Equilibrium Selection in the Stochastic Iterated Prisoner's Dilemma

Susan M. Bennett

SFI WORKING PAPER: 1996-08-062

Equilibrium Selection in the Stochastic Iterated Prisoner's Dilemma

Susan M. Bennet

Department of Economics
Simon Fraser University
Burnaby, B.C., Canada
V5A 1S6

sbennet@sfu.ca

I. Introduction

> Many of the benefits sought by living things are disproportionately available to cooperating groups. While there are considerable differences in what is meant by the terms "benefits" and "sought", this statement, insofar as it is true, lays down a fundamental basis for all social life..." (Axelrod, 1984)

The Prisoner's Dilemma has received a great deal of attention in the literature because it captures the problems associated with fostering cooperation in social life. The game itself can be summarized by the following payoff matrix;

		Player B	
		Cooperate	Defect
Player A	Cooperate	c_1, c_2	s_1, t_2
	Defect	t_1, s_2	d_1, d_2

where $t_i > c_i > d_i > s_i$ and $(c_i + c_i) > (t_i + s_i)$. The temptation payoff, t_i, is greater than the payoff to agents if everyone cooperates, c_i. This cooperative payoff in turn is greater than d_i, the punishment for mutual defection, which is greater still than the sucker's payoff s_i. The second inequality simply restricts the payoffs to exclude the possibility that players may do better to alternate between being the "sucker" than to always cooperate.

For each player, Defect (D) dominates Cooperate (C). No matter what his opponent does, Defect is the player's unique best response. Thus, the iterated deletion of dominated strategies gives a unique answer to the one-shot PD game, namely {D, D}. This constitutes a Nash equilibrium, since no player, taking the strategy of his opponent as given, wishes to change his own strategy choice. Notice that while individually rational play prescribes the {D,D} outcome, this equilibrium is Pareto dominated by {C,C}, the collectively optimal outcome. It is this [tension] between individually rational behavior and the collectively optimal outcome that makes the PD game so interesting to economists.

Of course, many social interactions are not of the one-shot variety, but are rather played repeatedly. Thus, there has been great interest in studying repeated games, where players meet over an infinite (indefinite) horizon[1] to play the one-shot (stage) game. These games are of particular interest as they allow for reputation building, retaliation, and, in some instances, cooperation. There exists an extensive literature on these IPD game[2], focused largely, however,

[1] There is a substantial literature on the finitely repeated IPD, which will not be discussed in this paper. See for instance Selten, R. (1978) "The Chain-Store Paradox," *Theory and Decision* 9:127-159; Selten, R. and Stoecker, R. (1986), "End Behavior in Sequences of Finite Prisoner's Dilemma Supergames," *Journal of Economic Behavior and Organization* 7:47-61; or Nachbar (1992) "Evolution in the Finitely Repeated Prisoner's Dilemma," *Journal of Economic Behavior and Organization* 19:307-326.
[2] See for instance Friedman (1971) for theoretical work, and Axelrod (1984) for overview of theory and description of evolutionary approach.

on a fixed payoff game. That is, while the history of play may change over time, the payoffs do not.

Is this a useful way to represent social dilemmas? In some cases perhaps. I argue, however, that the nature of most repeated interactions is better captured in games where payoffs change with each meeting. Such a game is referred to as a Stochastic Iterated Prisoner's Dilemma (SIPD). Little is known about such games[3]. This paper attempts to shed light on the nature of these games, in particular on their equilibrium properties.

II. The Stochastic Iterated Prisoner's Dilemma (SIPD)

A. Motivation

It is not difficult to look out the window and find examples of games where stochastic payoffs - play an important role. One interesting example comes from recent work by Eaton and Eswaran (1996), who discuss know-how sharing within the context of an SIPD. In their model, producers discover better ways of producing their goods each period, with incremental innovations being drawn randomly from an exogenously specified distribution. The important thing to note is that producers never discover the same thing twice. Further, the value of their discoveries (to themselves and to rival producers) changes from period to period. The players problem under these circumstances is considerably more complex than under a fixed payoff scenario. Specifically, their decision over whether or not to share their discoveries with rival producers must take into account not only the possibility of future interactions[4], but also the stochastic nature of innovation.

Stochasticity also plays an important role in many social interactions of interest. Academics regularly exchange favors with colleagues; we read drafts of papers, help with computing difficulties, etc. The payoffs associated with these exchanges of favors, however, are not constant, but change from day to day. The same is true of the (anonymous) interactions we have in driving to work. How late am I in leaving the house? How important is the meeting I'm going to be late for? The game's payoffs - the cooperative payoff, and the payoff associated with defecting - will depend on the answer to these questions, which will change from day to day.

