead to different strict equilibria.

Proposition III.1. *For symmetric two-by-two matrix games, best response dynamics and replicator dynamics produce identical dynamics (Fudenberg and Levine, 1998).*

Proof. The best response dynamics reduces to

$$\dot{x}_i = x_j$$

$$\dot{x}_j = -x_j$$

when $\pi_i > \pi_j$, and to $\dot{\mathbf{x}} = 0$ when they payoffs are equal. The replicator dynamics reduces to

$$\dot{x}_1 = (\pi_1 - \pi_2)x_1 x_2$$

$$\dot{x}_2 = (\pi_2 - \pi_1)x_1 x_2.$$

In both dynamics, the action with the higher payoff increases until the two payoffs become equal or the other action is completely eliminated. $\square$

Our first theorem says that there are three-by-three matrix games such that the two learning dynamics lead to different outcomes, for nearly all initial conditions. The claim cannot hold for all initial conditions because of the case where the initial point is a Nash Equilibrium of the game.

Theorem III.2. *For any ϵ, there is a three-by-three game such that the fraction of the space of initial conditions from which best response dynamics and replicator dynamics lead to the same outcome is less than ϵ.*

We present a proof by construction. Consider the class of games with payoff matrix

$$(3.1) \qquad \begin{pmatrix} 1 & -N & -N^{-1} \\ 2 - N^3 & 2 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Lemma III.3. *For any $N > 1$, both best response dynamics and replicator dynamics have two stable fixed points at:* $\mathbf{x} = (1,0,0)$ *and* $\mathbf{x} = (0,1,0)$.

Proof. Both configurations are strict Nash Equilibria because both actions are strict best responses to themselves. Thus, a population in which all players take action 1 (resp. 2) would remain fixed. Strict Nash Equilibria are necessarily stable fixed points of both best response and replicator dynamics. The game also has an interior Nash Equilibrium which is unstable under either learning rule. These stable fixed points have to be Nash Equilibria, and no other Nash Equilibria exist.

Note that $(0,0,1)$ is not a Nash Equilibrium because action 2 is a best response. While it is a fixed point with respect to replicator dynamics, it cannot be stable. $\square$

Given two stable rest points, the eventual choice of one or the other depends on the initial distribution of play. The next result shows that for large N, best response dynamics almost always converge to all players taking action 2. The accompanying Figure 3.4 shows the flow diagram for the best response dynamics when $N = 5$.

Lemma III.4. *For any ϵ, there exists M such that for all $N \geq M$, the basin of attraction of $(0,1,0)$ given best response dynamics is at least $1 - \epsilon$ of the action space.*

Proof. First we show any point with $x_1 > \frac{1}{N}$ and $x_2 > \frac{1}{N}$ is in the basin of attraction of $(0,1,0)$, assuming $N > 2$. For such a point, action 3 is initially a best response because $\pi_3 = 0$ whereas $\pi_1 = x_1 - Nx_2 - \frac{1}{N}x_3 < 0$ and $\pi_2 = 2 - N^3 x_1 < 0$. Then, as

we show, action 1 never becomes a best response. So, eventually, the dynamic flows toward action 2.

Because actions which are not best responses have the same relative decay rate,

$$\frac{x_1(t)}{x_1(0)} = \frac{x_2(t)}{x_2(0)}$$

for t such that action 3 is still a best response. So $x_1(t) - Nx_2(t) < 0$ for all t because it holds for $t = 0$. Action 3 dominates action 1. Action 3 is not a Nash Equilibrium, so eventually another action must become the best response, and the only candidate is action 2. Once x_1 falls to $\frac{2}{N^3}$, action 2 dominates forever.

Thus, by choosing N large enough, the basin of attraction of $(0, 1, 0)$ can be made as large as desired. $\qquad\square$

The next lemma shows that for large N, replicator dynamics leads to all players taking action 1 for almost any initial condition. Figure 3.5 shows the replicator dynamics flow pattern when $N = 5$.

Lemma III.5. *For any ϵ, there exists M such that for all $N \geq M$, the basin of attraction of $(1, 0, 0)$ given replicator dynamics is at least $1 - \epsilon$ of the action space.*

Proof.

$$\dot{x}_1 = x_1\left((x_1 - Nx_2 - \frac{1}{N}x_3)(1 - x_1) - 2x_2 + N^3x_1x_2\right).$$

If $x_1 > \frac{1}{N}$, then $x_1 - \frac{1}{N}x_3 > 0$. For $N > 2$, $x_1 > \frac{1}{N}$ also implies $-Nx_2(1 - x_1) - 2x_2 + N^3x_1x_2 > 0$ because $N^3x_1 > N^2 > N(1 - x_1) + 2$.

So, for $N > 2$, if $x_1 > \frac{1}{N}$, then $\dot{x}_1 > 0$. This means the replicator dynamics will flow to action 1.

By choosing N large enough, the basin of attraction of $(1, 0, 0)$ can be made as large as desired. $\qquad\square$
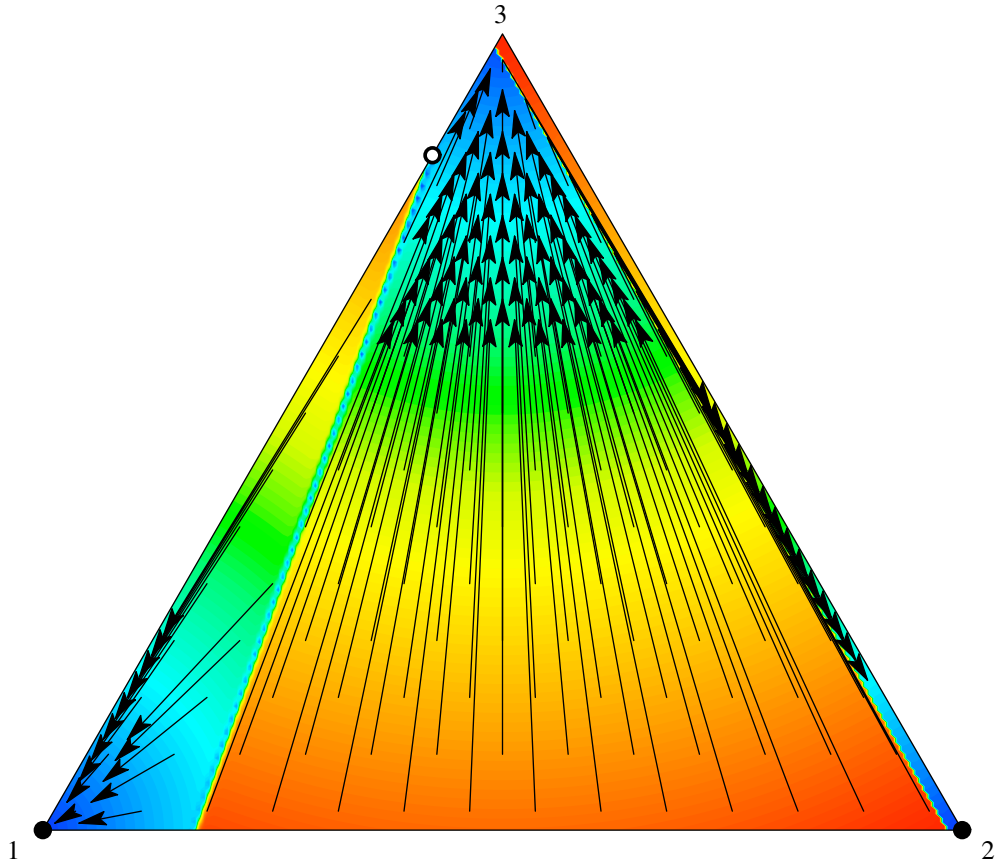
Figure 3.4: Phase diagram for the best response dynamics in the game used to prove Theorem III.2, setting $N = 5$. Black (white) circles are stable (unstable) rest points. Most trajectories initially move towards action 3, but from this corner then flow to action 2. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).
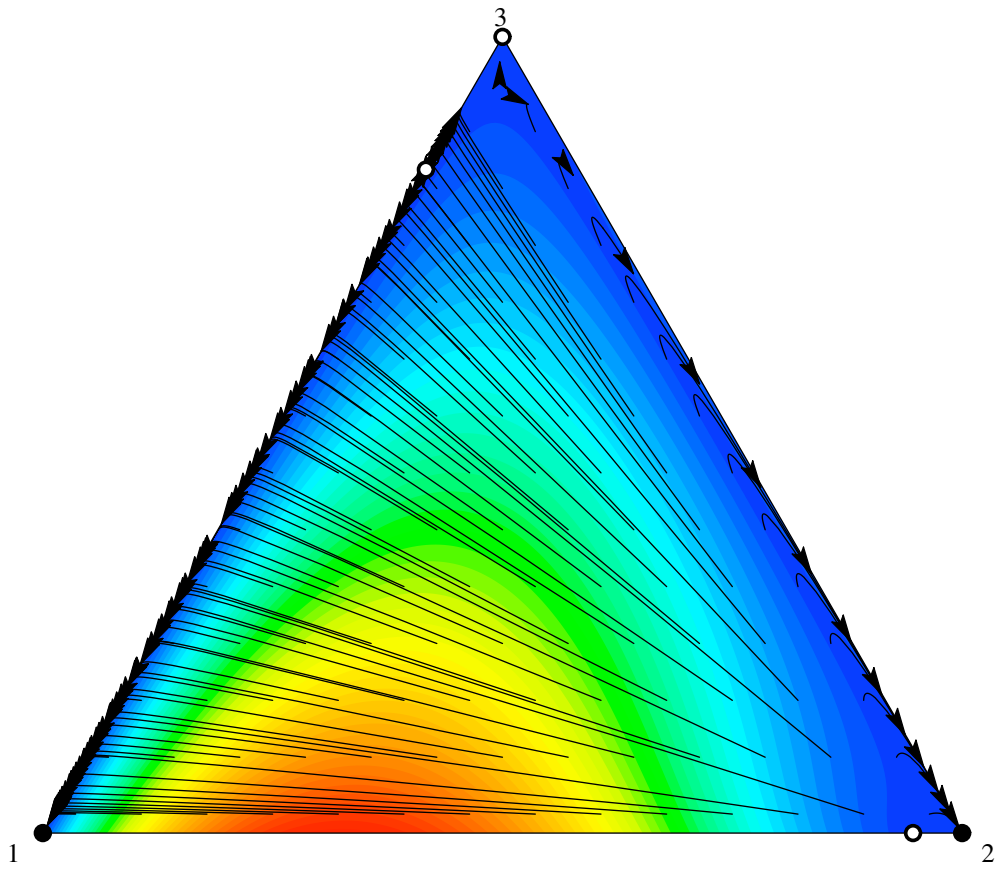
Figure 3.5: Phase diagram for the replicator dynamics in the game used to prove Theorem III.2, setting $N = 5$. Most trajectories flow away from action 2 and then towards action 1. Figure made by the Dynamo program (Sandholm and Dokumaci, 2007).

Thus, we have proved Proposition III.6, that as N approaches infinity, best response dynamics and replicator dynamics converge to different equilibria.

Proposition III.6. *In the limit as $N \to \infty$, the Lebesgue measure of the set of initial starting points for which best response dynamics and replicator dynamics flow to the same equilibrium tends to zero.*

This completes the proof of Theorem III.2 above. Notice that in the class of games used in the proof, neither of the equilibrium strategies is an initial best response almost anywhere in the action space when N grows large. We say that these strategies satisfy the *Never an Initial Best Response Property* for such a sequence of games. To formally define this property, we must introduce some notation.

Let m be the Lebesgue measure on the action space. Given a vector of parameter values $\vec{P}$, let $G(\vec{P})$ be a class of games with payoffs that depend on those parameters. Let $\text{BR}^{-1}(\mathbf{s})$ be the set of points $\mathbf{x}$ for which strategy $\mathbf{s}$ is a best response.

Definition III.7. Strategy $\mathbf{s}$ satisfies the *Never an Initial Best Response Property* at $\vec{\hat{P}}$ if

$$\lim_{\vec{P} \to \vec{\hat{P}}} m\left(\text{BR}^{-1}(\mathbf{s})\right) = 0.$$

Our next result makes use of the Never an Initial Best Response Property in establishing a necessary condition for there to be vanishing overlap in the basins of attraction created by best response dynamics and replicator dynamics. Before presenting this result, we need to lay down some more groundwork.

Recall that a strict equilibrium of a game is one in which each player's strategy is a strict best response to that equilibrium. We now extend the definition of a strict equilibrium to the limit of a sequence of games. Note that only pure Nash Equilibria can be strict.

Definition III.8. An equilibrium $\mathbf{s}$ is strict in the limit as $\vec{P} \to \vec{\hat{P}}$ if for all i such that $s_i > 0$,

(3.2) $\qquad \lim\limits_{\vec{P} \to \vec{\hat{P}}} f(\vec{P}) \left(\pi_i(\mathbf{s}) - \pi_j(\mathbf{s}) \right) > 0$ for all $j \neq i$ and some $f(\vec{P}) > 0$.

Condition (3.2) is equivalent to the following condition: for all $\vec{P} \neq \vec{\hat{P}}$ in some neighborhood of $\vec{\hat{P}}$,

$$\left(\pi_i(\mathbf{s}) - \pi_j(\mathbf{s}) \right) > 0 \text{ for all } j \neq i.$$

Strict equilibrium actions are also evolutionarily stable strategies (ESS), as we can see from Maynard Smith's original (1974) definition. An equilibrium $\mathbf{s}$ is an ESS if for all $\mathbf{s}' \neq \mathbf{s}$,

$$\mathbf{s} \cdot \vec{\pi}(\mathbf{s}) \geq \mathbf{s}' \cdot \vec{\pi}(\mathbf{s}),$$

with equality implying

$$\mathbf{s} \cdot \vec{\pi}(\mathbf{s}') > \mathbf{s}' \cdot \vec{\pi}(\mathbf{s}').$$

We can think of an ESS as an equilibrium satisfying an evolutionary stability condition that says that once it is fixed in the population, it will do better than any invading strategy as long as this invader is rare. Thomas (1985) reformulates this definition to allow for payoff functions that might be nonlinear.

Definition III.9. An equilibrium $\mathbf{s}$ is an ESS if for all $\mathbf{s}' \neq \mathbf{s}$ in some neighborhood of $\mathbf{s}$,

$$\mathbf{s} \cdot \vec{\pi}(\mathbf{s}') > \mathbf{s}' \cdot \vec{\pi}(\mathbf{s}').$$

We would like to extend this definition of an ESS to the limit of a sequence of games, but there are two ways to do this, depending on whether a different neighborhood of $\mathbf{s}$ may be chosen for each game in the sequence or a single neighborhood

of **s** is chosen for the entire sequence. We are interested in the latter concept, which is a stronger condition, and we call it a *uniformly evolutionarily stable strategy.*

Definition III.10. An equilibrium **s** is a uniformly ESS in the limit as $\vec{P} \to \hat{\vec{P}}$ if there is a punctured neighborhood $\dot{U}(\mathbf{s})$ of **s** (i.e., a neighborhood from which the point **s** is removed) such that for all $\mathbf{s}' \in \dot{U}(\mathbf{s})$ and all $\vec{P} \neq \hat{\vec{P}}$ in some neighborhood of $\hat{\vec{P}}$,

$$\mathbf{s} \cdot \vec{\pi}(\mathbf{s}') > \mathbf{s}' \cdot \vec{\pi}(\mathbf{s}').$$

Note that if equilibrium **s** is strict in the limit as $\vec{P} \to \hat{\vec{P}}$, this implies that for all $\vec{P} \neq \hat{\vec{P}}$ in some neighborhood of $\hat{\vec{P}}$, the state **s** is an ESS, but it does *not* imply that **s** is a *uniformly ESS* in this limit.

An example of a uniformly ESS can be found in the class of games used to prove Theorem III.2, with payoff matrix given by (3.1). In the limit as $N \to \infty$, the equilibrium strategy $(1, 0, 0)$ is a uniformly ESS, but the equilibrium strategy $(0, 1, 0)$ is not.

Our next results will make use of some additional notation. Given a learning rule $\mathcal{R}$ and an equilibrium action a of the game $G(\vec{P})$, let $B(\mathcal{R}, a, \vec{P})$ denote the basin of attraction of $(x_a = 1, x_{-a} = 0)$. Let **R** denote the replicator dynamics and **B** the best response dynamics.

In Theorem III.11 below and the associated Corollary III.12, we show that requiring pure, uniformly ESS to satisfy the *Never an Initial Best Response Property* is necessary if best response dynamics and replicator dynamics are to have basins of attraction with vanishing overlap. In the examples put forth here, this necessary condition entails the existence of either a parasitic action – an action that feeds off other actions but cannot survive on its own – or a misleading action – an action that looks good initially but eventually becomes less attractive as the population evolves.

Theorem III.11. *Suppose for some action s,*

$$\lim_{\vec{P}\to\vec{\hat{P}}} m\left(B(\mathbf{R}, s, \vec{P}) \cap B(\mathbf{B}, s, \vec{P})\right) = 0.$$

Then, if $(x_s = 1, x_{-s} = 0)$ is a uniformly ESS, it satisfies the Never an Initial Best Response Property at $\vec{\hat{P}}$.[5]

Proof. We will denote the equilibrium point $(x_s = 1, x_{-s} = 0)$ by $\mathbf{s}$. Suppose that $\mathbf{s}$ is a uniformly ESS such that $m\left(\mathrm{BR}^{-1}(\mathbf{s})\right)$ remains strictly positive in the limit $\vec{P} \to \vec{\hat{P}}$. We will identify a nonvanishing region inside the basins of attraction of $\mathbf{s}$ for both replicator dynamics and best response dynamics.

Let $U(\mathbf{s})$ be a neighborhood of $\mathbf{s}$ such that $\dot{U}(\mathbf{s}) = U(\mathbf{s})\backslash\{\mathbf{s}\}$ satisfies the condition for $\mathbf{s}$ to be a uniformly ESS. Let $\nu = \sup_{\mathbf{x}\notin U(\mathbf{s})} x_s$. Define the neighborhood $W(\mathbf{s}) \subseteq U(\mathbf{s})$ of all points satisfying $x_s > \nu$. We have constructed $W(\mathbf{s})$ such that $\mathbf{x} \in \dot{W}(\mathbf{s})$ implies that $\dot{x}_s > 0$ under the replicator dynamics (because by the ESS condition, action s has better than average payoff here) and in turn, $\dot{x}_s > 0$ implies that $\mathbf{x}$ remains in $W(\mathbf{s})$.

We now observe that $\mathrm{BR}^{-1}(\mathbf{s})$ is a convex set because of the linearity of payoffs. Additionally, since $\mathbf{s}$ is a pure Nash Equilibrium, $\mathbf{s} \in \mathrm{BR}^{-1}(\mathbf{s})$. Thus, $\mathrm{BR}^{-1}(\mathbf{s})$ and $W(\mathbf{s})$ have positive intersection. By the fact that $W(\mathbf{s})$ is independent of $\vec{P}$ and our hypothesis that $\mathrm{BR}^{-1}(\mathbf{s})$ is nonvanishing, we conclude that $m\left(W(\mathbf{s}) \cap \mathrm{BR}^{-1}(\mathbf{s})\right)$ remains strictly positive in the limit $\vec{P} \to \vec{\hat{P}}$. Note that by the ESS condition and the linearity of payoffs, we can rule out the possibility that there are multiple best responses anywhere in the interior of $\mathrm{BR}^{-1}(\mathbf{s})$. For points $\mathbf{x}$ in the interior of $W(\mathbf{s}) \cap \mathrm{BR}^{-1}(\mathbf{s})$, best response dynamics flows to $\mathbf{s}$ because $\mathrm{BR}(\mathbf{x}) = \{\mathbf{s}\}$ and replicator dynamics flows to $\mathbf{s}$ because $\mathbf{x} \in W(\mathbf{s})$. $\qquad\square$

---

[5]If we were to suppose that best response dynamics and replicator dynamics share vanishing overlap in their basins of attraction for an interior equilibrium, we could immediately conclude that this equilibrium is not a uniformly ESS. Interior ESS are, as already mentioned, globally asymptotically stable for both replicator and best response dynamics.

Theorem III.11 leads directly to the following corollary.

Corollary III.12. *Suppose*

$$\lim_{\vec{P}\to\hat{\vec{P}}}\sum_s m\left(B(\mathbf{R},s,\vec{P})\cap B(\mathbf{B},s,\vec{P})\right) = 0.$$

Then every pure, uniformly ESS satisfies the Never an Initial Best Response Property at $\hat{\vec{P}}$.

Corollary III.12 provides a necessary condition for non-overlapping basins. We can also derive several different sets of conditions that are sufficient to generate vanishing overlap in the basins of attraction of strict equilibria with best response and replicator dynamics. We present one such set of sufficient conditions for a symmetric three-by-three game here. Observe that the conditions we present are satisfied by the class of games used in the proof of Theorem III.2.

To describe these conditions, we introduce some new notation and some simplifying assumptions. Let π_{ij} be the payoff to action i against action j, which by definition depends on the parameters $\vec{P}$. Since both dynamics are invariant under the transformations $\pi_{ij} \to \pi_{ij} + c$ for all i and fixed j and $\pi_{ij} \to k\pi_{ij}$ for all i, j with $k > 0$, we can set $\pi_{3j} = 0$ for all j and $|\pi_{11}| \in \{0,1\}$. Also without loss of generality we can renumber the three actions so that $(x_1 = 1, x_{-1} = 0)$ denotes the equilibrium attained by replicator dynamics and $(x_2 = 1, x_{-2} = 0)$ the equilibrium attained by best response dynamics. Because these equilibria are strict in the limit as $\vec{P} \to \hat{\vec{P}}$, we have that for $j \in \{1,2\}$, $i \neq j$, $\lim_{\vec{P}\to\hat{\vec{P}}} f_{jji}(\vec{P})(\pi_{jj} - \pi_{ij}) > 0$ for some functions $f_{jji} > 0$. And, by our choice of which equilibrium is to be found by each dynamic, we also have $\lim_{\vec{P}\to\hat{\vec{P}}} f_{321}(\vec{P})(\pi_{23} - \pi_{13}) > 0$ for some function $f_{321} > 0$.

Theorem III.13.

$$\lim_{\vec{P}\to\hat{\vec{P}}}\sum_{i=1}^{2} m\left(B(\mathbf{R},i,\vec{P})\cap B(\mathbf{B},i,\vec{P})\right) = 0$$

if: i) $\pi_{23} > 0$; *ii)* $\pi_{13} \leq 0$ *and* $\lim_{\vec{P}\to\vec{\hat{P}}} \pi_{13} = 0$;[6] *iii)* $\lim_{\vec{P}\to\vec{\hat{P}}} \pi_{12} = -\infty$; *iv)*

$\lim_{\vec{P}\to\vec{\hat{P}}} \frac{\pi_{21}}{\pi_{12}} = \infty$; *v)* $\lim_{\vec{P}\to\vec{\hat{P}}} \frac{\pi_{21}}{\pi_{22}} = -\infty$; *and vi)* $\lim_{\vec{P}\to\vec{\hat{P}}} \frac{\pi_{21}}{\pi_{23}} = -\infty$.

The proof relies on two lemmas, one for each learning dynamic.

Lemma III.14. *As $\vec{P}$ approaches $\vec{\hat{P}}$, the fraction of the action space inside $B(\mathbf{B}, 2, \vec{P})$ approaches 1.*

Proof. We first show that actions 1 and 2 satisfy the *Never an Initial Best Response Property* at $\vec{\hat{P}}$, that action 3 is initially a best response in all but an arbitrarily small part of the action space when $\vec{P}$ nears $\vec{\hat{P}}$. By the normalization condition, $\pi_3 = 0$. Therefore, it suffices to show $\pi_1 < 0$ and $\pi_2 < 0$.

1. $\boldsymbol{\pi_2 < 0}$. Assuming $x_1 > 0$, $\pi_2 = x_1\left(\pi_{21} + \frac{x_2}{x_1}\pi_{22} + \frac{x_3}{x_1}\pi_{23}\right)$. Condition (v) implies π_{21} dominates $\frac{x_2}{x_1}\pi_{22}$. Condition (vi) implies π_{21} dominates $\frac{x_3}{x_1}\pi_{23}$. And π_{21} is negative. So, for $\vec{P}$ near $\vec{\hat{P}}$, $\pi_2 < 0$.

2. $\boldsymbol{\pi_1 < 0}$. Assuming $x_2 > 0$, $\pi_1 = x_2\left(\pi_{12} + \frac{x_1}{x_2}\pi_{11} + \frac{x_3}{x_2}\pi_{13}\right)$. The normalization conditions imply $\pi_{11} = 1$. Condition (iii) states that π_{12} approaches $-\infty$ while condition (ii) states that $\pi_{13} \leq 0$. So, for $\vec{P}$ near $\vec{\hat{P}}$, $\pi_1 < 0$.

Thus, for any point in the interior of the action space, $\vec{P}$ can be chosen such that action 3 is initially a best response.

Now we show that under best response dynamics, action 3 dominates action 1 along the path towards $(0, 0, 1)$. Under best response dynamics, actions which are not best responses have the same relative decay rates. So $\frac{x_1}{x_2}$ remains constant along the path towards $(0, 0, 1)$. So π_1 remains negative along this path. By condition (i), action 3 is not a best response to itself. Eventually action 2 becomes the best response.

---

[6] Another set of sufficient conditions might allow $\pi_{13} > 0$, but would then require additional conditions to ensure that the best response dynamics avoids selecting $(1, 0, 0)$.

As the dynamic then moves toward $(0, 1, 0)$, π_1 remains negative because the π_{12} term becomes even more significant relative to the others. Action 1 never becomes the best response, so the best response dynamics lead to $(0, 1, 0)$. $\square$

Lemma III.15. *As $\vec{P}$ approaches $\hat{\vec{P}}$, the fraction of the action space inside $B(\mathbf{R}, 1, \vec{P})$ approaches 1.*

Proof. Under the replicator dynamics,

$$\dot{x}_1 = x_1 \left(\pi_{11}x_1(x_2 + x_3) + \pi_{12}x_2(x_2 + x_3) + \pi_{13}x_3(x_2 + x_3) - \pi_{21}x_1x_2 - \pi_{22}x_2^2 - \pi_{23}x_3x_2 \right).$$

Consider initial points that satisfy $x_1 > -\pi_{13}$ and $x_2 > 0$. Recalling that $\pi_{11} = 1$, this gives

$$(3.3) \qquad\qquad \pi_{11}x_1(x_2 + x_3) + \pi_{13}x_3(x_2 + x_3) > 0.$$

By conditions (iv), (v), and (vi), $|\pi_{21}|$ grows faster than $|\pi_{12}|$, π_{22}, and π_{23} as $\vec{P}$ nears $\hat{\vec{P}}$. Consequently, the term with π_{21} dominates the other remaining terms in the expansion of $\dot{x}_1$. So, for $\vec{P}$ near $\hat{\vec{P}}$,

$$(3.4) \qquad\qquad \pi_{12}x_2(x_2 + x_3) - \pi_{21}x_1x_2 - \pi_{22}x_2^2 - \pi_{23}x_3x_2 > 0.$$

Thus, initially $\dot{x}_1 > 0$. Moreover, by choosing $\vec{P}$ such that $\pi_{21} < \frac{1}{x_1(0)}(\pi_{12} - \pi_{22} - \pi_{23})$, we can be sure equation (3.4) holds as x_1 increases. As x_1 increases, it remains above $-\pi_{13}$, so equation (3.3) continues to hold as well. Thus, $\dot{x}_1 > 0$ at all times.

It remains to show that the fraction of the action space satisfying $x_1 > -\pi_{13}$ and $x_2 > 0$ approaches 1 as $\vec{P}$ approaches $\hat{\vec{P}}$. This follows from (ii), which states that $\lim_{\vec{P} \to \hat{\vec{P}}} \pi_{13} = 0$. This implies that a point $\mathbf{x}$ need only satisfy $x_1 > 0$ and $x_2 > 0$ to be in $B(\mathbf{R}, 1, \vec{P})$ for some $\vec{P}$ near $\hat{\vec{P}}$. $\square$

We have thus described a set of six conditions which generate vanishing overlap in basins of attraction with best response dynamics and replicator dynamics in a class of games with only three actions.

Admittedly, none of the games within this class may be likely to arise in the real world. However, if we widen our scope and allow for more strategies, we can find games that map more tightly to real world phenomena and exhibit this same behavior. Consider the following symmetric matrix game with four actions, selected from a class of generalized stag hunt games that we explore further in Chapter IV (Golman and Page, 2008):

$$
\begin{pmatrix}
2 & 2 & 2 & 2 \\
1 & N+1 & 1 & 1 \\
0 & 0 & 0 & N^2 \\
0 & 0 & -N^2 & 0
\end{pmatrix}.
$$

In this game, the first action is a safe, self interested action like hunting hare. The second action represents an attempt to cooperate, to hunt a stag, for example. The third action is predatory toward the fourth action, which can be thought of as a failed attempt at cooperation. This fourth action fails to protect itself from the predator, fails to accrue benefits from coordination, and fails to guarantee itself a positive payoff. Clearly, a rational player would not choose it, and it is not played in equilibrium. Nevertheless, introducing predation into the stag hunt enhances the strategic context. This game captures a choice between the security of self-reliance, the productivity of cooperation, or the temptation of exploiting those agents who haven't yet learned what not to do. As we now show, when N goes to infinity, best response dynamics flow to an equilibrium in which all players choose action 1, but

replicator dynamics flow to an equilibrium in which all players choose action 2.

Proposition III.16. *In the four-by-four game above, as $N \to \infty$, the Lebesgue measure of the set of initial starting points for which best response dynamics and replicator dynamics flow to the same equilibrium tends to zero.*

Once again, the proof relies on three lemmas, one to identify the stable equilibria and two to describe the behavior of the learning rules.

Lemma III.17. *Both best response dynamics and replicator dynamics have two stable fixed points:* $\mathbf{x} = (1, 0, 0, 0)$ *and* $\mathbf{x} = (0, 1, 0, 0)$.

Proof. Here again, both configurations are strict Nash Equilibria because each of action 1 and 2 is a strict best response to itself. The only other Nash Equilibrium, $\mathbf{x} = \left(\frac{N-1}{N}, \frac{1}{N}, 0, 0\right)$, is clearly unstable given either dynamics. Note that action 4 is strictly dominated, and if we apply iterated elimination of strictly dominated actions, action 3 becomes strictly dominated once action 4 is eliminated. □

The next lemma shows that for large N, best response dynamics leads to action 1 starting from almost any initial condition.

Lemma III.18. *For any ϵ, there exists M such that for all $N \geq M$, the basin of attraction of $(1, 0, 0, 0)$ given best response dynamics is at least $1 - \epsilon$ of the action space.*

Proof. First we show any point with $x_4 > \frac{2}{N}$ is in the basin of attraction of $(1, 0, 0, 0)$, assuming $N > 2$. For such a point, action 3 is initially a best response because $\pi_3 > 2N$ whereas $\pi_1 = 2$, $\pi_2 < 1 + N$, and $\pi_4 < 0$. Then, as we show, action 1 becomes a best response before action 2. Once it becomes a best response, it remains one forever, because its payoff is constant, while the payoffs to actions 2

and 3 are decreasing. So, once action 1 becomes a best response, the dynamic flows toward it thereafter.

Now we show that action 1 does become the best response before action 2. We define

$$\alpha(t) = \frac{x_1(t)}{x_1(0)} = \frac{x_2(t)}{x_2(0)} = \frac{x_4(t)}{x_4(0)}$$

for t such that action 3 is still a best response. The latter equalities hold because actions which are not best responses have the same relative decay rate. Note that $\alpha(t)$ is a strictly decreasing function. Now

$$\pi_1 = \pi_3 \text{ when } \alpha = \frac{2}{N^2 (x_4(0))}.$$

But

$$\pi_2 < \pi_3 \text{ if } \alpha > \frac{1}{N(Nx_4(0) - x_2(0))}.$$

Action 1 eventually becomes the best response because

$$\frac{2}{N^2 (x_4(0))} > \frac{1}{N(Nx_4(0) - x_2(0))},$$

as long as $Nx_4(0) > 2x_2(0)$. This condition holds if $x_4(0) > \frac{2}{N}$.

Thus, by choosing N large enough, the basin of attraction of $(1,0,0,0)$ can be made as big as desired. $\square$

Unlike best response dynamics, for large N, replicator dynamics leads to almost all players taking action 2 for almost any initial condition.

Lemma III.19. *For any ϵ, there exists M such that for all $N \geq M$, the basin of attraction of $(0,1,0,0)$ given replicator dynamics is at least $1 - \epsilon$ of the action space.*

Proof. We now have $\dot{x_2} = x_2\left((1 + Nx_2)(1 - x_2) - 2x_1\right)$. So $\dot{x_2} \geq 0$ if $x_2 \geq \frac{1}{N}$. By choosing N large enough, the basin of attraction of $(0,1,0,0)$ can be made as big as desired. $\square$

This completes the proof of Proposition III.16. In this class of games, replicator dynamics flows to the equilibrium with the higher payoff, whereas in the class of games used in the proof of Theorem III.2, the best response dynamics flows to the optimal equilibrium. Neither learning dynamic can find the optimal equilibrium in all classes of games because a different set of normalization conditions can change which equilibrium is optimal.

3.4 Broader Classes of Dynamics

We introduce two new classes of adjustment dynamics: one-sided payoff positive dynamics, which generalize the replicator dynamics, and threshold dynamics, a generalization of the best response dynamics. We then extend our results from the previous section to describe vanishing overlap in the basins of attraction of a one-sided payoff positive dynamic and a threshold dynamic.

As the name suggests, our *one-sided payoff positive dynamics* are closely related to the commonly known payoff positive dynamics (Weibull, 1995). Payoff positive dynamics assume that actions with above average payoffs have positive relative growth rates and actions with below average payoffs have negative relative growth rates.[7] The one-sided class of dynamics still captures the property that actions with above average payoffs grow in the population, but does not address what happens to actions with below average payoffs. Thus, the class of one-sided payoff positive dynamics includes all the payoff positive dynamics. They in turn contain the replicator dynamics, which prescribe a relative growth rate proportional to the difference between action's payoff and population mean payoff. Neither class of dynamics specifies precise rates of growth the way replicator does, making them both quite general.

Definition III.20. A *one-sided payoff positive dynamic* is one that satisfies the

---

[7]These dynamics are also termed sign-preserving (Nachbar, 1990).

following condition:

$$(3.5) \qquad\qquad \dot{x}_i > 0 \text{ if } \pi_i > \bar{\pi} \text{ and } x_i > 0$$

as well as the requirements that Nash Equilibria are rest points and that if **z** is the limit of an interior orbit for $t \to \infty$, then **z** is a Nash Equilibrium[8].

These basic requirements are satisfied by most adjustment dynamics as part of a folk theorem (see Cressman, 2003). The second requirement holds whenever relative growth rates are Lipschitz continuous, for example. The other statements of the folk theorem, namely that stable rest points are Nash Equilibria and that strict equilibria are asymptotically stable, can be shown to hold from our definition of one-sided payoff positive dynamics.

Lipschitz continuous one-sided payoff positive dynamics fit into the even broader class of weakly payoff positive dynamics, which assume that some action with above average payoff will have positive relative growth whenever there is such an action. The distinction is that weak payoff positivity does not guarantee growth for *all* the above average actions in the population.

In contrast to one-sided payoff positive dynamics, in which agents seek actions with above average payoffs, we can conceive of a learning rule in which agents switch actions when their payoffs are at or below the median. But, there is no need to hold the 50th percentile payoff in such special regard as the threshold for switching. We define *threshold dynamics* by the property that agents switch away from actions with payoffs at or below the Kth percentile as long as there is something better to switch to. We do not restrict ourselves to a particular threshold by setting a value for K. Instead, we allow K to vary over time within a range that is bounded below by some

---

[8]Theorem III.25 would hold without these requirements, but with the possibility that the one-sided payoff positive dynamics have measure zero basins of attraction for all strict equilibria. We want to focus on the case that the one-sided payoff positive dynamic selects a different equilibria than the threshold dynamic.

$\hat{K} > 0$. We sometimes omit writing the input t when dealing with a fixed instant in time and use K to mean the function $K(t)$. But $\hat{K}$ is always a constant. We want agents to be as averse as possible to actions whose payoffs fall below the threshold, but recognizing that the speed of the dynamics can be adjusted by an overall scaling factor, we only require that such an action has a relative rate of decline as quick as any in the population.

Definition III.21. Consider any $K(t) \geq \hat{K}$ where $\hat{K} > 0$. A *threshold dynamic* is one that satisfies the following condition:

If at time t

$$(3.6) \qquad \sum_{\mu:\, \pi_\mu < \pi_i} x_\mu < K \text{ and for some } l,\ \pi_l > \pi_i,$$

then when $x_i > 0$,

$$(3.7) \qquad \frac{\dot{x}_i}{x_i} \leq \frac{\dot{x}_j}{x_j} \text{ for all actions } j \text{ such that } x_j > 0$$

and when $x_i = 0$, $\dot{x}_i = 0$.

Note that if two actions both have payoffs below the Kth percentile, they must have the same relative rate of decline. In addition, it is always the case that the action with the worst payoff declines in the population. On the other hand, there is no guarantee that the best response grows in the population unless other actions are sufficiently rare.

Definition III.22. A threshold dynamic is *properly scaled* if the speed of the dynamic has a lower bound $v(\mathbf{x})$ such that $v(\mathbf{x}) \geq \kappa d$ where d represents the distance to the nearest equilibrium point and κ is some constant of proportionality.

As previously mentioned, the speed of a dynamic can be adjusted by an overall scaling factor. Usually we do not care about speed at all because the scaling of the

time parameter does not have physical significance. In this case, we must assume a threshold dynamic is properly scaled in order to ensure that it does not slow to a halt at an arbitrary point in the strategy space. In this chapter, we consider only properly scaled threshold dynamics.

Let us briefly consider a few particular constant functions we might use for $K(t)$ in the threshold dynamics. If we do want the median payoff to be the threshold, we can choose $K(t) = .5$ for all t. Then all actions with payoffs equal to or below the median will have the same relative rate of decline, and actions with payoffs above the median will do no worse. That is, an action with a higher payoff may still decline at the same relative rate as the former or may grow very quickly; the threshold dynamics allow for either. This example suggests that when $K(t)$ is small, the dynamics allow for quite a bit of freedom in the center of the strategy space.

Alternatively, we could set the threshold below which agents switch actions to be the 100th percentile payoff at all times, $K(t) = 1$. Obviously, every action has a payoff at or below the 100th percentile payoff, so inequality (3.7) applies to every action present in the population that is not a best response. An agent already playing a best response cannot find something better to switch to, and thus best responses are not subject to this inequality. Every action that is subject to inequality (3.7) has the same relative rate of decline, and if we set this rate to be -1, we then obtain the best response dynamics. Thus, threshold dynamics are a generalization of the best response dynamics.

We now show that requiring pure, uniformly ESS to satisfy the *Never an Initial Best Response Property* is necessary if a one-sided payoff positive dynamic and a threshold dynamic are to have basins with vanishing overlap, just as Theorem III.11 and Corollary III.12 showed it is for best response and replicator dynamics. Let

OSPP denote any one-sided payoff positive dynamic and **TD** any threshold dynamic.

Theorem III.23. *Suppose for some action s,*

$$\lim_{\vec{P} \to \vec{\hat{P}}} m\left(B(\mathbf{OSPP}, s, \vec{P}) \cap B(\mathbf{TD}, s, \vec{P}) \right) = 0.$$

Then, if $(x_s = 1, x_{-s} = 0)$ is a uniformly ESS, it satisfies the Never an Initial Best Response Property at $\vec{\hat{P}}$.

Proof. The proof here mirrors the one for Theorem III.11. We construct the neighborhood $W(\mathbf{s})$ in the same way, but with the additional condition that $x_s > 1 - \hat{K}$. We need only show that for $\mathbf{x} \in \text{int}\left(W(\mathbf{s}) \cap \text{BR}^{-1}(\mathbf{s})\right)$, both classes of dynamics flow to $\mathbf{s}$. Under one-sided payoff positive dynamics, $\dot{x}_s > 0$ for $\mathbf{x} \in W(\mathbf{s})$ because action s has an above average payoff, and such a flow cannot leave $W(\mathbf{s})$. Under threshold dynamics, when $\mathbf{x} \in \text{int}\left(W(\mathbf{s}) \cap \text{BR}^{-1}(\mathbf{s})\right)$, inequality (3.7) applies to all actions other than s because they have payoffs below the $\hat{K}^{\text{th}}$ percentile. All other actions must have the same negative growth rate, so $\dot{\mathbf{x}} = \kappa(\mathbf{s} - \mathbf{x})$ for some positive constant κ. $\qquad\qquad\square$

Corollary III.24. *Suppose*

$$\lim_{\vec{P} \to \vec{\hat{P}}} \sum_s m\left(B(\mathbf{OSPP}, s, \vec{P}) \cap B(\mathbf{TD}, s, \vec{P}) \right) = 0.$$

Then every pure, uniformly ESS satisfies the Never an Initial Best Response Property at $\vec{\hat{P}}$.

We also find that the same set of conditions used in Theorem III.13 is sufficient for a one-sided payoff positive dynamic and a threshold dynamic to share vanishing overlap in their basins. Recall that the setting for this theorem is a symmetric three-by-three game with two strict equilibria.

Without loss of generality we choose $(x_1 = 1, x_{-1} = 0)$ to be the equilibrium attained by the one-sided payoff positive dynamic and $(x_2 = 1, x_{-2} = 0)$ the equilibrium attained by the threshold dynamic. Because these equilibria are strict in the limit as $\vec{P} \to \vec{\hat{P}}$, we have that for $j \in \{1, 2\}$, $i \neq j$, $\lim_{\vec{P} \to \vec{\hat{P}}} f_{jji}(\vec{P})(\pi_{jj} - \pi_{ij}) > 0$ for some functions $f_{jji} > 0$. And, by our choice of which equilibrium is to be found by each dynamic, we also have $\lim_{\vec{P} \to \vec{\hat{P}}} f_{321}(\vec{P})(\pi_{23} - \pi_{13}) > 0$ for some function $f_{321} > 0$. Once again, we set $\pi_{3j} = 0$ for all j and $\pi_{11} = 1$, but no longer are our dynamics necessarily invariant under positive affine transformations of the payoffs. If the dynamics happen to retain this invariance, then this still amounts to a choice of payoff normalization. However, in general, we are making an additional assumption about payoffs here.

Theorem III.25.

$$\lim_{\vec{P} \to \vec{\hat{P}}} \sum_{i=1}^{2} m \left(B(\textbf{OSPP}, i, \vec{P}) \cap B(\textbf{TD}, i, \vec{P}) \right) = 0$$

if: i) $\pi_{23} > 0$; ii) $\pi_{13} \leq 0$ and $\lim_{\vec{P} \to \vec{\hat{P}}} \pi_{13} = 0$; iii) $\lim_{\vec{P} \to \vec{\hat{P}}} \pi_{12} = -\infty$; iv) $\lim_{\vec{P} \to \vec{\hat{P}}} \frac{\pi_{21}}{\pi_{12}} = \infty$; v) $\lim_{\vec{P} \to \vec{\hat{P}}} \frac{\pi_{21}}{\pi_{22}} = -\infty$; and vi) $\lim_{\vec{P} \to \vec{\hat{P}}} \frac{\pi_{21}}{\pi_{23}} = -\infty$.

Again, we break up the proof into two lemmas, one for each learning dynamic.

Lemma III.26. *As $\vec{P}$ approaches $\vec{\hat{P}}$, the fraction of the action space inside $B(\textbf{TD}, 2, \vec{P})$ approaches* 1.

Proof. Consider the threshold dynamics with any value of $\hat{K}$ and any threshold function $K(t) \geq \hat{K}$. We show that if initially $x_2 > 0$, then for $\vec{P}$ near $\vec{\hat{P}}$, $\pi_1 < \pi_3$ at all times. As $\pi_3 = 0$ by the normalization condition, this amounts to showing $\pi_1 < 0$ forever. We know $\pi_1 = \pi_{11}x_1 + \pi_{12}x_2 + \pi_{13}x_3$. Recall that $\pi_{11} = 1$. Condition (ii) states that $\pi_{13} \leq 0$. So $\pi_1 < 0$ as long as

$$(3.8) \qquad\qquad x_1 + \pi_{12}x_2 < 0.$$

Consider first the case that $x_2(0) \geq \hat{K}$. Take $\vec{P}$ near enough $\hat{\vec{P}}$ that $\pi_{12} < -\frac{1}{\hat{K}}$. Condition (iii) makes this possible. As long as $x_2 \geq \hat{K}$, equation (3.8) holds and we still have $\pi_1 < 0$.

In the case that $x_2(0) < \hat{K}$, condition (iii) allows us to take $\vec{P}$ near enough $\hat{\vec{P}}$ that $\pi_{12} < -\frac{1}{x_2(0)}$. This guarantees that $\pi_1 < 0$ initially.

If ever $x_2 < K$, then π_1 is below the Kth percentile and $\frac{\dot{x}_1}{x_1} \leq \frac{\dot{x}_2}{x_2}$, so equation (3.8) continues to hold. Still $\pi_1 < 0$.

In fact, the only way to avoid $x_2 < K$ at some time would involve $\pi_2 > 0$ pretty quickly. But, if indeed $x_2 < K$ at some time, then the decline in x_1 also would lead to $\pi_2 > 0$ eventually. So, one way or another, action 2 becomes the best response and x_1 has to decline. When $x_1 < \hat{K} \leq K$, the Kth percentile payoff is either 0 or π_2, and when additionally $\pi_2 > 0$, then only x_2 can grow. From then on, the dynamic moves straight toward $(0, 1, 0)$. $\qquad\square$

Lemma III.27. *As $\vec{P}$ approaches $\hat{\vec{P}}$, the fraction of the action space inside $B(\mathbf{OSPP}, 1, \vec{P})$ approaches 1.*

Proof. The proof of Lemma III.15, which applied replicator dynamics to this game, carries over here, applying to all one-sided payoff positive dynamics with only trivial changes. We no longer have an exact formula for $\dot{x}_1$, but the argument that it is always positive still applies because it was based on the fact that $\pi_1 > \bar{\pi}$ at all times for almost all initial points. The definition of a one-sided payoff positive dynamic requires that the limit of an interior orbit is a Nash Equilibrium, and the only one that can be approached with $\dot{x}_1 > 0$ is $(1, 0, 0)$. $\qquad\square$

We have thus extended our finding of vanishing overlap in basins of attraction for strict equilibria to entire classes of dynamics.

3.5   Discussion

In this chapter, we have shown that it is possible to construct three-by-three symmetric games in which two common learning rules – replicator dynamics, and best response dynamics – have vanishing overlap in their basins of attraction. That so few actions are required is surprising, making the game we have constructed of significant pedagogical value. Our more general results describe necessary and sufficient conditions for vanishing overlap. The necessary condition – that for any game in which the learning rules attain distinct strict equilibria from almost any starting point the initial best response cannot be a uniformly ESS - has an intuitive explanation. The initial incentives must be misleading. They should point the agents away from equilibria and in some other directions. In doing so, these initial incentives allow for even small differences in the dynamics to take root and drive the two learning rules to distinct equilibria.

We also derived a set of sufficient conditions for the basins of attraction of two stable equilibria under best response learning and replicator dynamics to have almost no overlap. Other sufficient conditions could also be constructed. What appears invariant to the construction is that some payoffs must grow arbitrarily large.

Our focus on basins of attraction differentiates this chapter from previous studies that consider stability. Nash was aware that the existence of an equilibrium is not sufficient proof that it will arise. Nor is proof of its local stability. We also need to show how to attain an equilibrium from an arbitrary initial point (Binmore and Samuelson, 1999). And, as we have just shown, the dynamics of how people learn can determine whether or not a particular equilibrium is attained. Richer models of individual and firm behavior can also support diverse choices of equilibria (Allen,

Strathern, and Baldwin, 2007). Here, we emphasized the minimal conditions necessary for the learning rule to matter.

We also focused on the extreme case of no overlap. That said, our general findings about the necessity of misleading actions and the nature of our sufficient conditions should help us to identify games in which learning rules might matter. In particular, the idea that temporary best responses create opportunity for differences in learning rules to accumulate would seem to have wide applicability. It provides logical foundations for the intuition that learning rules matter more in more complex environments.

In conclusion, we might add that games in which best response dynamics and replicator dynamics make such different equilibrium predictions would seem to lend themselves to experiments. These games would allow experimenters to distinguish among learning rules more decisively than games in which the learning rules converge to the same equilibrium.

BIBLIOGRAPHY

[1] Allen P.M, Strathern, M., Baldwin, J. S., 2007. Complexity and the limits to learning. *J. Evol. Econ.* 17, 401-431.

[2] Basov, S., 2004. Equilibrium selection in coordination games: Why do dominated strategies matter? *Econ. Bull.* 3, 1-3.

[3] Binmore, K., Samuelson, L., 1999. Evolutionary Drift and Equilibrium Selection. *Rev. Econ. Stud.* 66, 363-394.

[4] Brown, G.W.. 1951. Iterative Solutions of Games by Fictitious Play. In: Koopmans, T.C. (Ed.), *Activity Analysis of Production and Allocation.* Wiley, New York.

[5] Brown, G.W., von Neumann, J., 1950. Solutions of games by differential equations. *Ann. Math. Stud.* 24, 73-79.

[6] Camerer, C., 2003. *Behavioral Game Theory: Experiments on Strategic Interaction.* Princeton University Press, Princeton.

[7] Camerer, C., Ho, T.H., 1999. Experience Weighted Attraction Learning in Normal Form Games. *Econometrica* 67, 827-874.

[8] Cressman, R., 2003. *Evolutionary Dynamics and Extensive Form Games.* MIT Press, Cambridge.

[9] Ellison, G. 2000. Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Rev. Econ. Stud.* 67 (1), 17-45.

[10] Erev, I., Roth, A.E., 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Amer. Econ. Rev.* 88 (4), 848-881.

[11] Feltovich, N., 2000. Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games. *Econometrica* 68, 605-641.

[12] Foster, D., Young, H.P., 1990. Stochastic evolutionary game dynamics. *Theoretical Population Biology* 38, 219-232.

[13] Fudenberg, D., Levine, D.K., 1998. *The Theory of Learning in Games.* MIT Press, Cambridge.

[14] Gilboa, I., Matsui, A., 1991. Social Stability and Equilibrium. *Econometrica* 59, 859-867.

[15] Golman, R., Page, S., 2008. Individual and Cultural Learning in Stag Hunt Games With Multiple Actions.

[16] Govindan, S., Wilson, R., 2005. Refinements of Nash Equilibrium. In: Durlauf, S., Blume, L. (Eds.), *The New Palgrave Dictionary of Economics* 2. Palgrave Macmillan, London.

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

[18] Haruvy, E., Stahl, D.O., 1999. Empirical Tests of Equilibrium Selection based on Player Heterogeneity. mimeo, University of Texas at Austin.

[19] Haruvy, E., Stahl, D.O., 2000. Robust Initial Conditions for Learning Dynamics. mimeo, University of Texas at Austin.

[20] Hauert, C., Haiden, N., Sigmund, K., 2004. The Dynamics of Public Goods. *Discrete and Continuous Dynamical Systems - Series B* 4, 575-587.

[21] Hofbauer, J., 1995. Stability for the best response dynamics. mimeo, University College London.

[22] Hofbauer, J., 2000. From Nash and Brown to Maynard Smith: equilibria, dynamics and ESS, *Selection* 1, 81-88.

[23] Hofbauer, J., Schuster, P., Sigmund, K., 1979. A note on evolutionary stable strategies and game dynamics. *J. Theor. Biol.* 81, 609-612.

[24] Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics.* Cambridge University Press, Cambridge, U.K.

[25] Hofbauer, J., Sigmund, K., 2003. Evolutionary Game Dynamics. *Bull. Am. Math. Soc.* 40, 479-519.

[26] Hopkins, E., 1999. A Note on Best Response Dynamics. *Games Econ. Behav.* 29, 138-150.

[27] Kandori, M., Mailath, G., Rob, R., 1993. Learning, mutation and long run equilibria in games. *Econometrica* 61, 29-56.

[28] Kojima, F., 2006. Stability and Instability of the Unbeatable Strategy in Dynamic Processes. *Int. J. Econ. Theory* 2, 41-54.

[29] Maynard Smith, J., 1974. The Theory of Games and the Evolution of Animal Conflicts. *J. Theor. Biol.* 47, 209-221.

[30] Nachbar, J. H., 1990. "Evolutionary" Selection Dynamics in Games: Convergence and Limit Properties. *Int. J. Game Theory* 19, 59-89.

[31] Salmon, T., 2001. An Evaluation of Econometric Models of Adaptive Learning. *Econometrica* 69, 1597-1628.

[32] Samuelson, L., 1997. *Evolutionary Games and Equilibrium Selection.* MIT Press, Cambridge.

[33] Sandholm, W. H., 2009. Pairwise Comparison Dynamics and Evolutionary Foundations for Nash Equilibrium, working paper, University of Wisconsin. http://www.ssc.wisc.edu/~whs.

[34] Sandholm, W.H., Dokumaci, E., 2007. *Dynamo: Phase Diagrams for Evolutionary Dynamics.* Software suite. http://www.ssc.wisc.edu/~whs/dynamo.

[35] Schlag, K. H., 1998. Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-armed Bandits. *J. Econ. Theory* 78, 130-156.

[36] Skyrms, B., 1990. *The Dynamics of Rational Deliberation.* Harvard University Press, Cambridge.

[37] Swinkels, J. M., 1993. Adjustment dynamics and rational play in games. *Games Econ. Behav.* 5, 455-484.

[38] Taylor, P.D., Jonker, L., 1978. Evolutionarily stable strategies and game dynamics. *Math. Biosci.* 40, 145-156.

[39] Thomas, B., 1985. On Evolutionarily Stable Sets. *J. Math. Biol.* 22, 105-115.

[40] Vaughan, R. N. Equilibrium Selection: Basins of Attraction vs. Potential Maxima. mimeo, University College London.

[41] Walliser, B., 1998. A spectrum of equilibration processes in game theory. *J. Evol. Econ.* 8, 67-87.

[42] Weibull, J.W., 1995. *Evolutionary Game Theory.* MIT Press, Cambridge.

[43] Weibull, J.W., 1996. Contribution to Nobel Seminar: The Work of John Nash in Game Theory. *J. Econ. Theory* 69, 153-185.

CHAPTER IV

Individual and Cultural Learning in Stag Hunt Games With Multiple Actions

4.1   Introduction

Game theorists motivate the prediction of equilibrium outcomes with one of two assumptions. Agents can be rational, in which case they choose strategies that form an equilibrium. Or, alternatively, agents can learn. In this case, they eventually settle into an equilibrium. When a game possesses multiple equilibria, the assumption of rational agents requires the introduction of refinement criteria to select from among the equilibria. With learning agents, the explicit model of behavior determines a basin of attraction for each equilibrium.

Surprisingly, the type of learning rule generally does not matter for the stability of equilibria – for most games, the set of stable equilibria are invariant for broad classes of learning rules. However, stable equilibria can have small basins (Epstein, 2003) and basin size can vary depending on the learning rule. In fact, as we saw in Chapter III, games with as few as three actions can have basins of attraction with vanishing overlap for different learning rules (Golman and Page, 2008). [1] In sum, how people learn might not affect the existence or stability of an equilibrium, but it

---

[1] Basins of attraction have also been the focus of a related literature that considers dynamical models with persistent randomness and selects stochastically stable equilibria (Foster and Young, 1990; Kandori et al., 1993; Young, 1993; Kandori and Rob, 1995). These models favor risk dominant solutions.

can strongly influence which equilibrium gets chosen.

A natural practical question to ask is whether differences in learning rules can lead to qualitatively different outcomes. In particular, does one type of learning rule make cooperative behavior more likely to emerge in equilibrium than another? To get at that question, in this chapter, we explore the extent to which the type of learning rule influences equilibrium selection in a class of generalized stag hunt games. We compare two canonical learning models: *cultural learning* and *individual belief-based learning.* We find that the former more often settles into a cooperative equilibrium. This does not imply that cultural learning is better for all games (see Chapter III for a counterexample), but it does suggest that for games that involve coordination on a cooperative action, cultural learning may be a preferred learning rule.

The original stag hunt game traces back to 1773, when Rousseau proposed the story of a stag hunt to represent a choice in which the benefits of cooperation conflict with the security of acting alone. In the story, two individuals must each choose to hunt a stag or to hunt a hare. Hunting stags can only be successful with cooperation, while hunting a hare does not require the other player's help. The catch is that the stag offers both hunters a lot more meat than the hare. Thus, the stag hunt obliges a choice between productivity and security. Skyrms (2001) argues that the stag hunt captures the incentives present in choices whether to adopt or modify the social contract.

Rousseau's stag hunt has been modeled as a two-by-two game with two strict pure Nash Equilibria: an efficient one in which both hunt stag and an inefficient one in which both hunt hare. In playing a stag hunt, agents try to figure out which equilibrium action to choose. Equilibrium selection arguments can be invoked in

favor of either action. While hunting the stag is payoff dominant, hunting the hare is risk dominant (Harsanyi and Selten, 1988).

The stag hunt game, along with the prisoner's dilemma, is often invoked as a framework with which to study collective action problems (Medina, 2007). In a prisoner's dilemma or a public goods game, incentives lead to everybody defecting. In contrast, in a stag hung game, players have an incentive to cooperate provided enough of the other players do so as well. Stag hunt differs from the prisoners' dilemma in that achieving cooperation does not require higher order strategies such as tit for tat (Axelrod, 1984), trigger mechanisms (Abreu et al., 1990), or norm based strategies (Bendor and Swistak, 1977), all of which allow defectors to be punished outside of the context of the original game. Higher order strategies that produce cooperation in the prisoner's dilemma, in effect, transform that game into a stag hunt game, where the choice to defect corresponds to the inefficient, but safer action (Skyrms, 2001). Therefore, many of the results of this chapter can be interpreted through the lens of the repeated prisoner's dilemma.

The choice between stag and hare simplifies a more complex reality in which agents might choose between multiple stags and a hare. By that we mean a society or a community would often have more than a single cooperative action to pursue. So, here, we extend the canonical two-by-two stag hunt game to allow for more actions. This increase in actions also provides sufficient space for learning rules to matter, whereas in a two-by-two game, how agents learn has no effect. As in the canonical stag hunt game, the models we consider include an insulated *self interested* action that does not require agents to coordinate. It generates only a modest payoff, but it is safe. We differ from the canonical model in that we allow for multiple potentially cooperative actions.

In our basic model, any cooperative action would be optimal if the population coordinated on it. We then expand our model to consider the possibility that some of these actions would be *effective*, e.g., successful if followed, but that others, termed *naive*, would be undermined by a *predatory* action.[2] Thus, we add predation (Conlisk, 2001) to the stag hunt. The predatory action can be seen as a form of defection. It robs those agents who undertake naive actions. This framework captures situations in which a community may have several potential "stag" actions of which some would prove fruitful and others would not. To illustrate, a community may have several mechanisms to share water, but some of these may prove exploitable by predatory actions.

Within this class of models, we find that cultural learning more often locates the efficient cooperative equilibria than does individual belief-based learning. To be precise, we show that cultural learning dominates individual learning: given any starting point for which individual learning results in cooperation, so does cultural learning.

Within the expanded framework that includes naive and predatory strategies, we highlight three additional results. First, contrary to intuition, we find that the effect of the learning rule becomes amplified as the stakes increase. As the stakes grow infinitely large, cultural learning converges to full cooperation, always locating an effective action, while belief-based learning converges to a zero probability of finding an effective action. Thus, ramping up incentives makes the type of learning rule more, not less important. Second, we find that as the number of potentially cooperative actions increases, so does the probability of finding the self interested action. This coordination failure aligns with basic intuition that lots of options makes

---

[2]Though the naive actions are not played in equilibrium, such dominated strategies play a crucial role in equilibrium selection in coordination games (Ochs, 1995; Basov, 2004).

coordinating on a single one more difficult. Third, and counter to intuition, given cultural learning we find that changing one action from naive to effective can move some initial points from the basin of attraction of a cooperative equilibrium into the basin of attraction of the safe equilibrium.

4.2 Individual and Cultural Learning

Given the variety of potential learning rules, we feel it necessary to motivate our decision to compare *individual, belief-based learning* to *cultural evolutionary learning* (Camerer 2003, Fudenberg and Levine 1999). These two learning rules differ in how they characterize behavior. Belief-based learning is prospective and individualistic; cultural evolutionary learning is retrospective and social. By considering these two extreme forms of learning we investigate the possibility that the type of learning rule might matter for equilibrium selection.

We model individual learning using a simple best response learning rule (Gilboa and Matsui, 1991; Hofbauer and Sigmund, 2003). Elaborated models of individual learning, such as logit learning and quantal response learning, include noise terms and individual errors. The extra degree of freedom introduced with this error term implies that they can fit experimental data better than the simpler best response dynamic. Nevertheless, we stick here with best response learning owing to its foundational nature and analytic tractability.

Following convention, we use replicator dynamics (Taylor and Jonker, 1978) to capture cultural learning in a population of players (Henrich and Boyd, 2002). Replicator dynamics can be seen as capturing situations in which agents compare payoffs with each other and copy better performing agents. [3] Cultural learning is less *greedy*

---

[3]Cultural learning can be performance-based if it depends on payoffs or conformist if it depends on popularity. Here, we consider performance-based cultural learning.

than best response dynamics. Agents don't move only in the direction of the best action but towards all actions that have above average payoffs.

For the sake of analytic tractability, we consider continuous time dynamics.[4] These would arise in the limit of a large, well-mixed population. In this limit, both dynamics can be derived as the expected behavior of agents with stochastic protocols for switching their actions (Sandholm, 2009). A simple revision protocol in which agents occasionally switch to the best response to the current population state generates the best response dynamics. Imitative revision protocols, such as *imitation driven by dissatisfaction* (Björnerstedt and Weibull, 1996) or *pairwise proportional imitation* (Schlag, 1998), lead to the replicator dynamics.

The differences between belief-based learning rules and cultural evolutionary learning have been the subject of substantial theoretical, experimental, and empirical investigation. For the most part, the theoretical literature focuses on how the rules operate and, in particular, on the stability of equilibria under the two types of rules. That literature shows that in many games both rules produce the same stable equilibria (Hopkins, 1999; Hofbauer et al., 1979; Hofbauer, 2000). Counterexamples rely on knife edge assumptions.

The experimental and empirical literatures attempt to flesh out which rule people apply in practice. As the two rules differ in their informational and cognitive requirements, we should expect each rule to be better suited to some environments than the other. Cultural learning requires knowledge of the success of others. Given that information, a cultural learning rule doesn't require much cognitive effort: agents need only copy someone doing better than they are. Best response learning, on the other hand, does not require any information about the success of others – other

---

[4]Showing that our main results hold with discrete dynamics as well is a straightforward, though somewhat involved, exercise.

than the payoffs to the game – but it does require knowledge of the full distribution of actions and calculation of the payoff from each possible action. Thus, we shouldn't expect to see best responses unless people understand the game fully.

In two-by-two games, we might therefore expect best response learning to better predict behavior. In fact, that is the case. Cheung and Friedman (1998) find greater support for belief-based learning than for replicator dynamics. However, looking across experiments reveals that the performance of these and other learning rules is often so similar as to be almost indistinguishable (Feltovich, 2000; Salmon, 2001). What differences that do exist between the behavior predicted by these rules and the data can often be explained by considering a hybrid model that includes both belief-based and reinforcement learning (Camerer and Ho, 1999).[5]

Learning rules have also been studied in the field. The use of real world data has both advantages and disadvantages. The stakes are often higher, and the situations are real, not manufactured. However, real world studies often suffer from problems of messier data, confounding factors, and higher dimensional action spaces. Given those caveats, evidence from the real world generally tilts towards cultural learning. Henrich (2001), in surveying evidence on the adoption of innovations, finds S-shaped adoption curves to be prevalent. Cultural learning, which relies on imitation, produces S-shaped curves. Individual learning does not. A hybrid learning model would also produce S-shaped adoption curves. Therefore, the empirical evidence should not lead us to declare cultural learning the winner so much as it tells us that people do take into account how others act.[6]

Our interests here tend less toward the empirical question of what people do and more in the direction of the theoretical question of what would happen if people were

---

[5]Reinforcement learning can also give rise to the replicator dynamics (Börgers and Sarin, 1997).

[6]*Prestige bias*, the coupling of prestige to success, is further evidence that imitation is a part of how people learn (Henrich and Gil-White, 2001).

to follow one rule at the expense of the other. We find that cultural learning proves far better able to locate an efficient equilibrium than does best response learning, and this could imply that societies that have a more collective orientation might be better equipped to coordinate their efforts and cooperate in the face of strategic uncertainty.

We do not mean to imply that cultural learning is the only mechanism through which a society can achieve efficient, coordinated cooperation. Many other mechanisms have been shown sufficient, including focal points (Schelling, 1960) and, in the context of an evolutionary dynamic, cheap talk (Skyrms, 2004). Alternatively, if groups or bands of people engage in battles with neighboring groups, then group selection could produce altruistic strategies that ensure cooperation (Gintis et al., 2003). Even though altruism may not be individually rational, groups with more altruists may be stronger than groups of self interested actors and therefore more likely to emerge victorious from conflict.

In what follows, we introduce our model of a stag hunt game with multiple cooperative actions, provide an example, and prove some general results. We then modify our model to allow some of the cooperative actions to fail and be undermined by a predatory action. We present the bulk of our results, analyzing this model. We conclude with a discussion of the types of learning we consider.

4.3 The Basic Model

In our basic model, we assume a *self interested* action that offers agents a risk-free return. We also assume *n cooperative* actions. Each offers the agent a reward that is assuredly positive and increases with the number of others taking the same action. In the canonical example, hunting rabbits would be a self interested action. It can

be done alone just as well as with others. Hunting stags would be a cooperative action, as would participating in a collective whale hunt. In these cases, hunting cooperatively greatly increases the chances of a hearty meal, while hunting alone for a stag or a whale gives the player a small (but nonzero) chance at finding something to eat. (Perhaps while being faithful to the larger cause, the hunter can still bag a small animal or fish on the side.) Taking a cooperative action has *positive spillovers* in that it helps others who join in. Thus, taking a cooperative action becomes the better choice (only) when a significant percentage of other agents also take this action.

Formally, our game consists of $n+1$ actions, with the action set $\mathcal{A} = \{1, \cdots, n, S\}$. We denote the set of cooperative actions $\mathcal{C} = \{1, \cdots, n\}$. We consider a single unit-mass population of agents. The state space is therefore the n-dimensional unit simplex $\triangle^n$, with a point $\mathbf{x} \in \triangle^n$ denoting the fraction of the population choosing each action. The vector $\mathbf{x} = (x_1, \ldots, x_n, x_S)$ is the population mixed strategy.

Payoffs in a population game are a function of the population state $\mathbf{x}$. We normalize payoff magnitudes by attributing a cooperative action taken in isolation a payoff of 1, assuming it's the same for any cooperative action, and we let the parameter $\beta > 0$ capture the relative benefit of coordinating on a cooperative action. Larger β imply greater benefits from achieving cooperation. The cooperative actions vary in their efficiency according to a family of parameters, θ_i for each $i \in \mathcal{C}$. The maximum reward, if everybody coordinates on effective action i, is an additional $\theta_i \beta$. Finally, the insulated, self interested action S receives a payoff of $c > 1$ regardless of the actions of the other agents. The payoffs can therefore be written as follows:

$$\pi(i, \mathbf{x}) = 1 + \theta_i \beta x_i \text{ for } i \in \mathcal{C}$$

$$\pi(S, \mathbf{x}) = c > 1.$$

We make a technical assumption, (A1) $\theta_i\beta + 1 > c$ for all $i \in \mathcal{C}$, to create the proper ordering over payoffs. This guarantees that successfully coordinating on a cooperative action yields a higher payoff than taking the self interested action. Thus, we have a strict pure Nash Equilibrium at every action in the game.

In what follows we compare two learning rules: *best response dynamics* and *replicator dynamics*. In the continuous time best response dynamics, some infinitesimal proportion of the agents are always switching their action to match the current best response.

Best Response Dynamics *(Individual Learning)* $\dot{\mathbf{x}} \in \text{BR}(\mathbf{x}) - \mathbf{x}$ where $\text{BR}(\mathbf{x})$ is the set of best replies to $\mathbf{x}$.

The second learning rule, *replicator dynamics*, describes agents who learn from the success of others (Henrich and Boyd 2002).

Replicator Dynamics *(Cultural Learning)* $\dot{x}_i = x_i(\pi_i - \bar{\pi})$ where π_i is the payoff to action i and $\bar{\pi}$ is the average payoff.

4.3.1 An Example

To show how to apply these learning rules, we create a simple example with $n = 2$, meaning there are three total actions, including the self interested one. We take $c = 3$ and $\theta_1\beta = \theta_2\beta = 5$. This game has three pure strategy equilibria, one for each action. Each equilibrium is strict and is therefore asymptotically stable for both learning rules. Even in this simple game, we see that replicator dynamics has larger basins for the cooperative actions.

Following convention, we can represent any distribution of actions as a point in the two-dimensional simplex $\triangle^2$. To locate the basins of attraction under best response dynamics, we identify the regions of the simplex $\triangle^2$ in which each action is a best
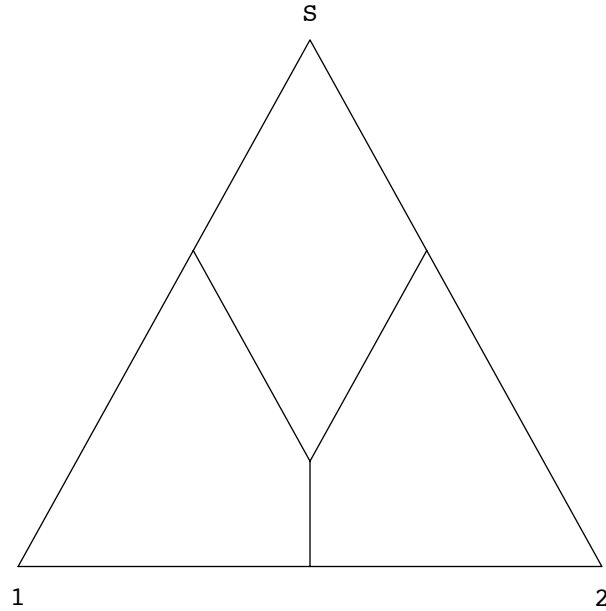
Figure 4.1: Basins of attraction under best response dynamics

response. These regions are defined by the lines where each pair of actions earns the same payoff. We find $\pi_1 = \pi_2$ when $x_1 = x_2$, $\pi_1 = \pi_S$ when $x_1 = \frac{2}{5}$, and $\pi_2 = \pi_S$ when $x_2 = \frac{2}{5}$. Because the payoff to a cooperative action increases as the action spreads, the equilibrium chosen under best response dynamics consists of the action that is initially a best response. This is a feature of our basic model, but it will not be true of our modified model. It means that the best response regions are the basins of attraction of the pure equilibria under best response dynamics. They are shown in Figure 4.1. The corresponding flow diagram for the best response dynamics is shown in Figure 4.2.

Figure 4.3 contains the flow diagram for the replicator dynamics, and then, in Figure 4.4, we characterize the basins of attraction for replicator dynamics. Here, the boundary separating the basins of attraction includes curves, not only lines. This curvature arises because under replicator dynamics a cooperative action can grow in the population even if the self interested action is the best response. As
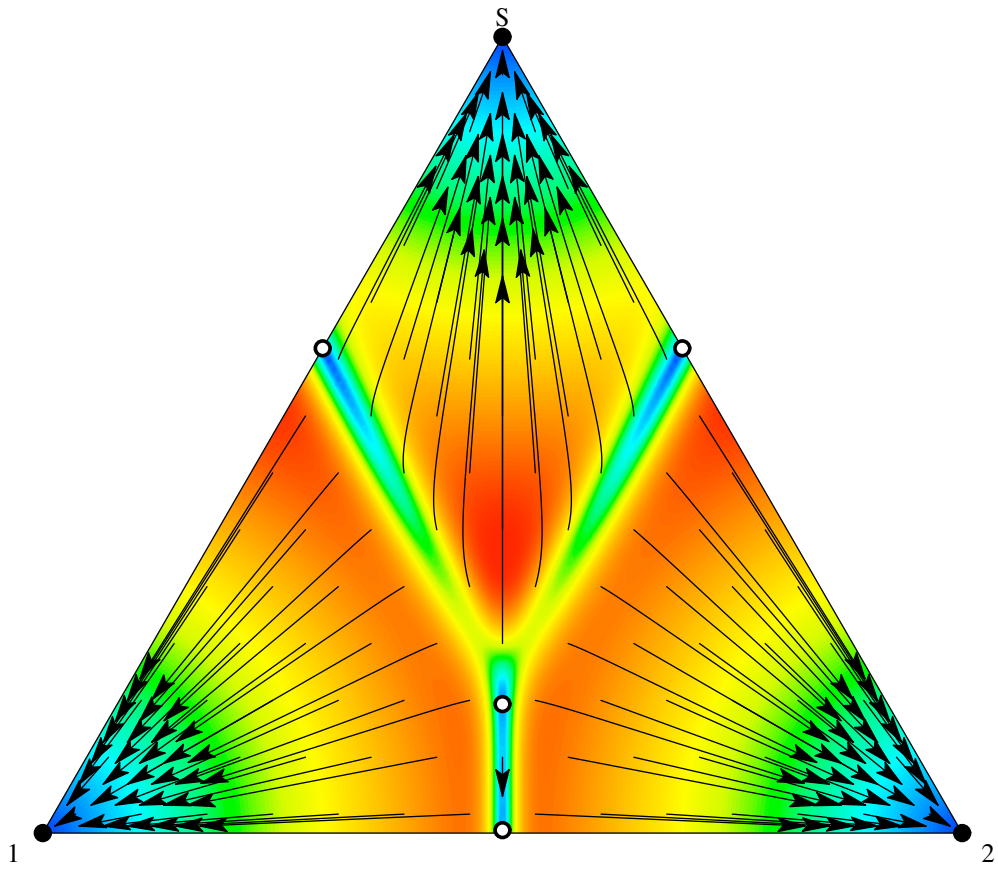
Figure 4.2: Phase diagram for the best response dynamics. Black (white) circles are stable (unstable) rest points. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).
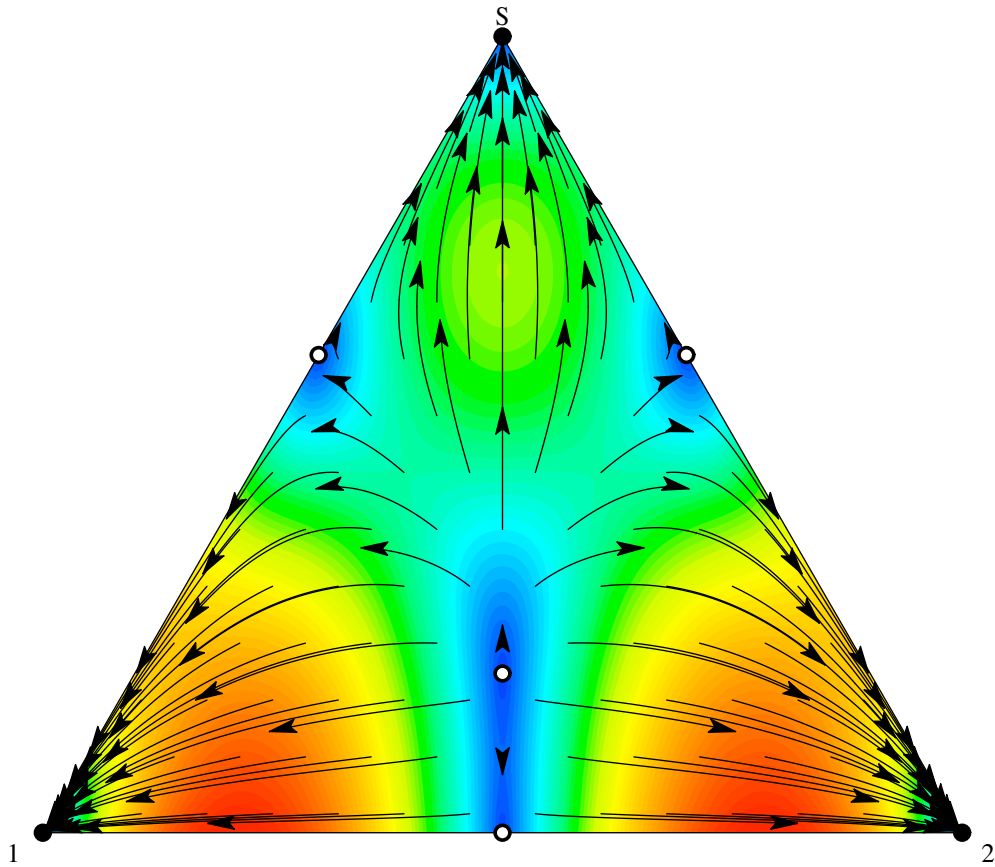
Figure 4.3: Phase diagram for the replicator dynamics. Black (white) circles are stable (unstable) rest points. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).

it grows, the cooperative action can then become the best response. As a result, the population can slip from the self interested action's best response region into a cooperative action's best response region. Thus, cooperation is more likely to arise under replicator dynamics.

4.3.2 General Results

We now show that the intuition developed in our example holds more generally. That is, the replicator dynamics is more likely to lead to cooperation, while the best response dynamics is more likely to lead to the worst outcome. In what follows, we assume an initial distribution of actions with full support, bounded density, and no
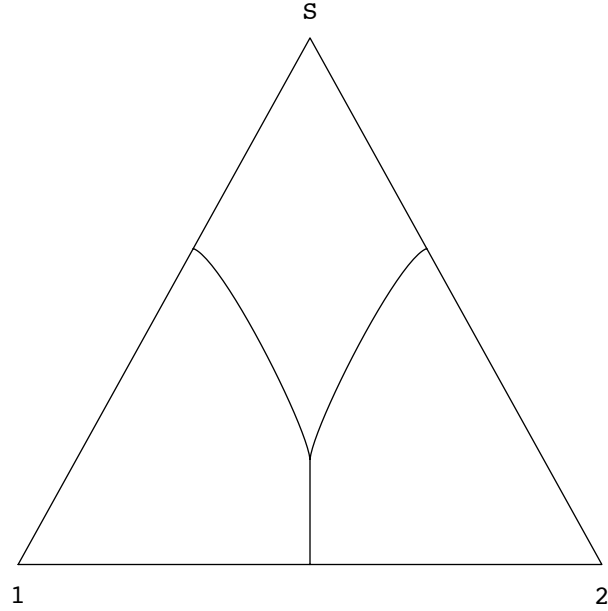
Figure 4.4: Basins of attraction under replicator dynamics

mass points on the boundary of the strategy space, and then apply two learning rules: best response dynamics and replicator dynamics. We first show for both learning rules that if a cooperative action ever has the best payoff, it remains best forever.

Lemma IV.1. *For both best response dynamics and replicator dynamics, if for some $i \in \mathcal{C}$, $\pi_i > \pi_j$ for all $j \neq i \in \mathcal{A}$ at some time, it remains so at all later times.*

Proof. For best response dynamics, the result is straightforward. Only a best response grows in the population. If a cooperative action is a best response, it becomes more widespread, and consequently its payoff increases. Meanwhile, other actions become less common, so the payoffs to the other cooperative actions decrease.

For replicator dynamics, suppose $\pi_i \geq \pi_j$ for all $j \in \mathcal{A}$. For $j \in \mathcal{C}$,

$$(4.1) \qquad \dot{\pi}_j = \theta_j \beta \dot{x}_j = \theta_j \beta x_j (\pi_j - \bar{\pi}) = (\pi_j - 1)(\pi_j - \bar{\pi}).$$

So $\dot{\pi}_i \geq \dot{\pi}_j$ for all $j \in \mathcal{C}$, and $\dot{\pi}_i > 0$. The payoff to the cooperative action which is the best response increases faster than the payoff to other cooperative actions. The

self interested action has a constant payoff, so it cannot become the best response. Cooperative action i must remain a best response forever. $\qquad\square$

Our first theorem identifies the basins of attraction under best response dynamics. In an abuse of notation, we refer to the equilibrium with everyone choosing cooperative action $i \in \mathcal{C}$, $(x_i = 1, x_{-i} = 0)$ as *cooperative equilibrium* $i \in \mathcal{C}$. Similarly, we use the term *self interested equilibrium* to stand for the equilibrium with everyone choosing self interested action, $(x_S = 1, x_{-S} = 0)$. For clarity in our presentation, we define some new parameters that help us compare the payoffs of a cooperative action and the self interested action. Let $T_i = \frac{c-1}{\theta_i \beta}$ be the threshold frequency for cooperative action $i \in \mathcal{C}$ to be better than the self interested action. That is, $\pi_i \geq \pi_S$ if and only if $x_i \geq T_i$, with equality in one following from equality in the other.

Theorem IV.2. *Given best response dynamics, a point $\mathbf{x}$ is in the basin of attraction of the self interested equilibrium if and only if for all $j \in \mathcal{C}$, $x_j < T_j$.*

A point $\mathbf{x}$ is in the basin of attraction of cooperative equilibrium $i \in \mathcal{C}$ if and only if $x_i > T_i$ and $x_i > \frac{\theta_j}{\theta_i} x_j$ for all $j \neq i \in \mathcal{C}$.

Proof. The inequalities given in the theorem define the best response regions. Lemma IV.1 tells us that the best response regions of cooperative actions are contained in the basins of attraction of cooperative equilibria. Similarly under best response dynamics, if the self interested action is a best response, cooperative actions decline in the population and cooperative payoffs decrease. The self interested action thus remains a best response, and the dynamic leads to the self interested equilibrium. $\qquad\square$

The next theorem states that the basins of attraction of the cooperative equilibria under best response dynamics are proper subsets of these basins under replicator dynamics.

Theorem IV.3. *Assume $n \geq 2$. The basin of attraction of any cooperative equilibrium $i \in \mathcal{C}$ under best response dynamics is a proper subset of the basin of attraction of this equilibrium under replicator dynamics.*

Proof. Lemma IV.1 and Theorem IV.2 together imply that the basin of attraction of a cooperative equilibrium under best response dynamics is contained in the basin under replicator dynamics because the former consists only of points for which the cooperative action is the best response. It remains to show that the basin of attraction of any cooperative equilibrium $i \in \mathcal{C}$ under replicator dynamics includes some points for which the self interested action is the best response. This is done in the appendix. □

Theorem IV.3 tells us that when there are multiple cooperative actions, replicator dynamics is more likely to select one than best response dynamics. (When there is just one cooperative action, the learning rules behave alike, as they do in all two-by-two games. Chapter III describes this general result.) Seeing that cultural learning outperforms belief-based learning in this environment, we next consider the question of how much better it can be.

4.4   A Model with a Predatory Strategy

We now modify our model by introducing a predatory strategy and making some of the cooperative strategies susceptible to it. This predatory strategy will temporarily be a best response, but it is not an equilibrium. We maintain the n potentially cooperative actions, but now we assume that only k of these cooperative actions are *effective*. These actions have positive spillovers and are immune from predation. The others we call *naive* because they are susceptible to attack from a *predatory* action. In the context of the literal stag hunt, consider a scenario in which players have the

option of domesticating animals and sharing the burden of looking after them. This is a potentially cooperative action, but in the case that nobody can be trusted to exercise costly vigilance while watching over somebody else's animals, it is naive. The animals may be stolen.

The predatory action can be thought of as a strong form of defecting. An agent who takes the predatory action benefits at the expense of those attempting naive actions. Neither the predatory action nor the self interested action aids the cooperative efforts of other players, but unlike the predatory action, the self interested one does not undermine the success of anybody who is trying to cooperate. An example clarifies this distinction. Hunting rabbits instead of stags indirectly lowers the payoffs to the stag hunters by denying them another participant. But, a player who steals the aforementioned domesticated animals is taking a *predatory* action. This directly harms those taking taking naive actions and creates no additional surplus.

This modified game consists of $n + 2$ actions: a *predatory* action, a *self interested* action, k *effective* actions, and $n - k$ *naive* actions. We assume $n > k \geq 1$ ensuring that both the set of effective actions and the set of naive actions are nonempty.

The **action set** $\mathcal{A} = \{1, \cdots, n, P, S\}$ *with partition* $f : \{1, \cdots, n\} \rightarrow \{E, N\}$ *where* $\mathcal{E} = \{i | f(i) = E\}$ *denotes the* **effective** *actions and* $\mathcal{N} = \{i | f(i) = N\}$ *denotes the* **naive** *actions.*

Note that $k = |\mathcal{E}|$. Effective actions $i \in \mathcal{E}$ have the same payoff as in the basic model when all cooperative actions were effective. Naive actions $i \in \mathcal{N}$ differ in their payoff structure. They each get a negative payoff per agent playing the predatory action. The parameter γ measures the value that could be lost to or gained by predation. If the population were to tend towards 100% predatory action, the payoff deducted

from each naive action would be γ. The payoffs lost by naive actions are gained by agents taking the predatory action. This predatory action P gets nothing from everybody else, so its payoff scales with the prevalence of naive actions, approaching γ as the population tends toward 100% naive actions. The payoffs from the various actions can be written as follows:

$$\pi(i, \mathbf{x}) = 1 + \theta_i \beta x_i \text{ for } i \in \mathcal{E}$$

$$\pi(i, \mathbf{x}) = 1 + \theta_i \beta x_i - \gamma x_P \text{ for } i \in \mathcal{N}$$

$$\pi(P, \mathbf{x}) = \gamma \sum_{i \in \mathcal{N}} x_i$$

$$\pi(S, \mathbf{x}) = c > 1.$$

Assumption A1 from the basic model still holds for all $i \in \mathcal{E}$, but not necessarily for $i \in \mathcal{N}$. We now have (A1') $\theta_i \beta + 1 > c$ for all $i \in \mathcal{E}$. We make two additional technical assumptions, (A2) $\gamma > c$ and (A3) $\gamma > \max_{i \in \mathcal{N}} \{\theta_i\} \beta \frac{c}{c-1}$. A2 guarantees that in a population full of naive actions, the predatory action has a higher payoff then the self interested action. A3 guarantees that a naive action is never best, i.e., a predator can steal even more than the positive spillovers generated by a naive action.

4.5 Results

We now turn to our main results: how the equilibrium attained depends strongly on the learning rule. We begin with an obvious theorem about the set of pure strategy equilibria to this class of games.

Theorem IV.4. *Given assumptions A1' and A3 and $c > 1$, this game has $k + 1$ pure strategy equilibria: one in which all players take the self interested action and k equilibria in which they all take the same effective action.*

We maintain our abuse of notation in referring to one of these k *effective equilibria*

as effective equilibrium $i \in \mathcal{E}$. We now show that Lemma IV.1 still applies to the effective cooperative actions under best response dynamics, and can be broadened to include the self interested action as well. However, we will need to slightly modify the lemma for the replicator dynamics. For the best response dynamics, we can say that if an effective action or the self interested action ever has the best payoff, it remains best forever. This is not the case for the predatory action. (By assumption A3, a naive action is never best.)

Lemma IV.5. *For best response dynamics, if for some $i \in \mathcal{E} \cup \{S\}$, $\pi_i > \pi_j$ for all $j \neq i \in \mathcal{A}$ at some time, it remains so at all later times.*

Proof. Note that naive actions are dominated by a mixed strategy that plays the self interested action with probability $\frac{1}{c}$ and the predatory action with probability $\frac{c-1}{c}$. So, naive actions never have the highest payoff and are always decreasing. This means the payoff to the predatory action is always decreasing. So, exempting the naive actions, an effective action has the only increasing payoff when it is the strict best response, and the self interested action has the only nondecreasing payoff when it is the strict best response. $\qquad\square$

Our next result identifies a sufficient condition for replicator dynamics to yield effective cooperation. We do not have a similar sufficient condition for fixation of the self interested action. This lemma is similar to Lemma IV.1 in that it applies to an effective action that at some time is better than the self interested action and has the most positive spillovers, but we do not require it to be better than the predatory action. Of particular importance is the fact that this condition is independent of the predation parameter γ.

Lemma IV.6. *Under the replicator dynamics, if for some $i \in \mathcal{E}$,*

1. $\pi_i > \pi_S$ and

2. $x_i > \frac{\theta_j}{\theta_i} x_j$ for all $j \neq i \in \mathcal{E} \cup \mathcal{N}$

at some time, then conditions 1 and 2 remain in effect at all later times and the replicator dynamics leads to the equilibrium $(x_i = 1, x_{-i} = 0)$.

Proof. A piece of the average payoff, $x_P \pi_P + \sum_{j \in \mathcal{N}} x_j \pi_j$, partially cancels, leaving only $\sum_{j \in \mathcal{N}} (1 + \theta_j \beta x_j) x_j$. Thus, the average payoff $\bar{\pi} = x_S \pi_S + \sum_{j \in \mathcal{E} \cup \mathcal{N}} (1 + \theta_j \beta x_j) x_j$. Now, conditions 1 and 2 together imply that $\pi_i > \pi_j$ for all $j \in \mathcal{A} \setminus \{i, P\}$, and in turn,

$$(4.2) \qquad \qquad \pi_i - \bar{\pi} > 0.$$

Condition 2 alone implies that

$$(4.3) \qquad \qquad \pi_i - \bar{\pi} > \pi_j - \bar{\pi} \text{ for all } j \neq i \in \mathcal{E} \cup \mathcal{N}.$$

Inequality (4.3) gives action i the highest relative growth rate in $\mathcal{E} \cup \mathcal{N}$, ensuring that condition 2 continues to hold. Inequality (4.2) means that action i does indeed have positive growth, maintaining condition 1 and leading to the equilibrium $(x_i = 1, x_{-i} = 0)$. $\qquad \square$

We now derive the basins of attraction under best response dynamics. We will make use of a new parameter $r(\mathbf{x})$ that denotes the ratio of the predatory payoff to the self interested payoff. (It depends on the prevalence of naive actions.) We have $r(\mathbf{x}) = \frac{1}{c} \gamma \sum_{j \in \mathcal{N}} x_j$. Also, recall that $T_i = \frac{c-1}{\theta_i \beta}$, now for $i \in \mathcal{E}$, is the threshold frequency for effective action i to be better than the self interested action.

Theorem IV.7. *Given best response dynamics, a point $\mathbf{x}$ is in the basin of attraction of the self interested equilibrium if and only if for all $j \in \mathcal{E}$,*

$$(4.4) \qquad \qquad x_j < T_j$$

or

(4.5) $$x_j < r(\mathbf{x})T_j.$$

A point $\mathbf{x}$ is in the basin of attraction of effective equilibrium $i \in \mathcal{E}$ if and only if the following three conditions hold:

i) $x_i > T_i$;

ii) $x_i > r(\mathbf{x})T_i$; and

iii) $x_i > \frac{\theta_j}{\theta_i}x_j$ for all $j \neq i \in \mathcal{E}$. [7]

Proof. See appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Conditions (i) and (iii) in Theorem IV.7, namely that an effective action does better than the self interested action and all other effective actions, prove to be necessary but not on their own sufficient for best response dynamics to attain the equilibrium featuring this effective cooperative action. On the other hand, inequality (4.4), which says that the self interested action initially does better than any effective action, is sufficient (and not even necessary) for the best response dynamics to lead to universal self interested action.

The next claim states that cultural learning more often achieves effective cooperation. *Whenever best response dynamics attains an effective equilibrium, replicator dynamics attains it as well.* The converse will not be true. Thus, cultural learning is strictly preferred to individual learning.

Theorem IV.8. *The basin of attraction of any effective equilibrium $i \in \mathcal{E}$ under best response dynamics is contained in the basin of attraction of this equilibrium under replicator dynamics.*

---

[7]In the zero probability event that an equality holds exactly and neither set of conditions applies, the best response dynamics finds a mixed equilibrium.

Proof. Using Lemma IV.6 and Theorem IV.7, we show that the necessary conditions for the best response dynamics to be in the basin of $(x_i = 1, x_{-i} = 0)$ for some $i \in \mathcal{E}$ are sufficient conditions for the replicator dynamics to be in the basin of this equilibrium. Condition (i) in Theorem IV.7 is equivalent to condition 1 in Lemma IV.6. Condition (ii) in the theorem along with assumption A3 implies that $x_i > \max_{j \in \mathcal{N}} \frac{\{\theta_j\}}{\theta_i} \sum_{l \in \mathcal{N}} x_l$, by plugging the latter inequality into the former. This easily gives us $x_i > \frac{\theta_j}{\theta_i} x_j$ for all $j \in \mathcal{N}$. Condition (iii) in Theorem IV.7 fills in for all $j \neq i \in \mathcal{E}$ and thus satisfies condition 2 in Lemma IV.6. $\qquad \square$

Just like Theorem IV.3 for the basic model, Theorem IV.8 tells us that replicator dynamics is more likely to produce effective cooperation than best response dynamics when there is a predatory strategy. In fact, the possibility of predation makes outcomes significantly more sensitive to the learning dynamic operating in the population. In Section 4.6.1 we will show how much of a deciding factor the learning style can be, but as a preview of these results, we present Figures 4.5 and 4.6, which show sample trajectories of the two learning dynamics for a game with just one effective action and one naive action. The images have a three-dimensional perspective because the strategy space for a game with four actions is the three-dimensional simplex $\triangle^3$. The parameter values ($c = 2$, $\gamma = 100$, $\beta = 10$, $\theta_E = \theta_N = 1$[8]) are inspired by our upcoming results. As we can see in Figure 4.5, the best response dynamics often approach the predatory action at first, before flowing directly to the self interested action; although not shown, some points near the effective action would flow to that equilibrium as well. In Figure 4.6, we see that the replicator dynamics often curves towards the effective action as the naive action decays; again, some trajectories (not shown) in the neighborhood of the self interested action would reach that

---

[8]We denote the single effective action by E and the naive action by N.
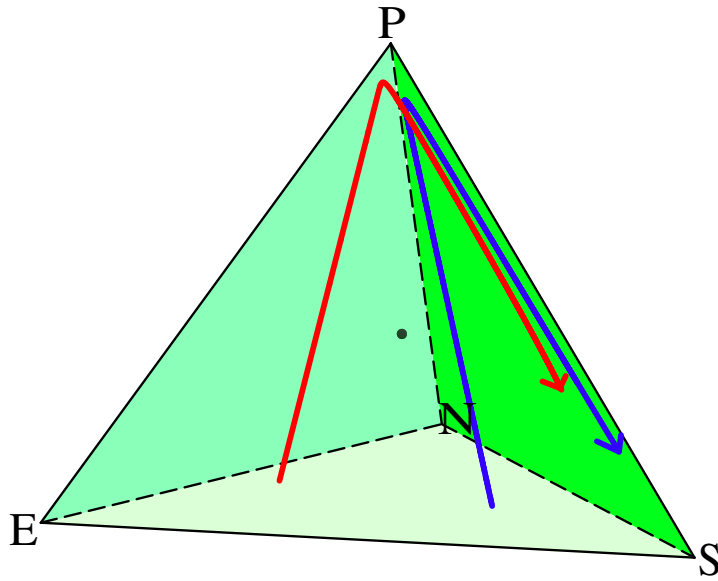
Figure 4.5: Sample trajectories for the best response dynamics in a game with just one effective action and one naive action, setting $c = 2$, $\gamma = 100$, $\beta = 10$, $\theta_E = \theta_N = 1$. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).
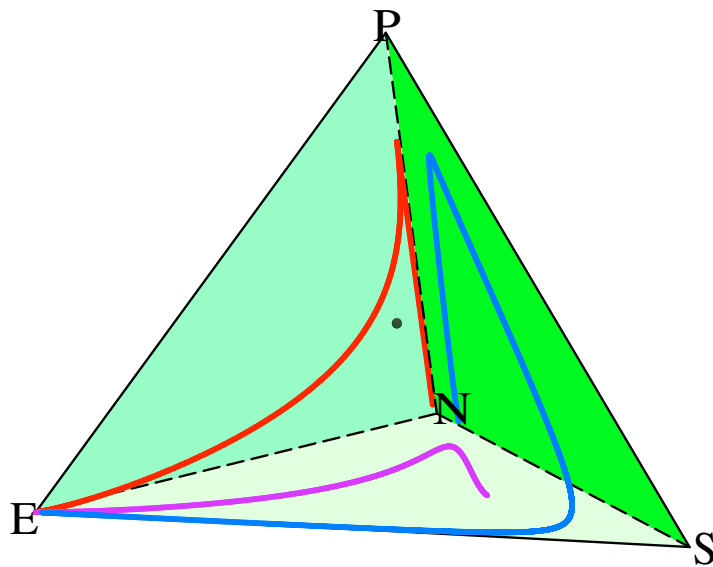
Figure 4.6: Sample trajectories for the replicator dynamics in a game with just one effective action and one naive action, setting $c = 2$, $\gamma = 100$, $\beta = 10$, $\theta_E = \theta_N = 1$. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).

equilibrium, too.

At this point, we have precisely determined the basins of attraction under best response dynamics, but we have only a partial characterization of basins under the replicator dynamics. Because the replicator dynamics are nonlinear, we cannot solve for the basins exactly. In Section 4.6, we will make a simplifying assumption that allows us to proceed without an exact description of the replicator dynamics' basins. We would like to make analysis of the replicator dynamics more tractable without losing generality from our results for the best response dynamics. To see how this is possible, let us refer back to Theorem IV.7, which characterizes the best response dynamics' basins of attraction. Observe that Theorem IV.7 makes no reference to the naive actions' efficiency parameters. As long as the $\{\theta_j\}_{j \in \mathcal{N}}$ obey assumption A3, they have no effect on the best response dynamics. These parameters do influence the replicator dynamics, but it appears that the behavior of the replicator dynamics is qualitatively similar for a range of possible values. We show in Figures 4.7 and 4.8 sample trajectories of the replicator dynamics while varying θ_N and retaining the other parameter values used in Figure 4.6.

4.6 Comparative Statics

We now consider the case that naive actions fail to produce positive spillovers for others taking the same action. In what follows, we assume $\theta_j = 0$ for all $j \in \mathcal{N}$. This gives each of the naive actions the same payoff,

$$\pi(i, \mathbf{x}) = 1 - \gamma x_P \text{ for } i \in \mathcal{N}.$$

While we may still refer to naive actions as potentially cooperative, in this framework they are not actually cooperative in the sense that they do not create positive externalities. As we have indicated, results for the best response dynamics in this
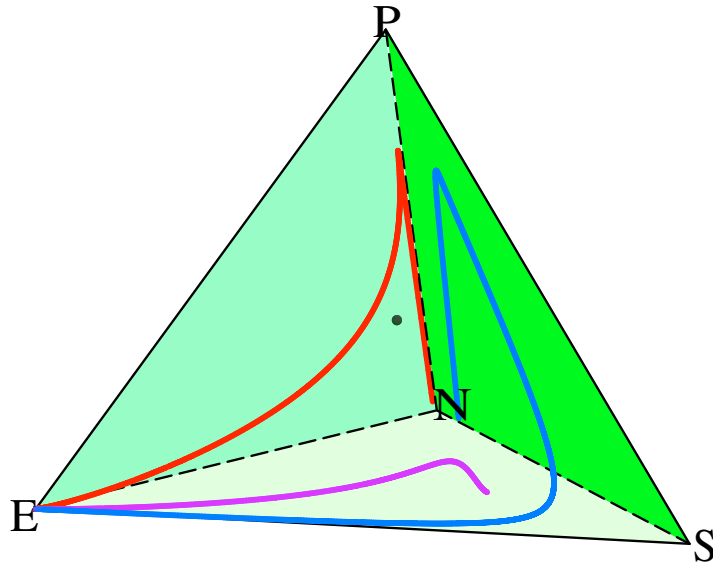
Figure 4.7: Sample trajectories for the replicator dynamics, setting $c = 2$, $\gamma = 100$, $\beta = 10$, $\theta_E = 1$, $\theta_N = 0$. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).
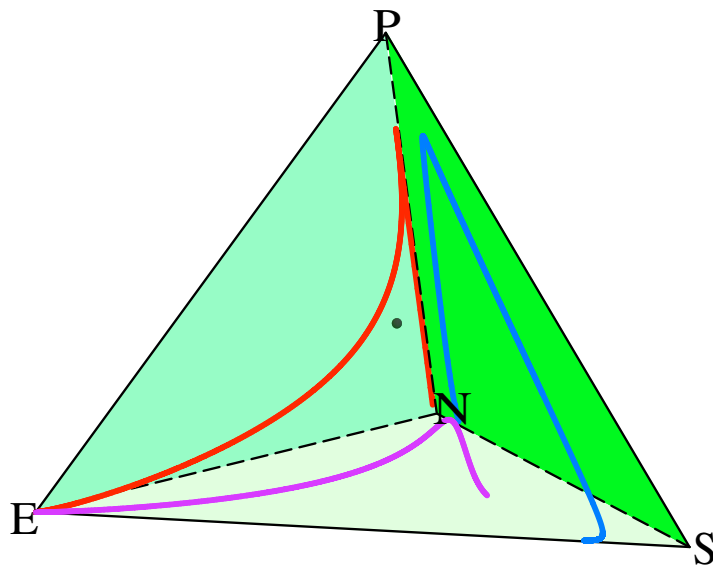
Figure 4.8: Sample trajectories for the replicator dynamics, setting $c = 2$, $\gamma = 100$, $\beta = 10$, $\theta_E = 1$, $\theta_N = 2$. Figure made by the game dynamics simulation program Dynamo (Sandholm and Dokumaci, 2007).

section would still hold for nonzero $\{\theta_j\}_{j \in \mathcal{N}}$, but with our assumption analysis of the replicator dynamics is more tractable.

4.6.1 Changing the Incentives to Coordinate and Predate

The next two results consider the effects of increasing the stakes by ramping up both the incentives to coordinate on an effective action and the benefits of predation. The result in the first claim aligns with the intuition that increasing incentives increases the probability of an efficient outcome. It states that as β, the benefit from coordinating on an effective action, grows large, replicator dynamics almost always achieves coordinated, effective cooperation. However, the cooperative action taken may not be the most efficient one. Since the result follows from an application of Lemma IV.6, it holds regardless of whether the incentives to predate are large as well.

The second claim states that under best response dynamics, as the benefits to predation grow, the basin of attraction of the self interested action expands to the entire space. This holds even if β goes to infinity as well, so long as the benefits of predation are sufficiently larger.

Together, these two results imply that as the stakes rise, we need not worry about the initial distribution of strategies. As long as an effective action is played occasionally, no matter how rarely, a population using replicator dynamics will learn to coordinate effective cooperation. And as long as naive actions cannot be ruled out entirely, if the incentive to predate rises faster than the incentive to cooperate, then a population using best response dynamics will learn to take safe, self interested action.

Theorem IV.9. *As $\beta \to \infty$, the basins of attraction of the equilibria featuring*

only effective actions $(x_S = 0, \sum_{i \in \mathcal{E}} x_i = 1)$ *approach the entire strategy space under replicator dynamics. This holds even if $\gamma \to \infty$ as well.*

Proof. Let $m = \arg\max_j(\theta_j x_j)$ be the set of effective actions with highest payoff. By the logic that proved Lemma IV.1 and the fact that equation (4.1) still holds for our modified model, this set of best effective actions remains constant over time. If for $i \in m$, $x_i > T_i$, then Lemma IV.6 applies in the case that $|m| = 1$ and a straightforward extension of it applies when multiple effective actions tie for the highest payoff. In short, the replicator dynamics flow to an equilibrium that satisfies $(\sum_{i \in m} x_i = 1, x_j = 0 : j \notin m)$. As $\beta \to \infty$, every $T_i \to 0$ and the set of points satisfying $x_i > T_i$ for $i \in m$ approaches the entire strategy space. $\qquad\square$

Theorem IV.10. *Under best response dynamics the basin of attraction of the self interested equilibrium monotonically increases in γ.*

Proof. As γ increases, the condition $x_j < T_j \frac{\gamma}{c} \sum_{l \in \mathcal{N}} x_l$ for all $j \in \mathcal{E}$ is satisfied for more initial points. By Theorem IV.7, if this condition is met, the best response dynamics flow to $(x_S = 1, x_{-S} = 0)$. $\qquad\square$

Corollary IV.11. *As $\gamma \to \infty$, the basin of attraction of $(x_S = 1, x_{-S} = 0)$ approaches the entire strategy space under best response dynamics. This conclusion holds even if $\beta \to \infty$ as well, as long as $\frac{\gamma}{\beta} \to \infty$.*

Proof. Points with $\sum_{i \in \mathcal{N}} x_i > \frac{c\theta_j\beta}{(c-1)\gamma}$ (and thus, $r(\mathbf{x})T_j > 1$) for all $j \in \mathcal{E}$ satisfy inequality (4.5) and are in the basin of attraction of $(x_S = 1, x_{-S} = 0)$ given best response dynamics. So, as $\gamma \to \infty$ faster than β, the basin of attraction of $(x_S = 1, x_{-S} = 0)$ approaches the entire strategy space. $\qquad\square$

In the case where β and $\frac{\gamma}{\beta}$ approach infinity, Theorem IV.9 and Corollary IV.11 taken together show that best response dynamics and replicator dynamics predict en-

tirely different outcomes. Replicator dynamics *always* achieves effective cooperation. Best response dynamics *never* does.

4.6.2 Changing the Number of Effective Actions

Up to this point, we have taken the partitioning function f as exogenous and examined the effects of varying payoff parameters on the basins of attraction. In practice, some stag hunt games may have many effective actions, whereas others may have only a few. To analyze the effect of relatively easy and hard coordination problems, we now examine the effect of adding more potentially cooperative actions to the game or changing how many of those actions are effective. The following claim states that when the self interested action is sufficiently attractive and there are many potentially cooperative actions, both learning dynamics likely find the inefficient outcome, in which the agents all take the self interested action. Note that the condition we place on the payoff to the self interested action is only sufficient and could be weakened.

Theorem IV.12. *Assume $c > 3$. As $n \to \infty$, the fraction of the strategy space in the basin of attraction of the self interested equilibrium grows to 1 under both best response and replicator dynamics.*

Proof. See appendix. $\qquad \square$

Now, we fix the payoff parameters and the number of potentially cooperative actions and compare basins of attraction across different partition functions. Given partition f_i, let $B(f_i)$ be the union of the basins of attraction of the equilibria featuring only effective actions $(x_S = 0, \sum_{j \in \mathcal{E}_i} x_j = 1)$.

Definition IV.13. Partition f_2 *effectively contains* f_1 if $\mathcal{E}_1 \subset \mathcal{E}_2$ and for all $j \in \mathcal{E}_1$, θ_j is the same for both partitions.

If f_2 effectively contains f_1, then $k_2 > k_1$. That is, more of the n potentially cooperative actions are effective with partition f_2 than with f_1. We might expect that making more of the potentially cooperative actions effective would improve the probability of locating such an equilibrium. And, indeed, the basins of attraction of these good equilibria do grow under best response dynamics as more actions are made effective, as the next claim states.

Theorem IV.14. *If f_2 effectively contains f_1, then under best response dynamics, any initial point that results in coordination on an effective action under partition f_1 also does under partition f_2, i.e., $B(f_1) \subset B(f_2)$.*

Proof. From Theorem IV.7 we know that under best response dynamics a point is in $B(f_i)$ if and only if for some $l \in \mathcal{E}_i$, (i) $x_l > T_l$ and (ii) $x_l > T_l \frac{\gamma}{c} \sum_{j \in \mathcal{N}_i} x_j$. If these inequalities are satisfied for a given x_l under partition f_1, they must still be satisfied for x_l under f_2 because (i) is unchanged and (ii) is weaker because $\mathcal{N}_2 \subset \mathcal{N}_1$. Moreover, there are additional actions in $\mathcal{E}_2$ for which these inequalities may be satisfied. So $B(f_1) \subset B(f_2)$. $\qquad\square$

Surprisingly, a similar result does not hold for replicator dynamics. Increasing the number of effective actions creates a *crowding effect*. It can raise the average payoff in the population and therefore prevent any effective action from growing in the population.

Theorem IV.15. *If f_2 effectively contains f_1, then under replicator dynamics there can exist initial points that lead to coordination on an effective action under partition f_1 that do not under partition f_2, i.e., $B(f_1) \backslash B(f_2)$ need not be empty.*

Proof. See appendix. $\qquad\square$

In interpreting these last two claims, we must keep in mind our earlier result that the basin of attraction for effective actions is always larger under replicator dynamics. Thus, when we change one cooperative action from naive to effective, any initial points that we move out of the basin of attraction of an effective action and into the basin of attraction of the self interested action under replicator dynamics must remain in the basin of attraction of the self interested action under best response dynamics as well.

4.7   Discussion

In this chapter, we have shown that for a class of generalized stag hunt games, cultural learning dominates individual learning: any initial condition for which individual, belief-based learning achieves effective cooperation necessarily leads to effective cooperation under replicator dynamics too. Moreover, we have shown that as the stakes grow large, cultural learning, as captured by replicator dynamics, almost always achieves coordinated, effective cooperation. In contrast, individualistic, belief-based learning captured by best response dynamics almost never does.

These dichotomous limiting results are clearly a product of our specification. Nevertheless, the core intuition holds generally: cultural learning allows moderately successful attempts at coordination to gain a foothold and potentially grow in the population, while best response learning only rewards the best action, which rarely is a nascent attempt at cooperation.[9] It is the greediness of the best response dynamics that distinguishes its behavior from the replicator dynamics. In general, we would expect the greediness of a learning dynamic to be the key factor in determining outcomes here, even more so than whether the underlying revision protocol is imitative

---

[9]Given that as a general rule, small changes in payoffs have correspondingly small effects on the basins of attraction, we can expect that slight variations in the payoff structure of our model have only mild effects on our results.

or direct.[10]

Our findings have several implications for institutional design. Most importantly, they reveal the importance of learning rules for equilibrium selection. How agents learn may not seem to be a choice variable, and yet, each learning style requires information about the payoffs or actions of others. That information can be made more or less available through institutional choices. Second, our results show that ramping up incentives may not be sufficient to produce the desired equilibrium. Finally, our finding that increasing the number of effective cooperative actions can move some initial conditions into the basin of an inefficient equilibrium demonstrates the costs of abundance. More effective actions can produce a crowding effect.

It would be surprising if learning rules did not differ across societies. Recent work has found that behavior in a common experimental setting varies widely across cultures and that some of that variation can be explained by features of those cultures (Henrich et al., 2001). For example, cultures that engage in collective enterprises, like whale hunting, appear more likely to share. These findings do not prove that distinct cultures learn differently, but they are consistent with such an assumption.

Relatedly, a substantial body of survey and case study research shows that cultures vary in their levels of individualism and collectivism (Inglehart, 1997). In more collectivist societies, people may have richer social networks giving them better knowledge of the actions of others and the payoffs of those actions. This suggests a possible link between collectivism and cultural learning. Comparably, in individualistic societies, people may be less informed about others and more concerned with acting optimally. These characteristics would point to best response learning. Seeing that, in our framework, cultural learning induces cooperation, which could reinforce

---

[10]Sandholm (2009) defines a direct revision protocol as one in which a revising agent's choice of an alternative action to consider is without regard to its popularity.

underlying collectivist sentiment, it is conceivable that cooperation, collectivism, and cultural learning form a positive feedback loop that could amplify initial differences between societies.

While there is logic to the idea that more collectivist societies might have an easier time mustering cooperation, levels of collectivism and individualism do not independently determine how societies learn. Recall from the introduction the different informational and cognitive requirements of the two learning rules. Best response learning requires greater cognitive effort than cultural learning. The members of a collectivist society with strong attachment to rationality could indeed use best response learning. And members of an individualistic society might turn to cultural learning when a game becomes too complicated to think through. A surprising feature of our results is that additional cognitive effort could be counterproductive for the society if it means belief-based learning prevails over cultural learning in contexts such as ours.

Our results also have implications for the evolution of learning styles. Stag hunt games are common, and societies that use cultural learning will have greater success in them. Group selection, either cultural or genetic, could promote cultural learning over belief-based learning. This hypothesis requires further study into the mechanism through which societies adopt new learning styles (see Henrich and McElreath, 2003).

Overall, we advocate a modest application of our results. Rather than arguing over which learning rule better describes human behavior or even which rule fits particular cultures, social scientists might better explore *how* members of different societies and cultures learn. Learning may well include transference of behaviors learned in one context to other contexts (Bednar and Page 2007). We can then explore whether those differences in learning processes can result in significant differences

in outcomes. If they do, we can try to design the games created by economic and political institutions so that they produce better outcomes.

We conclude by reiterating the observation that the type of learning rule in use matters most when a model admits multiple equilibria. Except in those cases where learning rules do not attain an equilibrium (Hofbauer and Swinkels, 1996), single equilibrium models leave little room for how people learn, or for culture more generally, to have any effect. Thus, if we want to understand the implications of variation in learning rules, we need to consider games with more than two strategies.

4.8   Appendix

Completing the Proof of Theorem IV.3.

We now identify points for which the self interested action is the best response, but for which the replicator dynamics leads to the cooperative equilibrium ($x_i = 1, x_{-i} = 0$). Consider points of the form $x_i = \frac{c-1-\epsilon}{\theta_i \beta}$, $x_l = \frac{c-1}{2\theta_l \beta}$ for some other $l \in \mathcal{C}$, and $x_j < \frac{\theta_i}{\theta_j} x_i$ for all $j \neq i \in \mathcal{C}$. Assume ϵ is small. Such points have been chosen so that $\pi_i = c - \epsilon$ and $\pi_l = \frac{c+1}{2}$. The self interested action is the best response at this point, but cooperative action i is very close and better than all other cooperative actions. Plugging in for π_i and parts of $\bar{\pi}$, we have

$$
\begin{aligned}
\dot{x}_i &= x_i \left((c - \epsilon)(1 - x_i) - c x_S - \sum_{j \neq i \in \mathcal{C}} \pi_j x_j \right) \\
&= x_i \left(-\epsilon(1 - x_i) + \sum_{j \neq i \in \mathcal{C}} (c - \pi_j) x_j \right).
\end{aligned}
$$

The second step here used the fact that $\sum_{j \in \mathcal{A}} x_j = 1$. Dropping some positive terms from the right hand side, we get the inequality $\dot{x}_i > x_i ((c - \pi_l) x_l - \epsilon)$. Plugging in for π_l and simplifying, we have

$$
(4.6) \qquad\qquad \dot{x}_i > x_i \left(\frac{c-1}{2} x_l - \epsilon \right).
$$

As long as ϵ is small, the right hand side of inequality (4.6) is positive and x_i increases past $\frac{c-1}{\theta_i \beta}$ right away. This makes cooperative action i a best response, and then Lemma IV.1 applies and we know the dynamics leads to the cooperative equilibrium $(x_i = 1, x_{-i} = 0)$. $\qquad \square$

Proof of Theorem IV.7.

First, we point out that if for some $i \in \mathcal{E}$, $x_i > T_i$ and $x_i < r(\mathbf{x})T_i$, then $x_i < \frac{1}{\theta_i \beta} \left(\gamma \sum_{l \in \mathcal{N}} x_l - 1 \right)$. (Actually, if equality holds in one but not both of the conditions, we still obtain the desired inequality.) This fact follows just from rearranging terms, using a fair bit of basic algebra. It means that if we are relying on inequality (4.5) to establish that a point is in the basin of the self interested equilibrium (i.e., when inequality (4.4) fails and some effective action is initially better than the self interested action), then inequality (4.5) ensures that the predatory action is initially better than this effective action. Alternatively, if we are establishing that a point is in the basin of attraction of an effective equilibrium, we can say that in order for an effective action to be initially better than the self interested action and the predatory action, condition (ii) must hold.

Now, assume best response dynamics. We will invoke Lemma IV.5, which tells us that if either the self interested action or an effective action is initially a best response or becomes one, it remains a best response forever, so the best response dynamics flow towards the equilibrium featuring this action.

For effective action i to initially be the best response, conditions (i) and (iii) obviously must hold and the need for condition (ii) is described in the proof's first paragraph. For the self interested action S to initially be the best response, inequality (4.4) is clearly necessary.

If neither the self interested action nor any of the effective actions are initially a

best response, then the predatory action P must initially be the best response. In this case, the equilibrium attained depends on which action next becomes the best response. So, let us now consider points for which the predatory action P is the best response. If for all $j \in \mathcal{E}$, inequality (4.4) holds and $x_j < T_j$, then the self interested action S has a higher payoff than any of the effective actions. As the dynamic moves toward P, the payoffs to the predatory and effective actions decrease, so eventually S becomes the best response. Alternatively, suppose for some $i \in \mathcal{E}$, condition (i) holds and $x_i > T_i$. Then we define $\alpha(t) = \frac{x_j(t)}{x_j(0)}$ for $j \neq P$ and t such that action P is still a best response. This definition is independent of j because actions which are not best responses have the same relative decay rate. Note that $\alpha(t)$ is a strictly decreasing function. Now either

$$(4.7) \qquad\qquad \pi_S = \pi_P \text{ when } \alpha = \frac{c}{\gamma \sum_{l \in \mathcal{N}} x_l(0)}$$

or for some $i \in \mathcal{E}$,

$$(4.8) \qquad\qquad \pi_i = \pi_P \text{ when } \alpha = \frac{1}{\left(\gamma \sum_{l \in \mathcal{N}} x_l(0) - \theta_i \beta x_i(0)\right)},$$

whichever happens first. Equation (4.7) follows from $\pi_P = \gamma \sum_{l \in \mathcal{N}} \alpha(t) x_l(0)$. Equation (4.8) depends on this as well as on $\pi_i = 1 + \theta_i \beta \alpha(t) x_i(0)$ for $i \in \mathcal{E}$. If inequality (4.5) applies, i.e., if for all $j \in \mathcal{E}$, $x_j(0) < \frac{(c-1)\gamma}{c\theta_j \beta} \sum_{l \in \mathcal{N}} x_l(0)$, then rearranging terms produces

$$\frac{c}{\gamma \sum_{l \in \mathcal{N}} x_l(0)} > \frac{1}{\left(\gamma \sum_{l \in \mathcal{N}} x_l(0) - \theta_j \beta x_j(0)\right)},$$

and this means action S eventually becomes the best response. On the other hand, if for some $i \in \mathcal{E}$, $x_i(0) > \frac{(c-1)\gamma}{c\theta_i \beta} \sum_{l \in \mathcal{N}} x_l(0)$ and $x_i(0) > \frac{\theta_j}{\theta_i} x_j(0)$ for all $j \neq i \in \mathcal{E}$, conditions (ii) and (iii) respectively, then action i always has the highest payoff of all the effective actions and becomes the best response before the self interested action does. $\qquad \square$

Proof of Theorem IV.12.

Let $M\{*\}$ denote the fraction of the strategy space satisfying $\{*\}$. As $n \to \infty$, $M\{x_j < T_j$ for all $j \in \mathcal{E}\}$ approaches 1. By Theorem IV.7, all points satisfying this condition are in the basin of attraction of $(x_S = 1, x_{-S} = 0)$ with the best response dynamics.

Assume replicator dynamics. A sufficient condition to be in this basin of attraction is at some time t,

$$(4.9) \qquad\qquad cx_S(t) > 1 + \theta_j \beta x_j(t) \text{ for all } j \in \mathcal{E},$$

as this ensures that $\pi_j < \bar{\pi}$ then and at all future times. An alternative condition is that

$$(4.10) \qquad\qquad \bar{\pi} - \sum_{i \in \mathcal{N}} x_i \geq 2\pi_j - 1 \text{ for all } j \in \mathcal{E}$$

at some time. Inequality (4.9) is self enforcing because it ensures that x_S increases while π_j decreases for all $j \in \mathcal{E}$. To see that inequality (4.10) is self enforcing is slightly more involved. It too ensures that π_j decreases for all $j \in \mathcal{E}$, but now we must take a time derivative of $\bar{\pi} - \sum_{i \in \mathcal{N}} x_i$ and show that it is positive. We get

$$
\begin{aligned}
\dot{\bar{\pi}} - \sum_{i \in \mathcal{N}} \dot{x}_i &= \sum_{l \in \mathcal{A}} \dot{x}_l \pi_l + x_l \dot{\pi}_l - \sum_{i \in \mathcal{N}} \dot{x}_i \\
&= \sum_{l \in \mathcal{A}} \dot{x}_l (\pi_l - \bar{\pi}) + x_l \dot{\pi}_l - \sum_{i \in \mathcal{N}} \dot{x}_i.
\end{aligned}
$$

The last step here uses $\sum_{l \in \mathcal{A}} \dot{x}_l = 0$. We can write $\dot{x}_l(\pi_l - \bar{\pi})$ as $x_l(\pi_l - \bar{\pi})^2$. For $l \in \mathcal{E}$, $x_l \dot{\pi}_l = x_l(\pi_l - 1)(\pi_l - \bar{\pi})$, and $x_l(\pi_l - \bar{\pi})^2 > x_l(\pi_l - 1)(\bar{\pi} - \pi_l)$ by inequality (4.10) itself. So

$$(4.11) \qquad\qquad \sum_{l \in \mathcal{E}} \dot{x}_l(\pi_l - \bar{\pi}) + x_l \dot{\pi}_l > 0.$$

We can plug in the payoffs for the predatory and the naive actions and as a shortcut use $\sum_{l \in \{P\} \cup \mathcal{N}} x_l \pi_l = \sum_{i \in \mathcal{N}} x_i$, taking the time derivative of both sides, to write

$$\sum_{l \in \{P\} \cup \mathcal{N}} \dot{x}_l (\pi_l - \bar{\pi}) + x_l \dot{\pi}_l = \sum_{l \in \{P\} \cup \mathcal{N}} \dot{x}_l (-\bar{\pi}) + \sum_{i \in \mathcal{N}} \dot{x}_i.$$

We find that $\sum_{l \in \{P\} \cup \mathcal{N}} \dot{x}_l \leq 0$ as long as $\bar{\pi} \geq 1$ because $\sum_{l \in \mathcal{N}} \dot{x}_l = \sum_{l \in \mathcal{N}} x_l (1 - \gamma x_P - \bar{\pi})$ and $\dot{x}_P = x_P \left(\sum_{l \in \mathcal{N}} \gamma x_l - \bar{\pi} \right)$ imply that

$$\sum_{l \in \{P\} \cup \mathcal{N}} \dot{x}_l = \sum_{l \in \mathcal{N}} x_l (1 - \bar{\pi}) + x_P (-\bar{\pi}).$$

And we know $\bar{\pi} \geq 1$ from inequality (4.10). So

(4.12)
$$\sum_{l \in \{P\} \cup \mathcal{N}} \dot{x}_l (\pi_l - \bar{\pi}) + x_l \dot{\pi}_l - \sum_{i \in \mathcal{N}} \dot{x}_i > 0.$$

Finally, $x_S (\pi_S - \bar{\pi})^2$ is clearly positive and $\dot{\pi}_S = 0$, so

(4.13)
$$\dot{x}_S (\pi_S - \bar{\pi}) + x_S \dot{\pi}_S > 0.$$

Thus, piecing together inequalities (4.11), (4.12) and (4.13), we get $\dot{\bar{\pi}} - \sum_{i \in \mathcal{N}} \dot{x}_i > 0$.

Let j be a best effective action. We can place an upper bound on the rate at which action j spreads, $\dot{x}_j = x_j (1 + x_j \theta_j \beta - \bar{\pi}) < x_j (1 + x_j \theta_j \beta)$. This bound has the form of a logistic differential equation. The solution is then bounded by the logistic function,

(4.14)
$$x_j(t) \leq \frac{x_j(0)}{(1 + \theta_j \beta x_j(0)) e^{-t} - \theta_j \beta x_j(0)}.$$

Because inequality (4.10) is sufficient for the replicator dynamics to flow to $(x_S = 1, x_{-S} = 0)$, we consider an assumption that $\bar{\pi} < 2\pi_j - 1 + \sum_{i \in \mathcal{N}} x_i \leq 2\pi_j$. This allows us to place a lower bound on the rate at which action S spreads, $\dot{x}_S = x_S (c - \bar{\pi}) > x_S (c - 2 - 2\theta_j \beta x_j)$. Then, plugging in inequality (4.14) for $x_j(t)$ and integrating,

$$x_S(t) \geq x_S(0) e^{(c-2)t} (1 + \theta_j \beta x_j(0)(1 - e^t))^2.$$

Applying this to our first sufficient condition, inequality (4.9), it is sufficient to show

$$cx_S(0)e^{(c-2)t}(1 + \theta_j\beta x_j(0)(1 - e^t))^2 > 1 + \frac{\theta_j\beta x_j(0)}{(1 + \theta_j\beta x_j(0))e^{-t} - \theta_j\beta x_j(0)}$$

or equivalently,

(4.15) $$cx_S(0)e^{(c-2)t}(1 + \theta_j\beta x_j(0)(1 - e^t))^3 > 1 + \theta_j\beta x_j(0).$$

This last step requires some algebra. The left hand side of (4.15) is maximized at

$$t = \ln\left(\frac{1 + \theta_j\beta x_j(0)}{\theta_j\beta x_j(0)\frac{c+1}{c-2}}\right).$$

Plugging in for t in (4.15), the sufficient condition becomes

$$cx_S(0)\left(\frac{(1 + \theta_j\beta x_j(0))(c - 2)}{\theta_j\beta x_j(0)(c + 1)}\right)^{c-2}\left(1 + \theta_j\beta x_j(0)\left(1 - \frac{(1 + \theta_j\beta x_j(0))(c - 2)}{\theta_j\beta x_j(0)(c + 1)}\right)\right)^3$$

$$> 1 + \theta_j\beta x_j(0).$$

As $n \to \infty$, $x_j(0)$ becomes small, so we keep only terms of lowest order in $x_j(0)$. This simplifies our sufficient condition to

$$cx_S(0)\left(\frac{c - 2}{\theta_j\beta x_j(0)(c + 1)}\right)^{c-2}\left(\frac{3}{c + 1}\right)^3 > 1.$$

It remains only to show that this condition is met almost everywhere when n is large.

Our sufficient condition holds if

(4.16) $$x_j(0) \le \frac{1}{n^{\frac{c}{3(c-2)}}} \quad \text{and} \quad x_S(0) > \frac{1}{c}\left(\frac{\theta_j\beta(c + 1)}{c - 2}\right)^{c-2}\left(\frac{c + 1}{3}\right)^3\frac{1}{n^{\frac{c}{3}}}.$$

Because these two inequalities are positively correlated, [11]

$$M\{\text{Constraint (4.16)}\} \ge$$

$$M\left\{x_j(0) \le \frac{1}{n^{\frac{c}{3(c-2)}}}\right\} \cdot M\left\{x_S(0) > \frac{1}{c}\left(\frac{\theta_j\beta(c + 1)}{c - 2}\right)^{c-2}\left(\frac{c + 1}{3}\right)^3\frac{1}{n^{\frac{c}{3}}}\right\}.$$

---

[11] Recall that $M\{*\}$ denotes the fraction of the strategy space satisfying $\{*\}$.

Consider the first of these inequalities. We have

$$M\left\{x_j(0) > \frac{1}{n^{\frac{c}{3(c-2)}}}\right\} \leq M\left\{x_i(0) > \frac{1}{n^{\frac{c}{3(c-2)}}} \text{ for some } i \in \mathcal{E} \cup \mathcal{N}\right\}$$

$$\leq n \cdot M\left\{x_1(0) > \frac{1}{n^{\frac{c}{3(c-2)}}}\right\}$$

$$= n\left(1 - \frac{1}{n^{\frac{c}{3(c-2)}}}\right)^{n+1}.$$

Here, and again in equation (4.17), we evaluate the fraction of the strategy space satisfying a given inequality simply by integrating over the strategy space. Now, $\frac{c}{3(c-2)} < 1$ because we assumed $c > 3$, so

$$\lim_{n\to\infty} n\left(1 - \frac{1}{n^{\frac{c}{3(c-2)}}}\right)^{n+1} = 0.$$

Thus,

$$\lim_{n\to\infty} M\left\{x_j(0) \leq \frac{1}{n^{\frac{c}{3(c-2)}}}\right\} = 1.$$

Now consider the second inequality. We have

$$(4.17) \quad M\left\{x_S(0) > \frac{1}{c}\left(\frac{\theta_j\beta(c+1)}{c-2}\right)^{c-2}\left(\frac{c+1}{3}\right)^3\frac{1}{n^{\frac{c}{3}}}\right\}$$

$$= \left[1 - \frac{1}{c}\left(\frac{\theta_j\beta(c+1)}{c-2}\right)^{c-2}\left(\frac{c+1}{3}\right)^3\frac{1}{n^{\frac{c}{3}}}\right]^{n+1}.$$

This approaches 1 as $n \to \infty$ because $\frac{c}{3} > 1$. Thus,

$$\lim_{n\to\infty} M\left\{cx_S(0)\left(\frac{c-2}{\theta_j\beta x_j(0)(c+1)}\right)^{c-2}\left(\frac{3}{c+1}\right)^3 > 1\right\} = 1.$$

The fraction of the strategy space satisfying a condition that puts it in the basin of attraction of $(x_S = 1, x_{-S} = 0)$ approaches 1. $\qquad\square$

Proof of Theorem IV.15.

We construct a specific counterexample for the case $n = 3$ that can be extended to a more general case. Let $\mathcal{E}_1 = \{1\}$, $\mathcal{E}_2 = \{1,2\}$, $\theta_1 = 1$, $\theta_2 = 1$ under partition f_2

(whereas, of course, $\theta_2 = 0$ under partition f_1), and $\beta > 2c$. Then any point which satisfies

$$x_1 = x_2 = \frac{c-1}{\beta} - \epsilon, \; x_S = \frac{c - \beta\epsilon}{c}\left(1 + 2\epsilon - 2\frac{c-1}{\beta}\right)$$

for small enough ϵ will be in $B(f_1)$ but not in $B(f_2)$.

Consider partition f_1. Recall that $\dot{x}_1 = x_1(\pi_1 - \bar{\pi})$. By construction $\pi_1 = (c - \beta\epsilon)$ and still $\pi_S = c$. Plugging in and simplifying, we get the average payoff

$$\bar{\pi} = (c - \beta\epsilon)\left(1 - \frac{c-1}{\beta} + \epsilon\right) + \frac{c-1}{\beta} - \epsilon + x_3.$$

We combine terms and find that at our initial point, $\pi_1 - \bar{\pi} = \beta\left(\frac{c-1}{\beta} - \epsilon\right)^2 - x_3$. Therefore, initially, $\dot{x}_1 = \beta\left(\frac{c-1}{\beta} - \epsilon\right)^3 - \left(\frac{c-1}{\beta} - \epsilon\right)x_3$. From the fact that $\sum_{j \in \mathcal{A}} x_j = 1$, we know our initial point satisfies $x_3 \leq \frac{\beta\epsilon}{c}\left(1 + 2\epsilon - 2\frac{c-1}{\beta}\right)$. This gives us a minimum initial value for $\dot{x}_1$,

$$(4.18) \qquad \dot{x}_1 \geq \beta\left(\frac{c-1}{\beta} - \epsilon\right)^3 - \left(\frac{c-1}{\beta} - \epsilon\right)\frac{\beta\epsilon}{c}\left(1 + 2\epsilon - 2\frac{c-1}{\beta}\right).$$

Observe that the right hand side of (4.18) has a positive leading order term with no ϵ dependence. As ϵ is small, x_1 soon grows larger than $\frac{c-1}{\beta}$. By Lemma IV.6, the point must be in $B(f_1)$.

Now consider partition f_2. The average payoff is larger with this partition. In fact, $\pi_1 = \pi_2 = c - \beta\epsilon$, and it turns out $\bar{\pi} = c - \beta\epsilon + x_3$ at our initial point. This means that initially $\pi_1 = \pi_2 = \bar{\pi} - x_3$. We will now see that the state of

$$(4.19) \qquad\qquad\qquad \pi_1 = \pi_2 \leq \bar{\pi} - x_3$$

must persist forever because it is self enforcing. Note that $x_1 = x_2$ and $\pi_1 = \pi_2$ at all times by the symmetry of their initial conditions. We can plug in $\bar{\pi} - x_3 = cx_S + 2\pi_i x_i$ with $i \in \{1, 2\}$ and then rewrite equation (4.19) as

$$(4.20) \qquad\qquad\qquad \pi_i\left(1 - 2x_i\right) \leq cx_S.$$

We now compare time derivatives of both sides of inequality (4.20) and show $c\dot{x}_S > \dot{\pi}_i(1 - 2x_i) + \pi_i(-2\dot{x}_i)$. In particular, $\dot{\pi}_i < 0$ because of inequality (4.19) itself, and clearly $1 - 2x_i > 0$. It remains to show $c\dot{x}_S + 2\pi_i\dot{x}_i \geq 0$. We have $c > \pi_i > 0$, so it is sufficient to show $\dot{x}_S + 2\dot{x}_i \geq 0$. And, recognizing that $\sum_{j \in \mathcal{A}} \dot{x}_j = 0$, it is fine to show $\dot{x}_3 + \dot{x}_P \leq 0$. In the proof of Theorem IV.12, we show this will be negative as long as $\bar{\pi} > 1$. We know $\bar{\pi} > cx_S$, and we know x_S has been increasing because $c > \bar{\pi}$. (To check this last inequality, just examine the formula for the average payoff and recall that we have already argued that $c > \pi_i$ in our persistent state.) Finally, we obtain

$$
\begin{aligned}
cx_S(0) &= (c - \beta\epsilon)\left(1 + 2\epsilon - 2\frac{c-1}{\beta}\right) \\
&= c - \frac{2c}{\beta}(c - 1) + O(\epsilon) > 1,
\end{aligned}
$$

using in the last step the facts that ϵ is small and $2c < \beta$. Because the average payoff always remains above the payoff to either of the effective actions, the effective actions become rarer, and it follows that the initial point is not in $B(f_2)$. $\square$

BIBLIOGRAPHY

[1] Abreu, D. , Pearce, D., Stacchetti, E., 1990. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica* 58, 1041-1063.

[2] Axelrod, R., 1984. *The Evolution of Cooperation.* New York: Basic Books.

[3] Basov, S., 2004. Equilibrium selection in coordination games: Why do dominated strategies matter? *Economics Bulletin* 3, 1-3.

[4] Bednar, J., Page, S.E., 2007. Can Game(S) Theory Explain Culture? *Rationality and Society* 19(1) 65-97. forthcoming.

[5] Bendor, J., Swistak, P., 1997. The Evolutionary Stability of Cooperation. *American Political Science Review* 91(2), 290-307.

[6] Björnerstedt, J., Weibull, J.W., 1996. Nash Equilibrium and Evolution by Imitation. *The Rational Foundations of Economic Behavior.* Arrow, K. J., et. al., 155-181. New York: St. Martins Press.

[7] Börgers, T., Sarin, R., 1997. Learning Through Reinforcement and Replicator Dynamics. *Journal of Economic Theory* 77, 1-14.

[8] Camerer, C., 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton: Princeton University Press.

[9] Camerer, C., Ho, T.H., 1999. Experience Weighted Attraction Learning in Normal Form Games. *Econometrica* 67, 827-874.

[10] Cheung, Y.-W, Friedman, D., 1998. A Comparison of Learning and Replicator Dynamics Using Experimental Data. *Journal of Economic Behavior and Organization* 35 (3), 263-280.

[11] Conlisk, J., 2001. Costly Predation and the Distribution of Competence. *The American Economic Review* 91 (3), 475-484.

[12] Epstein, J., 2003. *Generative Social Science.* Princeton: Princeton University Press.

[13] Feltovich, N., 2000. Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games. *Econometrica* 68, 605-641.

[14] Foster, D., Young, H.P., 1990. Stochastic evolutionary game dynamics. *Theoretical Population Biology* 38, 219-232.

[15] Fudenberg, D., Levine, D., 1999. *The Theory of Learning in Games.* Cambridge: MIT Press.

[16] Gilboa, I., Matsui, A., 1991. Social Stability and Equilibrium. *Econometrica* 59, 859-867.

[17] Gintis, H., Bowles, S., Boyd, R., Fehr, E., 2003. Explaining Altruistic Behavior in Humans. *Evolution and Human Behavior* 24, 153-172.

[18] Golman, R., Page, S., 2008. Basins of Attraction and Equilibrium Selection Under Different Learning Rules. *Journal of Evolutionary Economics* forthcoming.

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

[20] Henrich, J. 2001. Cultural Transmission and the Diffusion of Innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change and much of sociocultural evolution. *American Anthropologist*, 103: 992-1013.

[21] Henrich, J., Boyd, R., 2002. On Modeling Cognition and Culture: Why Cultural Evolution Does Not Require Replication of Representations. *Journal of Cognition and Culture* 2(2), 87-112.

[22] Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review* 91, 73-78.

[23] Henrich, J., Gil-White, F., 2001. The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission. *Evolution and Human Behavior* 22, 165-196.

[24] Henrich, J., McElreath, R., 2003. The Evolution of Cultural Evolution. *Evolutionary Anthropology* 12, 123-135.

[25] Hofbauer, J., 2000. From Nash and Brown to Maynard Smith: Equilibria, Dynamics and ESS. *Selection* 1, 81-88.

[26] Hofbauer, J., Schuster, P., Sigmund, K., 1979. A Note on Evolutionary Stable Strategies and Game Dynamics. *Journal of Theoretical Biology* 81, 609-612.

[27] Hofbauer, J., Sigmund, K., 2003. Evolutionary Game Dynamics. *Bulletin of the American Mathematical Society* 40, 479-519.

[28] Hofbauer, J., Swinkels, J.M., 1996. A Universal Shapley Example. Unpublished manuscript, University of Vienna and Northwestern University.

[29] Hopkins, E., 1999. A Note on Best Response Dynamics. *Games and Economic Behavior* 29, 138-150.

[30] Inglehart, R., 1997. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies.* Princeton: Princeton University Press.

[31] Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29-56.

[32] Kandori, M., Rob, R., 1995. Evolution of equilibria in the long run: a general theory and applications. *Journal of Economic Theory* 65, 383-414.

[33] Medina, L., 2007. *A Unified Theory of Collective Action and Social Change.* Ann Arbor: University of Michigan Press.

[34] Ochs, J., 1995. Coordination Problems. *Handbook of Experimental Economics.* Kagel, J., Roth, A. Princeton: Princeton University Press.

[35] Rousseau, J., 1773. *A Discourse on Inequality.*

[36] Salmon, T., 2001. An Evaluation of Econometric Models of Adaptive Learning. *Econometrica* 69, 1597-1628.

[37] Sandholm, W. H., 2009. Pairwise Comparison Dynamics and Evolutionary Foundations for Nash Equilibrium, working paper, University of Wisconsin. http://www.ssc.wisc.edu/~whs.

[38] Sandholm, W.H., Dokumaci, E., 2007. *Dynamo: Phase Diagrams for Evolutionary Dynamics.* Software suite. http://www.ssc.wisc.edu/~whs/dynamo.

[39] Schelling, T., 1960. *The Strategy of Conflict.* Harvard University Press.

[40] Schlag, K. H., 1998. Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-armed Bandits. *Journal of Economic Theory* 78, 130-156.

[41] Skyrms, B., 2001. The Stag Hunt. *Proceedings and Addresses of the American Philosophical Association* 75(2), 31-41.

[42] Skyrms, B., 2004. *The Stag Hunt and the Evolution of Social Structure.* Cambridge University Press.

[43] Taylor, P.D., Jonker, L., 1978. Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences* 40, 145-156.

[44] Young, H.P., 1993. The evolution of conventions. *Econometrica* 61, 57-84.

CHAPTER V

Why Learning Doesn't Add Up: Equilibrium Selection and Compositions of Learning Rules

5.1 Introduction

Models of individual learning in games often produce the same dynamical systems as population-based evolutionary adjustment processes. For example, an appropriate form of reinforcement learning leads to the replicator dynamics (Borgers and Sarin 1997), originally proposed by theoretical biologists to describe the growth of a haploid population (Taylor and Jonker 1978). Similarly, the best response dynamics (Gilboa and Matsui 1991) can arise as the continuous time limit of a belief-based learning rule like fictitious play or in a population of myopic best responders. The models that lead to the replicator dynamics and those that lead to the best response dynamics make mutually incompatible assumptions about how strategies are updated, offering no middle ground.

In this chapter, we adopt a population-based framework and propose three ways to combine best response and replicator dynamics into tractable composite rules. We find that a population is sensitive to the presence of more than one learning style, as well as to how the learning rules are put together, in that composite learning dynamics may have basins of attraction that share little overlap with each other or with the homogeneous learning dynamics. We focus especially on a straightforward

linear combination of best response and replicator learning rules, which we show to select an equilibrium not attained by homogeneous use of either pure rule in certain matrix games. We then show for specific classes of coordination games that introducing even arbitrarily small weight on an alternative learning rule causes the resulting linear combination of rules to locate a different equilibrium than the pre-existing pure rule. A common feature of the classes of games we consider is the existence of a temporary initial best response – an action that looks good at first, but that cannot survive against itself. The temptation to try such an action instead of an equilibrium action creates complex dynamics, which allow the effects of subtle differences in learning styles to build up over time.

Categorization of individual learning rules as belief-based or reinforcement learning reflects an idealization that people follow all-or-nothing rules. It makes sense that people consider both what has been successful in the past and what might be a good response to the expected opponent's play. The breakthrough of experience weighted attraction (Camerer and Ho 1999), a learning model with strong experimental support, is to combine belief-based and reinforcement learning. The combination is not simply a weighted average though. While experience weighted attraction does reduce to these particular learning models as special cases, it is nonlinear in the parameters that distinguish them. This allows experience weighted attraction to capture the best parts of both models. It also means we have multiple free parameters and a complicated formula that often hinders theoretical analysis. By adopting the population learning framework in this chapter, we try to capture the spirit of experience weighted attraction while retaining analytical tractability.

Learning in games, at the population level, can be modeled by dynamic adjustment processes (Swinkels 1993, Fudenberg and Levine 1998). We might think of

best response dynamics as capturing a population of optimizers and replicator dynamics a population of imitators. In best response dynamics, players choose myopic best responses, ignoring the possibility that the population mixed strategy may soon change. A population of players slowly moves toward a best response to the current state of the population. For a matrix game, the resulting flows are piecewise linear. In replicator dynamics, players copy others' strategies, so more successful actions spread in the population. The dynamic is less greedy than best response dynamics: the population doesn't move straight to the best action but rather in a weighted direction of all above average actions. Actions initially not present in the population will never be tried with replicator dynamics. These population dynamics can be derived from the revision protocols that describe how the individual agents learn.

There is reason to think that optimizers and imitators coexist in the same population (Conlisk, 1980). We might also think that each agent relies on a hybrid of learning styles, as suggested by the success of experience weighted attraction. In this chapter, we explore three different ways of combining revision protocols that lead to best response and replicator dynamics respectively, from rules in which either learning protocol may be tried with some positive probability to a rule that incorporates aspects of both.

We first consider a population in which agents sometimes use a revision protocol that generates best response dynamics and other times use a protocol that generates replicator dynamics. With this composition, all the agents use the same rule, and this rule is a linear combination of the best response and replicator protocols. We then consider a population in which some agents always use the best response protocol and others always use the replicator protocol. This scenario has each agent use a pure rule, but allows different agents to use different protocols. These two versions

allow us to compare different sources for the heterogeneity of learning styles in the population. Lastly, we also consider a population in which all agents use a hybrid revision protocol that is in between the ones that generate pure best response and pure replicator dynamics. In all cases, an entirely new learning dynamic emerges.

We compare equilibrium selection following from the various compositions of learning styles as well as from pure best response and replicator dynamics, and we find they may be markedly different from each other in their long-run behavior. For specific classes of coordination games, we show that introducing even a small presence of a second revision protocol into a population changes which equilibrium the population locates. Moreover, combining different types of learning rules in the same population leads to still different behavior from allowing agents to use a hybrid rule. These results underscore the importance of accurately determining exactly how agents learn.

Our results build off of Chapter III and IV's comparisons of basins of attraction and equilibrium selection under the best response dynamics and replicator dynamics (Golman and Page 2008a, Golman and Page 2008b). The contribution of this chapter is to explore the behavior of compositions of these dynamics. We find that heterogeneity across learning styles may bring about new outcomes not foreseeable from analysis of the homogeneous dynamics induced by the component learning styles.

Our treatment of equilibrium selection relies on analysis of the deterministic dynamical system. The learning rules define paths through the strategy space that attain equilibria in the games we consider. The equilibrium that is selected is thus a function of the initial population mixed strategy point and the learning dynamic. Nevertheless, we obtain results that hold throughout the interior of the strategy space by focusing on cases in which the basin of attraction of a particular equilibrium ap-

proaches the entire space. Our approach does not assume random shocks that allow the population to move from one equilibrium to another, as in the stochastic stability literature (Foster and Young 1990, Young 1993, Kandori et al. 1993, Ellison 2000, Binmore and Samuelson 1999). But, our findings complement that sort of ultralong run analysis because an equilibrium whose basin of attraction approaches the entire strategy space will be uniquely stochastically stable.

The rest of the chapter is organized as follows. The next section defines the revision protocols and the learning dynamics derived from them. In Section 5.3, we present our results comparing equilibrium selection of these rules. Section 5.4 concludes with a discussion of the importance of models that capture heterogeneity of learning styles. The Appendix contains proofs.

5.2 The Learning Dynamics

We assume a symmetric game, with agents recurrently randomly matched from a single population. The set of actions is finite, $\mathcal{A} = \{1, \ldots, n\}$. As we have done in previous chapters, we let x_i refer to the fraction of the society choosing action $i \in \mathcal{A}$. The population mixed strategy $\mathbf{x} = (x_1, \ldots, x_n)$ is an element of $\triangle^{n-1}$, the $(n-1)$-dimensional simplex where $x_i \geq 0$ for all i and $\sum_i x_i = 1$. We denote by π_i the payoff to action i. As we have mentioned in previous chapters, payoffs are a function of the population mixed strategy, $\pi_i : \triangle^{n-1} \to \Re$, but we omit the function's argument for ease of notation, writing π_i in place of $\pi_i(\mathbf{x})$. We denote the vector of these payoffs by $\vec{\pi} = (\pi_1, \ldots, \pi_n)$. We let $\mathrm{BR}(\mathbf{x})$ be the set of best replies to $\mathbf{x}$,

$$\mathrm{BR}(\mathbf{x}) = \arg\max_{\mathbf{v} \in \triangle^{n-1}} \mathbf{v} \cdot \vec{\pi}.$$

Best response dynamics can be written as

$$(5.1) \qquad\qquad \dot{\mathbf{x}} \in \mathrm{BR}(\mathbf{x}) - \mathbf{x}.$$

The best response dynamics describe a population always moving toward a best response to its current state. The replicator dynamics are

(5.2) $$\dot{x}_i = x_i(\pi_i - \bar{\pi})$$

where $\bar{\pi} = \mathbf{x} \cdot \vec{\pi}$ is the average payoff.

The best response dynamics and replicator dynamics are defined above in terms of the behavior of the population as a whole in the limit as the number of agents are large. But these dynamics can also arise by defining the behavior of the individual agents. Sandholm (2009) envisions a population of agents equipped with Poisson alarm clocks and allowed to switch strategies when their clocks ring. A revision protocol determines the probability of switching to another action as a function of the population state and the current payoffs. The expected motion of this stochastic process defines the learning dynamic. The best response dynamics, for example, emerges when each individual agent switches to the action with highest current payoff when its alarm goes off.

On its face, the replicator dynamics describes a population in which agents can survive and replicate proportionally to their current payoff. Thus, agents taking actions with below average payoff die off, while agents taking actions with above average payoff replicate. But the replicator dynamics can also emerge when individual agents follow the proportional imitation revision protocol (Schlag 1998). When an agent's alarm goes off, the agent picks an opponent at random and observes this opponent's action. (The selection of an opponent to observe is entirely independent of the random matching of opponents to play against.[1]) The agent observes action j with probability x_j. The agent then switches to the opponent's strategy only if the opponent's payoff is higher than his own, and even then only with probability

---

[1] See Boylan (1992) and Gilboa and Matsui (1992) for justification of random matching in a countably infinite population and Alos-Ferrer (1999) for the case of a continuum of agents.

proportional to the payoff difference. Thus, the probability of a switch from action i to action j is

$$\frac{x_j(\pi_j - \pi_i)}{\pi_{\max} - \pi_{\min}} \text{ if } \pi_j > \pi_i$$

and 0 otherwise.

It is worth noting that superficially different payoff matrices may generate the same flows for these learning dynamics. Specifically, the paths of the best response and replicator dynamics are invariant under positive affine transformations of the payoff matrix. (Transformations of payoffs that do not affect the dynamics can still determine which of multiple equilibria is payoff dominant. Thus, this traditional equilibrium selection criterion is unrelated to the dynamical equilibrium selection we presently consider.) Invariance of the best response dynamics is trivial – best responses do not change when a constant is added to all payoffs or when a positive constant is multiplied by all payoffs. Under replicator dynamics, the transformation affects π_i and $\bar{\pi}$ the same way. If a constant is added to all the payoffs, $\pi_i - \bar{\pi}$ remains invariant. If the payoffs are all multiplied by $\alpha > 0$, then $\dot{\mathbf{x}}$ becomes $\alpha\dot{\mathbf{x}}$. This affects the speed of the flow, but not the path the dynamic takes. Flows are considered identical if their paths are the same, because to apply a learning dynamic in a practical situation, the time parameter must be scaled to the physical time.

When trying to combine best response and replicator dynamics, the time parameter becomes relevant. The first step is to put both learning dynamics on the same footing. The framework we adopt is a generalized imitation revision protocol. When an alarm clock rings, the agent chooses at random an opponent to observe and switches to the opponent's strategy with probability q that depends on the opponent's payoff. If $q(\pi_j) = \frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}}$, the revision protocol generates the replicator dynamics, with a continual rescaling of time. For this *replicator protocol*, the proba-

bility of imitation increases linearly in the observed payoff. Alternatively, if $q(\pi_j) = 1$ when $\pi_j = \pi_{\max}$ and 0 otherwise, the revision protocol generates the best response dynamics,[2] with a continual rescaling of time and the danger that the dynamics stop altogether if the best response is absent in the population. This danger can be avoided by assuming that the initial population state is in the interior of the strategy space, and consequently, that a best response is initially present in the population. This *best response protocol* amounts to imitating only the best actions.

One natural way to combine best response and replicator dynamics is to suppose that agents use the replicator protocol with probability p and the best response protocol with probability $1 - p$. A linear combination of best response and replicator dynamics emerges:

$$(5.3) \qquad \dot{x}_i = x_i \left[p \, \frac{\pi_i - \bar{\pi}}{\pi_{\max} - \pi_{\min}} + (1 - p) \left(I(i) - \sum_j I(j) \, x_j \right) \right]$$

where the best response indicator function $I(i) = 1$ if $\pi_i = \pi_{\max}$ and 0 otherwise.

We can recover homogeneous use of a single learning protocol by setting $p = 1$ or $p = 0$. With $p = 1$, equation (5.3) becomes

$$(5.4) \qquad \dot{x}_i = x_i \left(\frac{\pi_i - \bar{\pi}}{\pi_{\max} - \pi_{\min}} \right).$$

The factor $\frac{1}{\pi_{\max} - \pi_{\min}}$ comes from the replicator revision protocol, where it is necessary to ensure that we can interpret q as a probability, i.e., that $q \leq 1$. With the appropriate rescaling of time, it would drop out of equation (5.4), producing the replicator dynamics of equation (5.2). With $p = 0$, the linear combination dynamics (5.3) reduces to

$$(5.5) \qquad \dot{x}_i = x_i \left(I(i) - \sum_j I(j) \, x_j \right).$$

---

[2]It might be fairer to say that this revision protocol generates a version of the best response dynamics, as it can align only with a single element of the differential inclusion (equation (5.1)). This version of the best response dynamics has the additional property that every Nash Equilibrium (even in mixed strategies) must be a steady state.

When there is a unique best response – say action b – (which, for almost any pay-off matrix, there is at almost every point in the simplex) equation (5.5) simplifies to $\dot{\mathbf{x}} = x_{\mathrm{b}} \left(\mathrm{BR}(\mathbf{x}) - \mathbf{x} \right)$. The additional factor of x_{b} here, as compared with equation (5.1), arises from using an imitative revision protocol to generate the best response dynamics and could be eliminated (in the interior of the strategy space) with a rescaling of time.[3]

Another possibility, instead of assuming that each agent sometimes uses each protocol, is that there could be some agents who always use the best response protocol and other agents who always use the replicator protocol. The theoretical analysis is more complicated in this case, but by capturing heterogeneity of learning styles across the population, the model appears more realistic. Consider a partition of the population into subpopulations of best-responders and replicators, $\mathbf{y}$ and $\mathbf{z}$ respectively. For all i, $x_i = y_i + z_i$. Learning styles are fixed for all time, so $\sum_i z_i = \rho$ is the fraction of agents using the replicator protocol, and $\sum_i y_i = 1 - \rho$ is the fraction using the best response protocol. We assume the two subpopulations are well-mixed, both for the purposes of playing the game and for finding an agent to possibly imitate. Thus, the dynamics of these two subpopulations are coupled together. For the subpopulation learning with the replicator protocol, the following dynamics emerges:

$$
\begin{aligned}
\dot{z}_i = \sum_j z_j z_i \left(\frac{\pi_i - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right) - z_i \sum_j z_j \left(\frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right) + \\
\sum_j z_j y_i \left(\frac{\pi_i - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right) - z_i \sum_j y_j \left(\frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right).
\end{aligned}
$$

(5.6)

The first two terms of (5.6) represent learning from other members of this same subpopulation while the last two terms represent learning through imitation of agents who happen to be best-responders. Let $\bar{\pi}_\mu$ be the average payoff in subpopulation

---

[3]When there are multiple best responses, equation (5.5), adjusted for an overall factor of $\sum_j I(j) x_j$, is consistent with (though not implied by) equation (5.1).

μ, meaning that $\mathbf{z} \cdot \vec{\pi} = \rho \bar{\pi}_z$ and $\mathbf{y} \cdot \vec{\pi} = (1 - \rho)\bar{\pi}_y$. The dynamics (5.6) simplifies to

$$(5.7) \quad \dot{z}_i = \rho z_i \left(\frac{\pi_i - \bar{\pi}_z}{\pi_{\max} - \pi_{\min}} \right) + \rho y_i \left(\frac{\pi_i - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right) - (1 - \rho) z_i \left(\frac{\bar{\pi}_y - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right).$$

For the subpopulation learning with the best response protocol, the dynamics are:

$$(5.8) \qquad \dot{y}_i = I(i)(1 - \rho)(z_i + y_i) - y_i \sum_j I(j)(z_j + y_j).$$

We refer to the system defined by (5.7) and (5.8) as the two-subpopulation dynamics.[4]

An alternative generalization of the best response and replicator dynamics is to consider $q(\pi_j) = \frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}}$ if $\pi_j \geq k(\vec{\pi})$ and 0 otherwise. The function $k(\vec{\pi})$ parametrizes an entire class of revision protocols. Choosing $k = \pi_{\max}$ corresponds to the best response protocol, $k = \pi_{\min}$ to the replicator protocol, and k values inside this range to a hybridization of these protocols. The resulting hybrid learning dynamics can be written as follows:

$$(5.9) \qquad \dot{x}_i = \begin{cases} x_i \left(\frac{\pi_i - \bar{\pi}}{\pi_{\max} - \pi_{\min}} + \sum_{j: \pi_j < k(\vec{\pi})} x_j \frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right) & \text{if } \pi_i \geq k(\vec{\pi}); \\ -x_i \sum_{j: \pi_j \geq k(\vec{\pi})} x_j \frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}} & \text{if } \pi_i < k(\vec{\pi}). \end{cases}$$

In a three-by-three game, this hybrid dynamics simplifies to time-scaled versions of best response dynamics (5.5) if only one action has payoff at least k and replicator dynamics (5.4) if two or all three actions have payoff at least k. New and interesting behavior emerges in larger games.

5.3 Results

The first place to compare different learning rules is in the local stability of their rest points. Best response dynamics and replicator dynamics usually have the same asymptotic properties (Hopkins 1999). Strict pure Nash Equilibria, which we will

---

[4]We can recover homogeneous learning dynamics, as in equation (5.4) or (5.5), from the two-subpopulation dynamics by setting $\rho = 1$ or $\rho = 0$ respectively.

consider in this chapter, are asymptotically stable for both dynamics.[5] Theorem V.1 is a straightforward result that says that stability shared by the underlying dynamics extends to our compositions of these dynamics.

Theorem V.1. *Suppose a pure Nash Equilibrium is asymptotically stable under both best response and replicator dynamics. Then it is asymptotically stable under the linear combination of these dynamics (5.3), the two-subpopulation dynamics (5.7) and (5.8), and the hybrid dynamics (5.9).*

Proof. Denote the equilibrium action a so that the equilibrium is $x_a = 1$, $x_{-a} = 0$. Asymptotic stability under best response dynamics means that there is a neighborhood V around the equilibrium in which action a is a best response. Let $\nu = \sup_{\mathbf{x} \notin V} x_a$. Define the neighborhood $U \subseteq V$ of all points satisfying $x_a > \nu$. For both the linear combination dynamics (5.3) and the hybrid dynamics (5.9), $\dot{x}_a > 0$ inside U (except right at the equilibrium, of course) because a is a best response. The neighborhood U was defined so that such a trajectory cannot escape it. Thus, both dynamics approach the pure Nash Equilibrium. Proving the asymptotic stability of the two-subpopulation dynamics is more involved, and this is done in the appendix. $\square$

So, strict equilibria are asymptotically stable for the linear combination dynamics, the two-subpopulation dynamics, and the hybrid dynamics. However, agreement of local stability properties does not ensure similar long-run behavior. In games with multiple equilibria, we must consider basins of attraction.

Best response dynamics and replicator dynamics often produce very similar basins of attraction for their equilibria. However, Chapter III provides a class of three-by-

---

[5]See Hofbauer, et. al. (1979), Hofbauer (2000), Hofbauer and Sigmund (2003), and Hofbauer, et. al. (2009) for additional stability results concerning the best response and replicator dynamics.

three games in which these dynamics have arbitrarily different basins of attraction (Golman and Page, 2008a). Games with such divergence between the behavior of the replicator dynamics and the best response dynamics lend themselves to the analysis of our composite dynamics. Intuition might lead one to believe that a combination of best response and replicator dynamics would always share some overlap in basins of attraction with each component rule, but this is not so. The linear combination dynamics can be fundamentally different from its parts. Here, we identify conditions under which best response dynamics and replicator dynamics and the linear combination of the two all have vanishing overlap in their basins of attraction at the same time. That is, by combining a best response protocol and a replicator protocol, the population can locate an equilibrium that neither rule could find on its own.

We begin by reviewing the notation. Let m be the Lebesgue measure on the strategy space $\triangle^{n-1}$. Given a vector of parameter values $\vec{P}$, let $G(\vec{P})$ be a class of symmetric normal form games with payoffs that depend on those parameters. That is, the payoff to action i against action j, π_{ij}, is a function of $\vec{P}$. Then, the expected payoff to action i given random matching with population mixed strategy $\mathbf{x} \in \triangle^{n-1}$ is $\pi_i = \sum_j \pi_{ij} x_j$. Given a learning rule $\mathcal{R}$ and an equilibrium action a of the game $G(\vec{P})$, let $B(\mathcal{R}, a, \vec{P})$ denote the basin of attraction of $(x_a = 1, x_{-a} = 0)$. Let $\mathbf{R}$ denote the replicator dynamics, $\mathbf{B}$ the best response dynamics, $\mathbf{L}(\mathbf{p})$ the linear combination using the replicator protocol with probability p, $\mathbf{S}(\rho)$ the two-subpopulation dynamics with ρ the fraction of the population using the replicator protocol, and $\mathbf{H}(\mathbf{k})$ the hybrid dynamics with threshold value $k(\vec{\pi})$.

Consider a class of symmetric games with five actions and the following payoffs:

$$(5.10) \qquad\qquad \pi_1 = \alpha x_5$$

$$(5.11) \qquad\qquad \pi_2 = \beta - \gamma x_3 - \kappa x_4$$

$$(5.12) \qquad\qquad \pi_3 = x_3 + x_4 - \delta x_4$$

$$(5.13) \qquad\qquad \pi_4 = 1 - \delta + \epsilon x_4$$

$$(5.14) \qquad\qquad \pi_5 = -\lambda$$

where the seven parameters above depend on $\vec{P}$ and all are positive. Denote the set of parameters $\Xi = \{\alpha, \beta, \gamma, \delta, \epsilon, \kappa, \lambda\}$. If $\beta > 1 - \delta$, and $\beta - \gamma < 1$, and $\beta - \kappa < 1 - \delta + \epsilon$, then actions 2, 3, and 4 are all strict equilibrium actions.

The conditions on the payoff parameters in Theorem V.2 can get technical, but they permit a simple interpretation of the game. We should think of it as a coordination game, but with a couple of extra options. Action 5 is just a bad choice. So bad, in fact, that all other actions look good in comparison to it if an agent is using the replicator learning protocol. While such strongly dominated strategies are often ignored when modeling games, it's hard to argue that incredibly stupid strategies don't exist, and there is both theoretical and experimental evidence that they can affect equilibrium selection (Ochs 1995, Basov 2004). Action 1, on the other hand, looks tempting because it can exploit action 5. It has the possibility of achieving a much higher payoff than actions 3 and 4, and unlike action 2, its payoff is guaranteed to be positive. In fact, action 1 is almost always the initial best response, but it's only a temporary best response because its payoff shrinks as agents learn to avoid action 5. Action 2 is high risk, high reward; agents want to coordinate on it if and only if they almost never see actions 3 or 4 played. Actions 3 and 4 are safer, with moderate payoffs. Of the pair, action 3 would yield a slightly higher equilibrium

payoff, whereas action 4 has the highest worst-case payoff.

In Theorem V.2, we identify conditions sufficient for the best response dynamics to lead to everyone playing action 2, the replicator dynamics to everyone playing action 3, and the linear combination dynamics to everyone taking action 4. This means that the best response dynamics selects the action that is a best response to the temporary best response; the replicator dynamics selects the action that remains better than average as long as it spreads; and the linear combination dynamics selects something else entirely.

Theorem V.2. *Consider the class of games with payoffs given by equations (5.10) through (5.14). Suppose the following limits hold as* $\vec{P} \to \vec{\tilde{P}}$: *i)* $\frac{\lambda}{\iota} \to \infty$ *for any* $\iota \in \Xi \backslash \{\lambda\}$; *ii)* $\alpha \to \infty$; *iii)* $\beta \to \infty$; *iv)* $\frac{\gamma}{\beta} \to \infty$; *v)* $\frac{\alpha\beta}{\gamma} \to \infty$; *vi)* $\frac{\alpha\beta}{\kappa} \to \infty$; *vii)* $\frac{\alpha^{1-p}\beta}{\kappa} \to 0$; *viii)* $\delta \to 0$; *ix)* $\frac{\epsilon}{\delta} \to 0$; *and x)* $\frac{\delta}{\epsilon} \left[\frac{\delta}{1-\delta} \left(\frac{1}{\alpha} \right)^{1-p} C \right]^{\left(\frac{1-p}{p} \right)\lambda} \to 0$ *for any constant* C. *Then, for any* $\mathcal{R} \in \{\mathbf{B}, \mathbf{R}, \mathbf{L}(\mathbf{p})\}$,

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=2}^{4} m\left(B(\mathcal{R}, i, \vec{P}) \right) = m\left(\triangle^4 \right),$$

i.e., the best response dynamics, the replicator dynamics, and the linear combination dynamics almost everywhere flow to a strict equilibrium featuring action 2, 3, or 4. Still,

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=2}^{4} m\left(B(\mathbf{B}, i, \vec{P}) \cap B(\mathbf{R}, i, \vec{P}) \right) = 0$$

and moreover,

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=2}^{4} m\left(B(\mathbf{L}(\mathbf{p}), i, \vec{P}) \cap \left(B(\mathbf{B}, i, \vec{P}) \cup B(\mathbf{R}, i, \vec{P}) \right) \right) = 0,$$

i.e., all three dynamics share with each other vanishing overlap in their basins of attraction.

Proof. See appendix. □

Not only can the linear combination dynamics differ from its component parts, but this deviation in long-run behavior can even occur with almost all weight on one piece of the linear combination. We now show that for a subset of the class of three-by-three games considered in Chapter III (Golman and Page, 2008a), the linear combination dynamics with any positive probability of using the replicator protocol, $p > 0$, agrees with the pure replicator dynamics and shares vanishing overlap in its basins of attraction with the pure best response dynamics. Thus, the mere possibility of using the replicator protocol, however unlikely, is enough to completely shift the long-term behavior of a population of agents using primarily the best response protocol.

Consider a three-by-three game. We adopt the same normalization conditions as in Chapter III (Golman and Page, 2008a). Using invariance under positive affine transformations of the payoff matrix, we can set $\pi_{3j} = 0$ for all j and $|\pi_{11}| \in \{0, 1\}$. Also without loss of generality we can renumber the three actions so that $(x_1 = 1, x_{-1} = 0)$ denotes the equilibrium attained by replicator dynamics and $(x_2 = 1, x_{-2} = 0)$ the equilibrium attained by best response dynamics. So, for $j \in \{1, 2\}$, $i \neq j$, $\lim_{\vec{P} \to \vec{P}} f_{jji}(\vec{P})(\pi_{jj} - \pi_{ij}) > 0$ for some functions $f_{jji} > 0$. And we also have $\lim_{\vec{P} \to \vec{P}} f_{321}(\vec{P})(\pi_{23} - \pi_{13}) > 0$ for some function $f_{321} > 0$.

Theorem V.3 identifies payoff conditions sufficient for the best response dynamics to lead to everyone playing action 2 while the linear combination dynamics leads to everyone taking action 1. Here again, the payoff conditions get technical, but allow for some interpretation. The possibility that either equilibrium action may be chosen makes the other equilibrium action unpalatable. Action 2 is subject to the greater damage in this regard, but it is also the best response if everybody takes action 3, the temporary initial best response. The best response dynamics start out

heading straight for the corner featuring action 3, and eventually action 2 becomes the best response and spreads. The linear combination dynamics, just a touch less greedy, do not wipe out action 1 quickly enough while the temporary best response is spreading, and when action 3 becomes unsustainable, it is action 1 that spreads. Thus, we find that the linear combination dynamics and the best response dynamics share vanishing overlap in their basins of attraction even if the linear combination dynamics is weighted heavily towards the best response dynamics.

Theorem V.3. *Consider an arbitrary three-by-three payoff matrix with the normalization conditions described above. For any $p > 0$,*

$$\lim_{\vec{P} \to \tilde{\vec{P}}} \sum_{i=1}^{2} m \left(B(\mathbf{L}(\mathbf{p}), i, \vec{P}) \cap B(\mathbf{B}, i, \vec{P}) \right) = 0$$

if: i) $\pi_{23} > 0$; ii) $\pi_{13} = 0$; iii) $\lim_{\vec{P} \to \tilde{\vec{P}}} \pi_{12} = -\infty$; iv) $\lim_{\vec{P} \to \tilde{\vec{P}}} \frac{\pi_{21}}{(-\pi_{12})^d} = -\infty$ for all d; v) $\lim_{\vec{P} \to \tilde{\vec{P}}} \frac{\pi_{21}}{\pi_{22}} = -\infty$; and vi) $\lim_{\vec{P} \to \tilde{\vec{P}}} \frac{\pi_{21}}{\pi_{23}(-\pi_{12})^d} = -\infty$ for all d.

Proof. See appendix. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Theorem V.3 identifies conditions under which a linear combination of best response and replicator dynamics, even one in which the weight on replicator dynamics is arbitrarily small, agrees with the pure replicator dynamics and disagrees with the pure best response dynamics. Conversely, we now identify conditions under which the linear combination with arbitrarily small weight on best response dynamics still agrees with pure best response dynamics and disagrees with pure replicator dynamics.

We consider a symmetric matrix game with four actions, selected from Chapter IV's class of multi-action stag hunt games in which best response dynamics and replicator dynamics have vanishing overlap in their basins of attraction (Golman

and Page 2008b). The payoff matrix is

(5.15)
$$
\begin{pmatrix}
c & c & c & c \\
1 & \beta+1 & 1 & 1 \\
0 & 0 & 0 & \gamma \\
0 & 0 & -\gamma & 0
\end{pmatrix}
$$

where β and γ both depend on $\vec{P}$, and c is a constant greater than 1. The first action is safe, like hunting a hare. The second action represents an attempt to cooperate, such as hunting a stag. For this action to be successful, it requires coordination. The third action is predatory toward the fourth action, which can be thought of as a failed attempt at cooperation. The parameter γ can thus be interpreted as the potential spoils of predation and β as the benefits of cooperation.

Chapter IV analyzes the best response dynamics and the replicator dynamics for this game and finds that if $\lim_{\vec{P}\to\vec{\tilde{P}}}\beta = \infty$ and $\lim_{\vec{P}\to\vec{\tilde{P}}}\frac{\gamma}{\beta} = \infty$, then for $\vec{P}$ near $\vec{\tilde{P}}$, best response dynamics flows to an equilibrium in which all players choose action 1 (safe action), but replicator dynamics flows to an equilibrium in which all players choose action 2 (cooperation).[6] In Theorem V.4 here, we find that the linear combination dynamics flows to the equilibrium featuring action 1, given a stronger payoff condition regarding how severely the benefits to predation overshadow the benefits to cooperation. Just the smallest chance of best responding is enough to bring about a coordination failure: the population turns away from cooperation while predation – the temporary initial best response – spreads; then, when predation becomes unsustainable, cooperation is no longer attractive. The result is that the linear combination dynamics weighted almost entirely towards the replicator protocol

---

[6]The game we consider here has payoffs slightly changed from the games we considered in Chapter IV. This minor change in payoffs does not affect our analysis. An earlier draft of Chapter IV proved the same result for a payoff structure that does align with the game we now consider (Golman and Page, 2008b).

still shares vanishing overlap in its basins of attraction with the pure replicator dynamics.

Theorem V.4. *Consider the payoff matrix given by (5.15). If* $\lim_{\vec{P} \to \vec{P}} \beta = \infty$ *and* $\lim_{\vec{P} \to \vec{P}} \frac{\gamma}{\beta^d} = \infty$ *for all d,[7] then for any* $p < 1$,

$$\lim_{\vec{P} \to \vec{P}} \sum_{i=1}^{2} m\left(B(\mathbf{L}(\mathbf{p}), i, \vec{P}) \cap B(\mathbf{R}, i, \vec{P})\right) = 0.$$

Proof. See appendix. □

Theorem V.4 illustrates that introducing even the smallest possibility of best responding into a population obeying the replicator dynamics can drastically change long-term behavior, shifting the population from an equilibrium in which everybody cooperates to one in which everybody takes the safe action in isolation.

With twice the variables, the two-subpopulation dynamics is much more complicated than the linear combination dynamics, but for this game we can show they behave alike. They select the same equilibrium as the best response dynamics does because the replicator protocol tends to promote imitating any action other than the fourth (the worst response against all profiles) while the best response protocol is more directed and thus has more control of the population. Theorem V.5, which says the two-subpopulation dynamics agrees with the linear combination dynamics on almost all of the strategy space, is proved by showing the two-subpopulation dynamics almost always selects the equilibrium $(1, 0, 0, 0)$. This holds regardless of the relative sizes of the subpopulations, just as according to Theorem V.4 the linear combination dynamics finds this equilibrium no matter what probability weight is placed on each protocol.

[7]If γ is only polynomially bigger than β, then it appears that there is a nonvanishing region near the equilibrium featuring action 2 in which the linear combination dynamics with $p > 0$ (i.e., some positive weight on the replicator protocol) finds this equilibrium, in agreement with the pure replicator dynamics.

Theorem V.5. *Consider the payoff matrix given by (5.15). Fix $\rho < 1$ and $p < 1$.*
If $\lim_{\vec{P} \to \vec{\tilde{P}}} \beta = \infty$ and $\lim_{\vec{P} \to \vec{\tilde{P}}} \frac{\gamma}{\beta^d} = \infty$ for all d, then

$$\lim_{\vec{P} \to \vec{\tilde{P}}} m\left(B(\mathbf{S}(\rho), 1, \vec{P}) \cap B(\mathbf{L}(\mathbf{p}), 1, \vec{P})\right) = m\left(\triangle^3\right).$$

Proof. See appendix. □

Staying with this same game, we now show that the hybrid dynamics can disagree with the linear combination dynamics and the two-subpopulation dynamics. We consider various threshold value functions, including constant functions $k(\vec{\pi}) = K$ and convex combinations of maximum and minimum payoffs $k(\vec{\pi}) = \alpha\pi_{\min} + (1-\alpha)\pi_{\max}$. The hybrid dynamics almost always finds $(0, 1, 0, 0)$ with these simple threshold value functions, but for other functions, such as $k(\vec{\pi}) = \max\{\pi_1, \pi_3\}$, the hybrid dynamics would agree with the linear combination dynamics, the two-subpopulation dynamics, and the best response dynamics.

Theorem V.6. *Consider the payoff matrix given by (5.15). If $\lim_{\vec{P} \to \vec{\tilde{P}}} \beta = \infty$ and $\lim_{\vec{P} \to \vec{\tilde{P}}} \frac{\gamma}{\beta} = \infty$, then when $k(\vec{\pi}) = K$ or $k(\vec{\pi}) = \alpha\pi_{\min} + (1-\alpha)\pi_{\max}$, $\alpha > 0$,*

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=1}^{2} m\left(B(\mathbf{H}(\mathbf{k}), i, \vec{P}) \cap B(\mathbf{B}, i, \vec{P})\right) = 0.$$

Proof. See appendix. □

Theorems V.4, V.5, and V.6 imply the following corollary:

Corollary V.7. *Consider the payoff matrix given by (5.15). If $\lim_{\vec{P} \to \vec{\tilde{P}}} \beta = \infty$ and $\lim_{\vec{P} \to \vec{\tilde{P}}} \frac{\gamma}{\beta^d} = \infty$ for all d, then for any $p < 1$, any $\rho < 1$, and $k(\vec{\pi}) = K$ or*

$$k(\vec{\pi}) = \alpha\pi_{\min} + (1 - \alpha)\pi_{\max}, \ \alpha > 0,$$

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=1}^{2} m\left(B(\mathbf{L}(\mathbf{p}), i, \vec{P}) \cap B(\mathbf{H}(\mathbf{k}), i, \vec{P})\right) = 0$$

and

$$\lim_{\vec{P} \to \vec{\tilde{P}}} \sum_{i=1}^{2} m\left(B(\mathbf{S}(\rho), i, \vec{P}) \cap B(\mathbf{H}(\mathbf{k}), i, \vec{P})\right) = 0.$$

Theorem V.6 is telling us that in this stag hunt game, the hybrid dynamics with a threshold value very close to the parameter which reduces it to the best response dynamics can nonetheless share vanishing overlap in its basins of attraction with the best response dynamics. Corollary V.7 then tells us that in this case, the hybrid dynamics reaches a different equilibrium than the linear combination dynamics and the two-subpopulation dynamics. Corollary V.7 thus makes it clear that the presence of multiple learning rules in a population produces behavior that is entirely different from that generated by merging these learning styles into a single hybrid rule. And all three methods of combining learning rules within a population can lead to new predictions for long-term behavior.

5.4 Discussion

In a large population of agents, we should expect heterogeneity of behavior (Hommes 2006, Kirman 2006).[8] Camerer and Ho (1999) correctly recognized that individuals may combine diverse basic learning rules. One might infer that the dynamics produced by a hybrid rule like theirs corresponds to the average dynamics of a heterogeneous population of agents who use one rule or the other. Here we see that might not be the case.

---

[8]Heterogeneity can take many forms. Ely and Sandholm (2005) consider agents who all use a common learning rule, but who have diverse preferences, i.e., heterogeneous payoff functions.

We also see that for a specific class of coordination games, the outcome of a learning dynamic that is a convex combination of best response dynamics and replicator dynamics is completely different than a convex combination of the outcomes produced by these individual rules. The simple assumption that the combination of two learning dynamics behaves like one of its component pieces is wrong. The explanation is that in games with many actions and multiple equilibria, the trajectory of the combination dynamics can enter parts of the strategy space that would not be approached by either of the component dynamics. Thus, the basins of attraction can be entirely reshaped when agents use multiple learning rules.

These results demonstrate that heterogeneity of learning styles matters. The presence of multiple learning styles in the same population, and the specifics of how they all come into play, affects equilibrium selection. The fact that hybridizations of diverse learning rules can produce aggregate behavior that differs from the behavior produced by each individual learning rule as well as from the behavior of a population of homogeneous agents using the "average" learning rule suggests that one-size-fits-all estimations of learning rules and learning rates may be inaccurate.

Theorems V.3, V.4, and V.5 in this chapter focus on the extreme cases in which rare changes in how agents learn drastically alter population behavior. In many other classes of games, behavior may change more gradually with the makeup of the population. In general, we should expect long term behavior to be sensitive to different learning styles, so a complete analysis of a game requires an accurate determination of how the players learn.

5.5 Appendix

Completing the Proof of Theorem V.1.

To show asymptotic stability of the two-subpopulation dynamics, we proceed from the constructs of the first part of the proof and consider a neighborhood around the equilibrium smaller than U. Let $\omega = \frac{\nu}{2} + \sqrt{\left(\frac{\nu}{2}\right)^2 + \rho(1-\nu)}$. Note that $\omega > \nu$ and $\omega > \rho$ for $0 < \rho < 1$, and $\omega < 1$ because $\nu < 1$. Define W as the neighborhood around the equilibrium satisfying $x_a > \omega$. We will show that trajectories starting in W approach the equilibrium.

Because $W \subset U$, we know that a is a best response in this neighborhood, and because of the asymptotic stability of the replicator dynamics, we know it is the unique best response except possibly right at the equilibrium. Looking at (5.8), we see that

$$\dot{y}_a = (1 - \rho - y_a)(z_a + y_a)$$

(5.16)
$$\geq y_a(1 - \rho - y_a)$$

as long as a remains the best response. Solving this logistic differential inequality, we get

(5.17)
$$y_a \geq \frac{(1-\rho)y_a(0)e^{(1-\rho)t}}{1 - \rho - y_a(0) + y_a(0)e^{(1-\rho)t}}.$$

The dynamics of the other subpopulation, (5.7), gives us

(5.18)
$$\dot{z}_a \geq \rho y_a - (1 - \rho)z_a.$$

Plugging in (5.17), this becomes

$$\dot{z}_a \geq \frac{\rho(1-\rho)y_a(0)e^{(1-\rho)t}}{1 - \rho - y_a(0) + y_a(0)e^{(1-\rho)t}} - (1 - \rho)z_a.$$

Multiplying through by $e^{(1-\rho)t}$ produces

$$\frac{d}{dt}\left(e^{(1-\rho)t}z_a\right) \geq \frac{\rho(1-\rho)y_a(0)e^{2(1-\rho)t}}{1-\rho-y_a(0)+y_a(0)e^{(1-\rho)t}},$$

which can be integrated to find

$$z_a \geq$$
$$z_a(0)e^{-(1-\rho)t}+\rho\left(1-e^{-(1-\rho)t}\right)-\frac{\rho(1-\rho-y_a(0))}{y_a(0)}e^{-(1-\rho)t}\ln\left(\frac{1-\rho-y_a(0)+y_a(0)e^{(1-\rho)t}}{1-\rho}\right).$$

We can loosen this lower bound slightly to simplify it as

$$(5.19) \qquad z_a \geq \rho - (\rho - z_a(0))\,e^{-(1-\rho)t} - \frac{\rho(1-\rho)(1-\rho-y_a(0))}{y_a(0)}te^{-(1-\rho)t}$$

Using these lower bounds on y_a and z_a, we now argue that a remains the best response forever. This justifies the use of these bounds for all t. The first order condition for minimizing the right hand side of (5.19) over time is

$$t = \frac{1}{1-\rho}\left(1 - \frac{y_a(0)\,(\rho - z_a(0))}{\rho\,(1-\rho-y_a(0))}\right)$$

as long as this is non-negative. In this case,

$$z_a \geq \rho - \frac{\rho\,(1-\rho-y_a(0))}{y_a(0)}e^{\frac{y_a(0)(\rho-z_a(0))}{\rho(1-\rho-y_a(0))}-1}$$

and $\frac{y_a(0)(\rho-z_a(0))}{\rho(1-\rho-y_a(0))} \leq 1$. Otherwise, the minimum is $z_a(0)$ right at $t = 0$. It is easy to see that $y_a \geq y_a(0)$ as well. When we have $x_a(t) \geq x_a(0)$, the trajectory never leaves W and it's obvious a is always the best response. When the first order condition on z_a is binding, we have $x_a(t) \geq y_a(0) + \rho - \frac{\rho(1-\rho-y_a(0))}{y_a(0)}$. Because the initial point is in W, $y_a(0) > \omega - z_a(0) \geq \omega - \rho$. So

$$(5.20) \qquad\qquad\qquad x_a(t) > \omega - \frac{\rho(1-\omega)}{\omega-\rho}.$$

We chose ω to make the right hand side of (5.20) equal to ν. This means the trajectory stays inside U, ensuring that a remains the best response.

Having justified our lower bounds on z_a and y_a for all time, it only remains to observe that as $t \rightarrow \infty$ these bounds (5.19) and (5.17) approach ρ and $1 - \rho$ respectively. $\qquad\square$

Proof of Theorem V.2.

Take $\vec{P}$ close to $\hat{\vec{P}}$. We first show that from almost any starting point in the strategy space, the best response dynamics locates the equilibrium in which only action 2 is played. Observe that if action 2 is ever the best response, its payoff increases while the other payoffs do not, because x_i decreases for $i \neq 2$. This implies that if action 2 is the best response, it remains so forever.

Observe that $\pi_3 \leq 1$ and $\pi_4 \leq 1$ everywhere. Consider starting points that satisfy

$$(5.21) \qquad\qquad x_5(0) > \frac{\gamma}{\alpha\beta}$$

$$(5.22) \qquad\qquad x_5(0) > \frac{\kappa}{\alpha\beta}$$

$$(5.23) \qquad\qquad x_5(0) > \frac{1}{\beta},$$

conditions that are met almost everywhere as a result of (v), (vi), and (iii) respectively. If action 2 is not initially the best response, it must be action 1, because inequality (5.21) and condition (iv) imply that $x_5(0) > \frac{1}{\alpha}$ and so $\pi_1 > 1$. In this case, all x_i, $i \neq 1$, have the same relative decay rate. While $x_i \geq \min\left\{\frac{\beta}{\gamma}, \frac{\beta}{\kappa}\right\} x_i(0)$,

$$\pi_1 \geq \alpha \min\left\{\frac{\beta}{\gamma}, \frac{\beta}{\kappa}\right\} x_5(0)$$
$$> 1.$$

This last step follows from inequalities (5.21) and (5.22). If x_i is to reach this lower bound, at the same time

$$\pi_2 \geq \beta\left(1 - x_3(0) - x_4(0)\right)$$
$$> 1,$$

the last step following from inequality (5.23). Thus, π_2 grows above π_3 and π_4 before π_1 falls to that level. Eventually, action 2 must become the best response, and the dynamics heads towards it thereafter.

We now show that the replicator dynamics leads to the equilibrium in which action 3 is always played. Under the replicator dynamics,

$$
\begin{aligned}
\dot{x}_3 &= x_3 \left[x_3(1 - x_3) + (1 - \delta)x_4(1 - x_3) - (1 - \delta)x_4 - \epsilon x_4^2 + (\gamma x_3 + \kappa x_4 - \beta)x_2 + \right. \\
&\qquad \left. \lambda x_5 - \alpha x_5 x_1 \right] \\
&= x_3 \left[x_3(x_1 + x_2 + x_5) + \delta x_3 x_4 - \epsilon x_4^2 + (\gamma x_3 + \kappa x_4 - \beta)x_2 + \lambda x_5 - \alpha x_5 x_1 \right]
\end{aligned}
$$

and we claim this is positive. We need to restrict to starting points that satisfy

$$
\text{(5.24)} \qquad\qquad x_3(0) > \frac{\epsilon}{\delta}
$$

$$
\text{(5.25)} \qquad\qquad x_3(0) > \frac{\beta}{\gamma},
$$

and conditions (ix) and (iv) ensure that these inequalities hold almost everywhere. Inequality (5.24) implies that $\delta x_3 x_4 - \epsilon x_4^2 > 0$ and (5.25) that $\gamma x_3 + \kappa x_4 - \beta > 0$. We have $\lambda x_5 - \alpha x_5 x_1 > 0$ from condition (i). Thus, $\dot{x}_3 > 0$ under the replicator dynamics.

Lastly, we show that the linear combination dynamics finds the equilibrium in which everyone takes action 4. Condition (i) guarantees that π_5 is always $\pi_{\min}$. Due to condition (vii), almost all initial points satisfy

$$
\text{(5.26)} \qquad\qquad x_4(0) > \frac{\alpha^{1-p}\beta}{\kappa}.
$$

Among other things, this means that $\pi_2 < 0$ initially. Assuming additionally that $x_5(0) > \frac{1}{\alpha}$, as condition (ii) allows us to do, $\pi_1 = \pi_{\max}$ initially. While this remains

the case, the linear combination dynamics yield

$$(5.27) \qquad \dot{x}_1 = x_1 \left(p\frac{\alpha x_5 - \bar{\pi}}{\lambda + \alpha x_5} + (1-p)(1-x_1) \right)$$

$$(5.28) \qquad \dot{x}_3 = x_3 \left(p\frac{x_3 + (1-\delta)x_4 - \bar{\pi}}{\lambda + \alpha x_5} - (1-p)x_1 \right)$$

$$(5.29) \qquad \dot{x}_4 = x_4 \left(p\frac{1 - \delta + \epsilon x_4 - \bar{\pi}}{\lambda + \alpha x_5} - (1-p)x_1 \right)$$

$$(5.30) \qquad \dot{x}_5 = x_5 \left(p\frac{-\lambda - \bar{\pi}}{\lambda + \alpha x_5} - (1-p)x_1 \right).$$

While $x_5 > \frac{1}{\lambda - \alpha}$ and $x_4 > \frac{\beta}{\kappa}$, $\bar{\pi} < 0$, and we have the following bounds:

$$(5.31) \qquad \dot{x}_1 > (1-p)x_1(1-x_1)$$

$$(5.32) \qquad \dot{x}_3 > -(1-p)x_1 x_3$$

$$(5.33) \qquad \dot{x}_4 > -(1-p)x_1 x_4$$

$$(5.34) \qquad \dot{x}_5 > -x_5.$$

Also note that

$$\frac{-\lambda - \bar{\pi}}{\lambda + \alpha x_5} < \frac{-\lambda(1 - x_5) + (\gamma x_3 + \kappa x_4)x_2 - \alpha x_5 x_1}{\lambda + \alpha x_5}$$

$$< \frac{(-\lambda - \alpha x_5)x_1 + (-\lambda + \gamma x_3 + \kappa x_4)x_2}{\lambda + \alpha x_5}$$

$$< -x_1.$$

So we can bound the left hand side of equation (5.30),

$$(5.35) \qquad \dot{x}_5 < -x_1 x_5.$$

So

$$(5.36) \qquad x_1(t) > \frac{1}{1 + \frac{1-x_1(0)}{x_1(0)}e^{-(1-p)t}}$$

$$(5.37) \qquad x_3(t) > x_3(0)e^{-(1-p)\int_0^t x_1(s)ds}$$

$$(5.38) \qquad x_4(t) > x_4(0)e^{-(1-p)\int_0^t x_1(s)ds}$$

$$(5.39) \qquad x_5(t) < x_5(0)e^{-\int_0^t x_1(s)ds}$$

$$(5.40) \qquad x_5(t) > x_5(0)e^{-t}.$$

By condition (viii), almost all initial points satisfy $x_1(0) > \frac{\delta}{1-\delta}$. Observe that $\pi_4 > \pi_3 + (1-\delta)x_1 - \delta x_3$. As long as the dynamics as described by equation (5.27) apply, $x_1(t) \geq x_1(0)$, so $\pi_4(t) > \pi_3(t)$.

While $x_5(t) > \frac{1-\delta}{\alpha}$, inequality (5.39) implies $e^{-\int_0^t x_1(s)ds} > \frac{1-\delta}{\alpha x_5(0)}$. So, using inequality (5.38), $x_4(t) > x_4(0)\left(\frac{1-\delta}{\alpha x_5(0)}\right)^{1-p}$. By condition (viii), $1 - \delta > x_5(0)$ almost everywhere, and by inequality (5.26), this simplifies to $x_4(t) > \frac{\beta}{\kappa}$. So $\pi_2(t) < 0$ as long as $x_5(t) > \frac{1-\delta}{\alpha}$.

Let $t_c = t : \pi_1(t) = \pi_4(t)$. Thus, $x_5(t_c) > \frac{1-\delta}{\alpha}$. This means $\pi_4(t) > \pi_2(t)$ and $\pi_4(t) > \pi_3(t)$ for all $t \leq t_c$. The dynamics described by equations (5.31) through (5.35) hold until t_c, at which point π_4 becomes $\pi_{\max}$.

Because $\pi_4(t) > \pi_3(t)$ for all $t \leq t_c$, we can also conclude that

$$(5.41) \qquad \frac{x_4(t_c)}{x_4(0)} > \frac{x_3(t_c)}{x_3(0)}.$$

We know that $x_5(t_c) < \frac{1}{\alpha}$, too. Inequality (5.40) then implies $e^{-t_c} < \frac{1}{\alpha x_5(0)}$, and using inequality (5.36), we come to

$$(5.42) \qquad x_1(t_c) > \frac{1}{1 + \frac{1-x_1(0)}{x_1(0)}\left(\frac{1}{\alpha x_5(0)}\right)^{1-p}}.$$

After t_c, while π_4 remains $\pi_{\max}$, the linear combination dynamics yield

$$\begin{aligned}
\dot{x}_1 &= x_1\left(p\frac{\pi_1 - \bar{\pi}}{\lambda + \pi_4} - (1-p)x_4\right) \\
\dot{x}_3 &= x_3\left(p\frac{\pi_3 - \bar{\pi}}{\lambda + \pi_4} - (1-p)x_4\right) \\
\dot{x}_4 &= x_4\left(p\frac{\pi_4 - \bar{\pi}}{\lambda + \pi_4} + (1-p)(1-x_4)\right).
\end{aligned}$$

Let

$$f(t) = p\frac{\pi_3 - \bar{\pi}}{\lambda + \pi_4} - (1-p)x_4,$$

so that $\dot{x}_3 = x_3\, f(t)$. For $t \geq t_c$,

$$(5.43) \qquad x_3(t) = x_3(t_c)e^{\int_{t_c}^{t} f(s)ds}.$$

Because $\pi_3 \leq 1$ and $\pi_1 \geq 0$, we know that $\dot{x}_1 \geq x_1\left(f(t) - p\frac{1}{\lambda}\right)$. So

$$(5.44) \qquad x_1(t) \geq x_1(t_c)e^{\int_{t_c}^{t} f(s)ds} \cdot e^{-\frac{p}{\lambda}(t-t_c)}.$$

As $\pi_4 > \pi_3$, we know $\dot{x}_4 > x_4\left(f(t) + 1 - p\right)$, and

$$(5.45) \qquad x_4(t) \geq x_4(t_c)e^{\int_{t_c}^{t} f(s)ds} \cdot e^{(1-p)(t-t_c)}.$$

Recall that $\pi_4 > \pi_3$ if $(1-\delta)x_1 > \delta x_3$. Putting together equations (5.43) and (5.44), this holds as long as $(1-\delta)x_1(t_c)e^{-\frac{p}{\lambda}(t-t_c)} > \delta(1-x_1(t_c))$. Plugging in inequality (5.42) and arranging terms, we find that $\pi_4 > \pi_3$ if

$$(5.46) \qquad e^{-(t-t_c)} > \left[\frac{\delta}{1-\delta}\frac{1-x_1(0)}{x_1(0)}\left(\frac{1}{\alpha x_5(0)}\right)^{1-p}\right]^{\frac{\lambda}{p}}.$$

Note that we also obtain $\pi_4 > \pi_3$ if $\epsilon x_4 > \delta x_3$. Putting together equations (5.43), (5.45) and (5.41), we find that this holds when $\epsilon x_4(0)e^{(1-p)(t-t_c)} > \delta x_3(0)$. Arranging terms, we obtain $\pi_4 > \pi_3$ when

$$(5.47) \qquad e^{-(t-t_c)} < \left(\frac{\epsilon x_4(0)}{\delta x_3(0)}\right)^{\frac{1}{1-p}}.$$

By combining equations (5.46) and (5.47), we can be sure $\pi_4 > \pi_3$ for all t if

$$(5.48) \qquad \left(\frac{\epsilon x_4(0)}{\delta x_3(0)}\right)^{\frac{1}{1-p}} > \left[\frac{\delta}{1-\delta}\frac{1-x_1(0)}{x_1(0)}\left(\frac{1}{\alpha x_5(0)}\right)^{1-p}\right]^{\frac{\lambda}{p}}.$$

It follows from condition (x) that inequality (5.48) holds almost everywhere in the strategy space.

While $\pi_4 = \pi_{\max}$, obviously x_4 is increasing, and so π_2 remains negative for all t after t_c. Thus, action 4 remains the best response forever, and the linear combination dynamics locates the equilibrium in which only action 4 is played. $\qquad \square$

Proof of Theorem V.3.

The conditions in the theorem satisfy a more general set of conditions that Chapter III shows to be sufficient for best response dynamics to find $(0, 1, 0)$ from almost any initial point in the strategy space (Golman and Page, 2008a). Here, we show that the linear combination of dynamics finds $(1, 0, 0)$ from almost any initial point.

For $\vec{P}$ near enough to $\hat{\vec{P}}$, any initial point in the interior of the strategy space satisfies $x_2 > \frac{1}{-\pi_{12}}$ and $x_1 > \frac{C(-\pi_{12})+\pi_{22}+\pi_{23}}{-\pi_{21}}$, with $C > 1$. The first inequality follows from condition (iii) and the second from conditions (iv), (v), and (vi). The first inequality implies that initially $\pi_1 < 0$. From the second, we have $\pi_{21}x_1 + \pi_{22} + \pi_{23} < C\pi_{12}$, so initially $\pi_2 < C\pi_1$. By our normalization condition, $\pi_3 = 0$ always. Thus, we have established that initially $\pi_{\max} = 0$ and $\pi_{\min} = \pi_2$. While this remains the case, the linear combination dynamics from equation (5.3) can be written as:

$$(5.49) \qquad \begin{aligned} \dot{x}_1 &= x_1\left(p\frac{\pi_1(1-x_1)-\pi_2 x_2}{-\pi_2}-(1-p)x_3\right) \\ &= x_1\left(p\left(1-\frac{\pi_1}{\pi_2}\right)(1-x_1)-x_3\right) \end{aligned}$$

$$(5.50) \qquad \begin{aligned} \dot{x}_2 &= x_2\left(p\frac{-\pi_1 x_1+\pi_2(1-x_2)}{-\pi_2}-(1-p)x_3\right) \\ &= x_2\left(-p\left(1-\frac{\pi_1}{\pi_2}\right)x_1-x_3\right). \end{aligned}$$

These equations hold until $\pi_1 = 0$ or $\pi_1 = \pi_2$, and we now show that $\pi_1 = 0$ occurs first, even before $\frac{\pi_1}{\pi_2} = \frac{1}{C}$, for almost all initial points.

We can put a lower bound on the speed at which x_2 decays, $\dot{x}_2 < -x_2 x_3$. Thus, we have an upper bound on the fraction of the population taking action 2,

$$x_2(t) < x_2(0) \, e^{-\int_0^t x_3(s)ds}.$$

While $\frac{\pi_1}{\pi_2} < \frac{1}{C}$, we can place an upper bound on the speed at which x_1 decays, $\dot{x}_1 > x_1 x_3 \left(p(1 - \frac{1}{C}) - 1\right)$. Thus,

$$x_1(t) > x_1(0) \left(e^{-\int_0^t x_3(s)ds}\right)^{1-p(1-\frac{1}{C})}.$$

Putting these equations together, note that

$$\frac{x_1(t)}{x_2(t)} > \frac{x_1(0)}{x_2(0)} \left(e^{-\int_0^t x_3(s)ds}\right)^{-p(1-\frac{1}{C})}.$$

Using the normalization condition $\pi_{11} = 1$ as well as condition (ii), we find that $\pi_1 \geq 0$ if $\frac{x_1}{x_2} \geq -\pi_{12}$, and if differential equations (5.49) and (5.50) were to hold indefinitely, that would happen when

$$\frac{x_1(0)}{x_2(0)} \left(e^{-\int_0^t x_3(s)ds}\right)^{-p(1-\frac{1}{C})} \geq -\pi_{12}.$$

Rearranging terms, we find that $\pi_1 \geq 0$ when

$$e^{-\int_0^t x_3(s)ds} \leq \left(\frac{x_2(0)}{x_1(0)}(-\pi_{12})\right)^{\frac{1}{-p(1-\frac{1}{C})}}.$$

Thus, π_1 would hit 0 while

$$x_1(t) > x_1(0) \left(\frac{x_2(0)}{x_1(0)}(-\pi_{12})\right)^{\frac{1-p(1-\frac{1}{C})}{-p(1-\frac{1}{C})}}.$$

This simplifies to

$$x_1(t) > x_1(0) \left(\frac{x_2(0)}{x_1(0)}(-\pi_{12})\right)^{1-\frac{1}{p(1-\frac{1}{C})}}.$$

On the other hand, $\frac{\pi_1}{\pi_2} < \frac{1}{C}$ as long as $C\,\pi_{12} > \pi_{21}x_1 + \pi_{22}x_2 + \pi_{23}$ or equivalently $x_1 > \frac{C\,\pi_{12}-\pi_{23}}{\pi_{21}+\pi_{22}\frac{x_2}{x_1}}$. Recognizing that $\frac{x_2(t)}{x_1(t)} < \frac{x_2(0)}{x_1(0)}$, we find that this holds if $x_1(t) > \frac{C(-\pi_{12})+\pi_{23}}{-\pi_{21}-\pi_{22}\frac{x_2(0)}{x_1(0)}}$. We can be sure π_1 hits 0 before $\frac{\pi_1}{\pi_2} = \frac{1}{C}$ if

$$(5.51) \qquad x_1(0)\left(\frac{x_2(0)}{x_1(0)}(-\pi_{12})\right)^{1-\frac{1}{p(1-\frac{1}{C})}} > \frac{C(-\pi_{12})+\pi_{23}}{-\pi_{21}-\pi_{22}\frac{x_2(0)}{x_1(0)}}.$$

For $\vec{P}$ near enough to $\hat{\vec{P}}$, inequality (5.51) will hold for any interior initial point, because it holds whenever the following three inequalities apply: $\frac{x_2(0)}{x_1(0)} < \frac{-\pi_{21}}{2\pi_{22}}$; $\frac{x_2(0)}{x_1(0)} < -\pi_{12}$; and $x_1(0) > \frac{2(C(-\pi_{12})+\pi_{23})(-\pi_{12})^d}{-\pi_{21}}$ where $d = 2\left(\frac{1}{p(1-\frac{1}{C})}-1\right)$. The first follows from condition (v), the second from condition (iii), and the last from conditions (iv) and (vi).

When π_1 exceeds 0, $\pi_1 > \pi_3 > \pi_2$, and it remains to show that this ordering remains thereafter. The ordering implies that $\dot{x}_1 > 0$ and $\dot{x}_2 < 0$, which in turn implies that π_1 is increasing. If π_2 were to approach π_3, then $\dot{x}_3$ would become negative, and in turn π_2 would decrease. Thus, once π_1 exceeds 0, x_1 is forever increasing. $\qquad\square$

Proof of Theorem V.4.

We must show that the linear combination of dynamics finds $(1,0,0,0)$ from almost any initial point.

For $\vec{P}$ near enough to $\hat{\vec{P}}$, any initial point in the interior of the strategy space satisfies the following six inequalities:

$$(5.52) \qquad\qquad \gamma x_4(0) > \beta + 1$$

$$(5.53) \qquad\qquad (1-p)x_3(0) - p\frac{\beta+1}{\gamma x_3(0)} > 0$$

$$(5.54) \qquad x_4(0) > \left(\frac{2\beta}{c-1}\right)^d\frac{c}{\gamma} \text{ where } d = \frac{1}{(1-p)x_3(0) - p\frac{\beta+1}{\gamma x_3(0)}}$$

$$(5.55) \qquad \frac{x_3(0) \left(\frac{\gamma x_4(0)}{c}\right)^{1-p}}{1 - x_3(0) + x_3(0) \left(\frac{\gamma x_4(0)}{c}\right)^{1-p}} > \frac{c+1}{2c}$$

$$(5.56) \qquad \frac{p(\beta+1)}{\gamma \frac{c+1}{2c}} < (1-p)x_1(0) \left(x_3(0) \left(\frac{c}{\gamma x_4(0)}\right)^{1-p}\right)^{1+p}$$

and

$$(5.57) \qquad x_3(0) > \frac{\beta}{(1-p)\gamma}.$$

Inequality (5.52) implies that initially $\pi_3 > \pi_2$, and together with $\beta > c$, initially $\pi_3 > \pi_1$ as well. It is clear from the payoff matrix that $\pi_4 = \pi_{\min}$ for all points in the strategy space. While $\pi_3 > \pi_j > \pi_4$ for $j \in \{1, 2\}$,

$$\dot{x}_i = x_i \left(p\frac{\pi_i - \bar{\pi}}{\pi_3 - \pi_4} + (1-p)(\delta_{i3} - x_3) \right).$$

This leads to the following inequalities:

$$\dot{x}_1 \geq x_1 \left(-p\frac{\beta(x_2(0))^2}{\gamma x_3(0)} - (1-p) \right)$$

$$\dot{x}_2 \leq x_2 \left(p\frac{\beta+1}{\gamma x_3(0)} - (1-p)x_3(0) \right)$$

$$\dot{x}_3 \geq (1-p)x_3(1 - x_3)$$

$$\dot{x}_4 \leq x_4 \left(p\frac{-\gamma x_3}{\gamma(x_4 + x_3)} - (1-p)x_3 \right) \leq -x_4 x_3$$

$$\dot{x}_4 \geq -x_4.$$

So,

$$(5.58) \qquad x_1(t) \geq x_1(0) \left(e^{-t}\right)^{(1-p)+p\frac{\beta(x_2(0))^2}{\gamma x_3(0)}}$$

$$(5.59) \qquad x_2(t) \leq x_2(0) \left(e^{-t}\right)^{(1-p)x_3(0)-p\frac{\beta+1}{\gamma x_3(0)}}$$

$$(5.60) \qquad x_3(t) \geq \frac{x_3(0)e^{t(1-p)}}{1 - x_3(0) + x_3(0)e^{t(1-p)}}$$

$$(5.61) \qquad x_4(t) \leq x_4(0)e^{-\int_0^t x_3(s)ds}$$

$$(5.62) \qquad x_4(t) \geq x_4(0)e^{-t}.$$

Let $t_c = t : x_4(t) = \frac{c}{\gamma}$. Plugging into inequality (5.62), $\frac{c}{\gamma} \geq x_4(0)e^{-t_c}$, so $e^{-t_c} \leq \frac{c}{\gamma x_4(0)}$.

And plugging that in inequality (5.59), we find that

$$(5.63) \qquad\qquad x_2(t_c) \leq \left(\frac{c}{\gamma x_4(0)}\right)^{(1-p)x_3(0)-p\frac{\beta+1}{\gamma x_3(0)}}.$$

Note that inequality (5.63) requires a positive exponent on the right hand side, and that is exactly what we obtained with inequality (5.53). Inequality (5.63) can be simplified to $x_2(t_c) < \frac{c-1}{2\beta}$ by using inequality (5.54). We now have

$$(5.64) \qquad\qquad \beta x_2(t_c) + 1 < \frac{c+1}{2},$$

and recalling that $c > 1$, we see that this is less than c. At t_c, $\pi_3 = \pi_1$ by definition, but $\pi_2 < \pi_1$ by inequality (5.64). Thus, immediately after t_c, $\pi_1 = \pi_{\max}$, and it remains to show this is so at all future times as well.

Before we move on to $t > t_c$, it is necessary to establish a few more inequalities at t_c. From inequality (5.60), we have

$$x_3(t_c) \geq \frac{x_3(0)\left(\frac{\gamma x_4(0)}{c}\right)^{1-p}}{1 - x_3(0) + x_3(0)\left(\frac{\gamma x_4(0)}{c}\right)^{1-p}}.$$

By inequality (5.55), this becomes

$$(5.65) \qquad\qquad x_3(t_c) > \frac{c+1}{2c}.$$

From inequality (5.61), we have

$$(5.66) \qquad\qquad e^{-\int_0^{t_c} x_3(s)ds} \geq \frac{c}{\gamma x_4(0)}.$$

We can integrate inequality (5.60) to get

$$\begin{aligned}
\int_0^{t_c} x_3(s)ds &\geq \int_0^{t_c} \frac{x_3(0)e^{s(1-p)}}{1 - x_3(0) + x_3(0)e^{s(1-p)}}ds \\
&= \frac{1}{1-p}\ln\left(1 - x_3(0) + x_3(0)e^{t_c(1-p)}\right).
\end{aligned}$$

Plugging this into inequality (5.66), we have $\left(1 - x_3(0) + x_3(0)e^{t_c(1-p)}\right)^{\frac{1}{1-p}} \leq \frac{\gamma x_4(0)}{c}$.

So

$$e^{-t_c} \geq \left(\frac{x_3(0)}{\left(\frac{\gamma x_4(0)}{c}\right)^{1-p} - (1 - x_3(0))}\right)^{\frac{1}{1-p}} \geq \left(x_3(0)\left(\frac{c}{\gamma x_4(0)}\right)^{1-p}\right)^{\frac{1}{1-p}}.$$

Plugging this into inequality (5.58), we get

$$x_1(t_c) \geq x_1(0)\left(x_3(0)\left(\frac{c}{\gamma x_4(0)}\right)^{1-p}\right)^{1 + \frac{p\beta(x_2(0))^2}{(1-p)\gamma x_3(0)}}.$$

Applying inequality (5.57) to the exponent, we finally obtain

(5.67)
$$x_1(t_c) \geq x_1(0)\left(x_3(0)\left(\frac{c}{\gamma x_4(0)}\right)^{1-p}\right)^{1+p}.$$

For $t > t_c$, we have the following differential equations:

$$\dot{x}_i = x_i\left(p\frac{\pi_i - \bar{\pi}}{c + \gamma x_3} + (1 - p)(\delta_{i1} - x_1)\right).$$

These equations hold while $\pi_1 = \pi_{\max}$, and we will show that for $t > t_c$, that is always the case. In particular, we have

(5.68)
$$\dot{x}_2 \leq x_2\left(p\frac{\beta + 1}{\gamma x_3} - (1 - p)x_1\right).$$

Applying inequalities (5.65) and (5.67), note that

$$p\frac{\beta + 1}{\gamma x_3(t_c)} - (1 - p)x_1(t_c) \leq \frac{p(\beta + 1)}{\gamma\frac{c+1}{2c}} - (1 - p)x_1(0)\left(x_3(0)\left(\frac{c}{\gamma x_4(0)}\right)^{1-p}\right)^{1+p},$$

and by inequality (5.56), this is negative. So, the right hand side of equation (5.68) is negative at t_c. This means x_2 is decreasing immediately after t_c. And while x_2 is decreasing, $(x_1 + x_3)$ must be increasing. Let $t^* = t : x_1(t) = x_3(t_c)$ and $t > t_c$. For all t between t_c and t^*, $x_1(t)x_3(t) > x_1(t_c)x_3(t_c)$ because $(x_1 + x_3)$ has been increasing. Therefore, the right hand side of equation (5.68) must remain negative

until t^*. And thus, $x_2(t^*) < x_2(t_c)$. Observe that $\beta x_2(t_c) + 1 < cx_3(t_c)$ by combining inequalities (5.64) and (5.65). This means that

$$(5.69) \qquad\qquad \beta x_2(t^*) + 1 < cx_1(t^*).$$

So at t^*, $\pi_2 < \bar{\pi}$ and consequently $\dot{x}_2 < 0$. As t moves past t^*, x_1 increases and x_2 decreases, and that only reinforces inequality (5.69). So, this state of affairs persists forever. The combination dynamics approaches $(1, 0, 0, 0)$. $\qquad\qquad\square$

Proof of Theorem V.5.

We will show that starting at any point in the interior of the strategy space, the two-subpopulation dynamics eventually approaches $\mathbf{x} = (1, 0, 0, 0)$ for $\vec{P}$ close enough to $\vec{\hat{P}}$.

Many times throughout the proof, we rely on inequalities that hold for such $\vec{P}$ from the hypotheses of the theorem. To begin, we take $x_4(0) > \frac{\beta+1}{\gamma}$, a condition we will later refine. As it is, this ensures that action 3 is initially the best response. While this is so, we have: $\dot{y}_4 \geq -y_4$; $\dot{z}_4 \geq -z_4$; and $\dot{y}_2 \leq -y_2 \, y_3(0)$, making use of the fact that $y_3 \geq y_3(0)$. Solving these differential inequalities, we have $y_4 \geq y_4(0)e^{-t}$ and $z_4 \geq z_4(0)e^{-t}$, which can be summarized as

$$(5.70) \qquad\qquad x_4 \geq x_4(0)e^{-t},$$

as well as

$$(5.71) \qquad\qquad y_2 \leq y_2(0)e^{-y_3(0)t}.$$

Additionally, take

$$\frac{\beta(\beta + 1)}{\gamma} < \frac{(y_3(0))^2}{y_3(0) + y_4(0) + z_4(0)} \frac{(1 - \rho)}{\rho} \frac{(c - 1)}{2},$$

or equivalently,

$$\frac{\rho(\beta+1)}{\gamma y_3(0)} < \frac{c-1}{2\beta}(1-\rho)\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}.$$

With the notation

(5.72) $$\epsilon = \frac{c-1}{2\beta}(1-\rho)\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)},$$

we have $\frac{\rho\pi_2}{\pi_3-\pi_4} < \epsilon$. Notice that

$$\frac{-\pi_4}{\pi_3-\pi_4} = \frac{\gamma(x_3)}{\gamma(x_3+x_4)}$$

$$\geq \frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}.$$

Also notice that $-\rho\bar{\pi}_z - (1-\rho)\bar{\pi}_y = -\bar{\pi} < 0$. Putting this all together, we have

$\dot{z}_2 < \epsilon + \rho y_2 - (1-\rho)z_2\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}$. Plugging in (5.71) and multiplying through

by $e^{(1-\rho)\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t}$ produces

(5.73) $$\frac{d}{dt}\left(z_2\, e^{(1-\rho)\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t}\right) < e^{(1-\rho)\frac{y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t}\left(\epsilon + \rho\, y_2(0)\, e^{-y_3(0)t}\right).$$

To integrate, we must consider separately the cases: i) $1-\rho \neq y_3(0)+y_4(0)+z_4(0)$

and ii) they are equal. In case (i), we obtain

(5.74) $z_2 <$

$$z_2(0)e^{-\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t} + \frac{\epsilon\left(1-e^{-\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t}\right)}{\left(\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}\right)} + \frac{\rho y_2(0)\left(e^{-y_3(0)t} - e^{-\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}t}\right)}{\left(\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)} - y_3(0)\right)}.$$

Let $M = \min\left\{1, \frac{1-\rho}{y_3(0)+y_4(0)+z_4(0)}\right\}$ and take

$$\ln\left(\frac{\gamma x_4(0)}{\beta+1}\right) > \frac{1}{\left|\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)} - y_3(0)\right|}$$

to find a weaker upper bound in which the last term of (5.74) is replaced by

(5.75) $$\ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)\rho y_2(0)e^{-y_3(0)Mt}.$$

In case (ii), the first two terms of (5.74) are unchanged, but the last term is replaced by

$$(5.76) \qquad \rho y_2(0) t e^{-y_3(0)\,t}$$

because of cancellation of exponential pieces in (5.73).

The bounds on our dynamical variables that we have established so far only hold until a new action becomes a best response. We now show that it is action 1 that becomes the best response when the payoff to action 3 drops sufficiently. Specifically, by the time π_3 falls to the maximum possible payoff to action 2, the actual payoff π_2 has sunk irretrievably below π_1. We have $\gamma x_4 > \beta + 1$ when $x_4(0)e^{-t} > \frac{\beta+1}{\gamma}$, according to inequality (5.70). So action 3 is surely still the best response while $t \le \ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)$. We will show $\beta(y_2 + z_2) + 1 < c$ for $t \ge \ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)$ until the time that $\pi_3 = c$. We take

$$\frac{1}{y_3(0)} < \ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)$$

to ensure that our upper bound on z_2 in both cases is a decreasing function for $t \ge \ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)$. For such t, no matter which case applies, we have

(5.77)

$$z_2 < \frac{\epsilon}{\left(\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}\right)} + \left(z_2(0) - \frac{\epsilon}{\left(\frac{(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}\right)}\right)\left(\frac{\gamma x_4(0)}{\beta+1}\right)^{\frac{-(1-\rho)y_3(0)}{y_3(0)+y_4(0)+z_4(0)}}$$

$$+ \rho y_2(0) \ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)\left(\frac{\gamma x_4(0)}{\beta+1}\right)^{-y_3(0)\,M}.$$

The first two terms of (5.77) correspond to the first two terms in (5.74), rearranged and with $\ln\left(\frac{\gamma x_4(0)}{\beta+1}\right)$ plugged in for t. The final term of (5.77) is an upper bound on both (5.75) and (5.76). From (5.71), we also have

$$(5.78) \qquad y_2 \le y_2(0)\left(\frac{\gamma x_4(0)}{\beta+1}\right)^{-y_3(0)}$$

for $t \geq \ln \left(\frac{\gamma x_4(0)}{\beta + 1} \right)$. Putting together inequalities (5.77) and (5.78), and multiplying through by β, we have an upper bound on $\beta (y_2 + z_2)$ for applicable t. Yet, all but one of the terms in this upper bound can be made arbitrarily small because they contain a factor of γ raised to a negative exponent. The remaining term is the first term of (5.77) multiplied by β, which means

$$\beta (y_2 + z_2) < \frac{\beta \epsilon}{\left(\frac{(1-\rho)y_3(0)}{y_3(0) + y_4(0) + z_4(0)} \right)} + \text{ small terms.}$$

Recalling our definition of ϵ, from equation (5.72), we have

$$\beta (y_2 + z_2) < \frac{c-1}{2} + \text{ small terms,}$$

and thus $\beta(y_2 + z_2) + 1 < c$.

Moving forward, we will have use for an upper bound on y_2 at the time that action 1 becomes the best response. Refer to this time as $t = t_c$. It is defined by $\pi_3 = \pi_1$ or in other words $\gamma x_4 = c$ at $t = t_c$. By inequality (5.70), $c \geq \gamma x_4(0)e^{-t_c}$. So $t_c \geq \ln \left(\frac{\gamma x_4(0)}{c} \right)$. Plugging this into inequality (5.71) produces

(5.79)
$$y_2(t_c) \leq y_2(0) \left(\frac{\gamma x_4(0)}{c} \right)^{-y_3(0)}.$$

Now let us consider the dynamics after action 1 becomes the best response. We aim to show that action 1 remains the best response at all later times. Both y_4 and z_4 continue to decrease because action 4 remains the worst response, so $\gamma x_4 < c$ for $t > t_c$. This means action 3 can never be revived as the best response. Additionally, $y_2 < y_2(t_c)$ for $t > t_c$ because only the best response can grow in the subpopulation of best responders. We need an upper bound on z_2 for $t > t_c$. Looking at (5.7) and recognizing that $\rho \bar{\pi}_z + (1 - \rho)\bar{\pi}_y = \bar{\pi} > cx_1$, we get

$$\dot{z}_2 < z_2 \left(\frac{\rho \pi_2 - cx_1}{c + \gamma x_3} \right) + \rho y_2 \left(\frac{\pi_2 + \gamma x_3}{c + \gamma x_3} \right) - (1 - \rho)z_2 \left(\frac{\gamma x_3}{c + \gamma x_3} \right)$$

while action 1 is the best response. Since this means $\pi_2 < c$, we can make it

$$\dot{z}_2 < z_2 \left(\frac{c(\rho - x_1)}{c + \gamma x_3} \right) + \rho y_2 - (1 - \rho) z_2 \left(1 - \frac{c}{c + \gamma x_3} \right).$$

Grouping terms together,

$$\dot{z}_2 < z_2 \left(\frac{c(1 - x_1)}{c + \gamma x_3} - (1 - \rho) \right) + \rho y_2.$$

We use $1 - x_1 = z_2 + y_2 + x_3 + x_4$ and the previously mentioned bounds $\gamma x_4 < c$ and $y_2 < y_2(t_c)$ to write

$$\frac{1 - x_1}{c + \gamma x_3} < \frac{z_2 + y_2(t_c)}{c + \gamma x_3} + \frac{1}{\gamma}.$$

For simplicity, we weaken this bound by dropping the γx_3 from the denominator of the first term. This leaves us with

$$\dot{z}_2 < z_2 \left(c \left(\frac{z_2 + y_2(t_c)}{c} + \frac{1}{\gamma} \right) - (1 - \rho) \right) + \rho y_2(t_c)$$

$$= (z_2)^2 + z_2 \left(y_2(t_c) + \frac{c}{\gamma} - (1 - \rho) \right) + \rho y_2(t_c).$$

Completing the square, we get

$$\dot{z}_2 < \left(z_2 - \frac{1}{2} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right) \right)^2 - \frac{1}{4} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right)^2 + \rho y_2(t_c).$$

So

$$\dot{z}_2 < 0 \text{ if } \left(z_2 - \frac{1}{2} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right) \right)^2 < \frac{1}{4} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right)^2 - \rho y_2(t_c).$$

We can be sure the right hand side here is positive because by inequality (5.79), $y_2(t_c)$ can be made as small as necessary. Then $\dot{z}_2 < 0$ if

$$(5.80) \qquad z_2 < \frac{1}{2} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right) + \sqrt{\frac{1}{4} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right)^2 - \rho y_2(t_c)}$$

and

$$z_2 > \frac{1}{2} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right) - \sqrt{\frac{1}{4} \left(1 - \rho - y_2(t_c) - \frac{c}{\gamma} \right)^2 - \rho y_2(t_c)}.$$

Recall that we have already shown that $y_2 + z_2 < \frac{c-1}{\beta}$ for $\ln\left(\frac{\gamma x_4(0)}{\beta+1}\right) \leq t \leq t_c$. This means inequality (5.80) holds right at $t = t_c$, and by its own power, it holds thereafter as well. Thus, for $t \geq t_c$,

$$z_2 \leq \max\left\{z_2(t_c), \frac{1}{2}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right) - \sqrt{\frac{1}{4}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right)^2 - \rho y_2(t_c)}\right\}.$$

Our upper bounds on y_2 and z_2 for $t \geq t_c$ give us an upper bound on π_2 for this time, and having already shown that $\beta\left(y_2(t_c) + z_2(t_c)\right) + 1 < c$, we now seek to establish that

$$\beta\left(y_2(t_c) + \frac{1}{2}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right) - \sqrt{\frac{1}{4}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right)^2 - \rho y_2(t_c)}\right) + 1 < c.$$

With some algebra, this is equivalent to

$$y_2(t_c) + \frac{1}{2}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right) - \frac{c-1}{\beta} < \sqrt{\frac{1}{4}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right)^2 - \rho y_2(t_c)}$$

or

$$\left(y_2(t_c) - \frac{c-1}{\beta} + \frac{1}{2}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right)\right)^2 < \frac{1}{4}\left(1 - \rho - y_2(t_c) - \frac{c}{\gamma}\right)^2 - \rho y_2(t_c)$$

or then

$$\left(y_2(t_c) - \frac{c-1}{\beta}\right)\left(-\frac{c-1}{\beta} + 1 - \rho - \frac{c}{\gamma}\right) < -\rho y_2(t_c).$$

Finally, recognizing that $y_2(t_c) - \frac{c-1}{\beta} < 0$, we can rewrite our desired inequality as

(5.81) $$1 - \rho - \frac{c}{\gamma} - \frac{c-1}{\beta} > \frac{\rho y_2(t_c)}{\frac{c-1}{\beta} - y_2(t_c)}.$$

But, using inequality (5.79), we see that

$$\frac{\rho y_2(t_c)}{\frac{c-1}{\beta} - y_2(t_c)} \leq \frac{\rho y_2(0)\left(\frac{\gamma x_4(0)}{c}\right)^{-y_3(0)}}{\frac{c-1}{\beta} - y_2(0)\left(\frac{\gamma x_4(0)}{c}\right)^{-y_3(0)}},$$

and this can be made as small as necessary. So inequality (5.81) does indeed hold, and $\pi_2 < c$ for $t \geq t_c$. Action 1 remains the best response for all $t > t_c$.

Because action 1 is the best response, we can use inequalities (5.17) and (5.19) from the proof of Theorem V.1, translated to an initial time of t_c instead of 0, to obtain for $t > t_c$,

$$y_1 \geq \frac{(1-\rho)y_1(t_c)e^{(1-\rho)(t-t_c)}}{1 - \rho - y_1(t_c) + y_1(t_c)e^{(1-\rho)(t-t_c)}}$$

and

$$z_1 \geq \rho - (\rho - z_1(t_c))\, e^{-(1-\rho)(t-t_c)} - \frac{\rho(1-\rho)(1 - \rho - y_1(t_c))}{y_1(t_c)}\, (t - t_c)e^{-(1-\rho)(t-t_c)}.$$

Thus, as $t \to \infty$, z_1 and y_1 approach ρ and $1 - \rho$ respectively. □

Proof of Theorem V.6.

Since it has been shown in Chapter IV that the best response dynamics finds $(1, 0, 0, 0)$ in this game (Golman and Page 2008b), it suffices to show that the hybrid dynamics finds $(0, 1, 0, 0)$ from almost any initial point.

Consider first $k(\vec{\pi}) = K$. Anywhere in the interior of the strategy space we can obtain $\beta x_2(0) + 1 > K$ and $\beta x_2(0) + 1 > c$ by taking $\vec{P}$ close enough to $\vec{\hat{P}}$. While $\pi_2 > K$, the hybrid dynamics of equation (5.9) give us

$$\text{(5.82)} \qquad \dot{x}_2 = x_2 \left(\frac{\pi_2 - \bar{\pi}}{\pi_{\max} - \pi_{\min}} + \sum_{j:\pi_j < K} x_j \frac{\pi_j - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \right).$$

While $\pi_2 > c$, $\pi_2 - \bar{\pi} > 0$, and the remaining terms in equation (5.82) are always positive, so $\dot{x}_2 > 0$. As x_2 increases, π_2 increases, so this holds forever.

Now assume $k(\vec{\pi}) = \alpha \pi_{\min} + (1 - \alpha)\pi_{\max}$ with $\alpha > 0$. Consider initial points with $x_4(0) > \frac{\beta+1}{\gamma} > \frac{c}{\gamma}$ and $x_2(0) > \frac{c-1}{\alpha\beta}$, conditions that can be satisfied anywhere in the interior of the strategy space by taking $\vec{P}$ close enough to $\vec{\hat{P}}$. These conditions imply that initially $\pi_3 > \pi_2 > \pi_1 > \pi_4$. And initially $k(\vec{\pi}) = -\alpha\gamma x_3 + (1 - \alpha)\gamma x_4$. If initially $\pi_2 \geq k$, then equation (5.82) applies, and $\dot{x}_2 > 0$. As x_2 increases, π_2 increases. Meanwhile, x_3 increases because $\pi_3 = \pi_{\max}$ and x_4 decreases because

$\pi_4 = \pi_{\min}$, so k decreases. Thus, π_2 remains bigger than k until it becomes $\pi_{\max}$. And once action 2 becomes best, it's clear that x_2 will continue to grow forever.

Now consider the case that initially $\pi_2 < k$. The hybrid dynamics, equation (5.9), then reduce to $\dot{x}_i = -x_3 x_i$ for $i \neq 3$ and $\dot{x}_3 = x_3(1 - x_3)$ while $\pi_i < k$. This is a time-scaled version of the best response dynamics, and all x_i have the same relative decay rate. Let the decreasing function $\lambda(t) = \frac{x_i(t)}{x_i(0)}$ while these dynamics apply. Then $\pi_2 = k$ when

$$\beta x_2(0)\lambda(t) + 1 = \gamma \left[(1 - \alpha)x_4(0)\lambda(t) - \alpha \left(1 - \lambda(t) \left(1 - x_3(0) \right) \right) \right].$$

This occurs when

$$\lambda = \frac{\alpha + \frac{1}{\gamma}}{(1 - \alpha)x_4(0) + \alpha(1 - x_3(0)) - \frac{\beta x_2(0)}{\gamma}}.$$

All we needed was $\lambda > \alpha$, because then $x_2(t) > \frac{c-1}{\beta}$. Thus, when $\pi_2 = k$, $\pi_2 > \pi_1$, and this validates our use of these dynamics until this point. The argument in the previous paragraph, for the case when $\pi_2 \geq k$, now applies. From here on, x_2 will always grow. $\qquad\square$

BIBLIOGRAPHY

[1] Alos-Ferrer, C., (1999). Dynamical Systems with a Continuum of Randomly Matched Agents. *J. Econ. Theory* 86, 245-267.

[2] Basov, S., (2004). Equilibrium Selection in Coordination Games: Why do dominated strategies matter? *Econ. Bull.* 3, 1-3.

[3] Binmore, K., Samuelson, L., (1999). Evolutionary Drift and Equilibrium Selection. *Rev. Econ. Stud.* 66, 363-394.

[4] Borgers, T., Sarin, R., (1997). Learning Through Reinforcement and Replicator Dynamics. *J. Econ. Theory* 77, 1-14.

[5] Boylan, R., (1992). Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals. *J. Econ. Theory* 57, 473-504.

[6] Camerer, C., Ho, T.H., (1999). Experience Weighted Attraction Learning in Normal Form Games. *Econometrica* 67, 827-874.

[7] Conlisk, J., (1980). Costly Optimizers Versus Cheap Imitators. *J. Econ. Behav. Organ.* 1 (3), 275-293.

[8] Ellison, G., (2000). Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Rev. Econ. Stud.* 67 (1), 17-45.

[9] Ely, J., Sandholm, W.H., (2005). Evolution in Bayesian Games I: Theory. *Games Econ. Behav.* 53, 83-109.

[10] Foster, D., Young, H.P., (1990). Stochastic Evolutionary Game Dynamics. *Theoretical Population Biology* 38, 219-232.

[11] Fudenberg, D., Levine, D.K., (1998). *The Theory of Learning in Games.* Cambridge: MIT Press.

[12] Gilboa, I., Matsui, A., (1991). Social Stability and Equilibrium. *Econometrica* 59, 859-867.

[13] Gilboa, I., Matsui, A., (1992). A Model of Random Matching. *J. Math. Econ.* 21, 185-197.

[14] Golman, R., Page, S., (2008a). Basins of Attraction and Equilibrium Selection Under Different Learning Rules. *J. Evol. Econ.* forthcoming.

[15] Golman, R., Page, S., (2008b). Individual and Cultural Learning in Stag Hunt Games With Multiple Actions.

[16] Hofbauer, J., (2000). From Nash and Brown to Maynard Smith: Equilibria, Dynamics and ESS. *Selection* 1, 81-88.

[17] Hofbauer, J., Schuster, P., Sigmund, K., (1979). A Note on Evolutionary Stable Strategies and Game Dynamics. *J. Theor. Biol.* 81, 609-612.

[18] Hofbauer, J., Sigmund, K., (2003). Evolutionary Game Dynamics. *Bull. Am. Math. Soc.* 40, 479-519.

[19] Hofbauer, J., Sorin, S., Viossat, Y., (2009). Time Average Replicator and Best-Reply Dynamics. *Math. Oper. Res.* forthcoming.

[20] Hommes, C., (2006). Heterogeneous Agent Models in Economics and Finance. In: Judd, K., Tesfatsion, L. (eds) Handbook of Computational Economics, vol 2, Agent-Based Computational Economics. North Holland, Amsterdam.

[21] Hopkins, E., (1999). A Note on Best Response Dynamics. *Games Econ. Behav.* 29, 138-150.

[22] Kandori, M., Mailath, G., Rob, R., (1993). Learning, Mutation and Long Run Equilibria in Games. *Econometrica* 61, 29-56.

[23] Kirman, A., (2006). Heterogeneity in Economics. *Journal of Economic Interaction and Coordination* 1, 89-117.

[24] Ochs, J., (1995). Coordination Problems. *Handbook of Experimental Economics.* Kagel, J., Roth, A. (eds) Princeton: Princeton University Press.

[25] Sandholm, W. H., (2009). Pairwise Comparison Dynamics and Evolutionary Foundations for Nash Equilibrium. Working paper, University of Wisconsin. http://www.ssc.wisc.edu/~whs.

[26] Schlag, K. H., (1998). Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-armed Bandits. *J. Econ. Theory* 78, 130-156.

[27] Swinkels, J. M., (1993). Adjustment Dynamics and Rational Play in Games. *Games Econ. Behav.* 5, 455-484.

[28] Taylor, P.D., Jonker, L., (1978). Evolutionarily Stable Strategies and Game Dynamics. *Math. Biosci.* 40, 145-156.

[29] Young, H.P., (1993). The Evolution of Conventions. *Econometrica* 61, 57-84.

CHAPTER VI

Conclusion

Analysis of the aggregate behavior of individual agents is central to many areas of social science as well as theoretical biology. In economics, for example, a population's demand curve is the pointwise sum of each consumer's willingness to buy at a given price. In political science, a variety of voting rules aggregate information in different ways to determine a collective choice. In evolutionary biology, the varying fitness of individual animals determines how the species as a whole reacts to selective pressures. In all of these instances, our interest is in macroscopic behavior, but our theories at the population level must be consistent with our models of individual agents' behavior.

This dissertation examines behavior in population games. The idea of a population game is that players don't know exactly who their opponents are in a strategic situation; instead, they face a pool of potential opponents and react to the population-wide profile of behavior called the population mixed strategy. We derive predictions for population mixed strategies by analyzing the rules individual agents use to choose actions. While the individual agents themselves are not perfectly rational as traditionally assumed in economic theory, a population may end up coordinating on a single preferred outcome or may allow a diversity of suboptimal

behaviors to persist.

Our analysis relies on a variety of mathematical tools. In modeling heterogeneous quantal responders, we introduce a probability density functional. The proofs of our main results in this chapter make use of characteristic functions for random variables and Jensen's Inequality. In other chapters, we model population learning rules as nonlinear dynamical systems. We are able to characterize basins of attraction by introducing and solving differential inequalities. Our finding of vanishing overlap in basins of attraction under different learning dynamics can be understood as simultaneous bifurcations in the two dynamical systems, involving the basin of attraction of different sets of equilibria shrinking away, and in the limit, producing distinct global attractors.

Chapter II in this dissertation contributes a theoretical result characterizing the representative agent for a population of heterogeneous quantal responders and a more applied result describing bias in single-agent logit response models. Almost all applications fitting quantal response equilibrium to data have assumed iid payoff disturbances; our theoretical result is that the representative agent for a population of heterogeneous quantal responders often will not have an iid distribution of payoff disturbances. This suggests we should allow noise terms that are jointly dependent across actions. Our finding of downward bias in a mis-specified homogeneous logit parameter means that this frequently used single-agent model underestimates the average level of rationality in a heterogeneous player pool.

Chapter III shows how equilibrium selection depends on how agents learn. The finding that basins of attraction under different learning rules can share vanishing overlap lays the foundation for the analysis in subsequent chapters. In Chapter III, we focus on characterizing when this can occur. A necessary condition, worth high-

lighting for its simplicity, is that with probability one the initial best response cannot be a pure, uniformly evolutionarily stable strategy for the sequence of games. The existence of a parasitic or misleading action that serves as a temporary best response allows subtle differences in learning rules to accumulate.

Chapter IV focuses on a class of generalized stag hunt games, which provide a window into the evolution of cooperation. In these games, agents can choose from among multiple potentially cooperative actions or can take a secure, self interested action. We assume that a proportion of the cooperative actions prove effective, while others can be undermined by a predatory action. We find that the replicator dynamics, which relies on cultural learning, leads to an efficient cooperative outcome, whereas the individualistic best response dynamics does not. We caution, however, that this does not imply that cultural learning outperforms individualistic belief-based learning in all games. On the contrary, simple transformations of payoffs that do not affect the paths of the learning dynamics, such as adding a constant to all entries in a single column of the payoff matrix (in a symmetric game), can determine which of multiple equilibria is payoff dominant, i.e., most efficient for everybody involved. The strategic context a given player faces is invariant to a uniform change in how one of his actions affects other players' payoffs. Such a change thus cannot affect the learning dynamics, even while it can determine the efficiency of an equilibrium.

Chapter V can be viewed as a sensitivity analysis of equilibrium selection in relation to the specification of the learning rule. Indeed, we find that small changes in the way agents learn can produce drastic changes in outcomes. We also establish that convex combinations of different dynamics can select different equilibria than either of the two dynamics select alone.

The general lesson here is that predictions for outcomes in games require a com-

plete and precise understanding of how players learn and err, and for such a model to be true to life, it must capture the heterogeneity of different actors.