Are these considerations important? Do the equilibrium properties of the these stochastic games tell us anything about social dilemma's that we haven't already learned from study of the fixed-payoff IPD game? This paper answers 'yes' to both these questions.

B. Modeling

There are a number of approaches one can take toward modeling this game. One logical starting point is a standard game theoretical approach. Here, agents are modeled as fully rational, able to compute the subgame perfect Nash equilibrium of the game. However, in their recent paper, Eaton and Eswaran look at the SIPD game from this game theoretic point of view, and find that,

[3]I am aware of only one piece of research on this topic - see Eaton and Eswaran (1996).

[4]It is this possibility of meeting again which makes cooperation possible. Axelrod (1984) writes that "The future can therefore cast a shadow back upon the present, and thereby affect the current strategic situation." (p.12)

like its fixed payoff relative, the SIPD game is plagued by the problem of multiple equilibria. The standard notion of equilibrium then has no predictive (or prescriptive) power.

In the extant literature on fixed payoff IPD games, the response to the multiplicity problem has been to use the notion of boundedly rational agents to try and select among equilibria[5]. In context of standard IPD games, the work by Axelrod (1984, 1987), Marks (1989), and Foster and Young (1991) has provided some interesting equilibrium selection results. This notion of bounded rationality as an equilibrium selection criterion has been used in other games as well. Kandori, Mailath and Rob (1993), Young (1993) and Bergin and Lipman (1994) are all good examples of this literature. The remainder of this paper uses the notion of bounded rationality in a similar fashion. Specifically, the hope is that boundedly rational agents will coordinate on a unique equilibrium in the SIPD game, thus providing a resolution to the multiplicity of equilibria problem.

C. The Game

i. Agent Representation: Because payoffs are random, an agent's sharing decision must be a function of the payoff realization. Imagine a distribution of realizations, f(x), such as the one illustrated in figure 1a.

Each player will have a decision rule, X_g, which will stand for those realizations over which they are willing to cooperate. Players also need to form expectations regarding the behavior of others. Specifically, they need to decide at what point to cease cooperating with an opponent, changing to the Nash strategy of always defect[6]. This decision rule is embodied in the variable X_r. An example from the know-how sharing discussion will be helpful here. Each period a producer realizes an innovation, drawn from an exogenously specified distribution like f(x) illustrated above. Let $\{X_g^i, X_r^i\} = \{0.75, 0.66\}$. This means that agent i will share any realization less than or equal to 0.75, and will expect his opponents to share any realization less than or equal to 0.66. If his expectations are disappointed, i.e. if $X_g^j < 0.66$ and $X_g^j < X_j \leq 0.66$ (so that we are in region A in figure 1a) then agent i reverts to the Nash equilibrium of the stage game, namely always defect. Each agent can then be entirely defined by his two attributes, $\{X_g^i, X_r^i\}$.

ii. The Distribution: A Simple Discrete Case with Player Types: Our ultimate goal is to allow agents to interact in a more complex environment, with a possibly continuous distribution of payoff realizations. Using a simple distribution to begin with, however, will help us to examine some of the interesting dynamics of the game that are not as obvious in the more complex environment. Consider then the simple discrete distribution shown in figure 1b. Here, Pr(0) = Pr(2/3) = 1/2. The expected value of each drawing from this distribution is 1/3.

[5]The use of equilibrium refinements, such as Properness, or Strategic Stability are of no use in reducing the set of possible equilibria to these games.

[6]This permanent retaliation assumption, sometimes called a grim strategy, is consistent with the equilibrium suggested by Friedman (1971).

Using this simple scenario, it is possible to categorize players in terms of their *type* [7]

Type I: $X_g \geq 0.6667$ and $X_r \geq 0.6667$
Type II: $X_g \geq 0.6667$ and $X_r < 0.6667$
Type III: $X_g < 0.6667$ and $X_r \geq 0.6667$
Type IV: $X_g < 0.6667$ and $X_r < 0.6667$

We can combine Types III and IV into a single player type, called III, since the cooperative (or noncooperative, as the case may be) behavior of these two types is indistinguishable in this environment. Consider for a moment a Type I player. He might be described by an $\{X_g, X_r\}$ combination such as $\{0.9, 0.8\}$. Given the simple, discrete nature of the distribution that is being considered here, this player will share any non-zero innovation, and expects any 'opponent's' to do likewise. We call the Type I player a *demanding cooperator*, since he is willing to behave cooperatively as long as those with whom he interacts are also cooperative. A Type II player, on the other hand, might be described by $\{0.8, 0.3\}$. This player, given the above distribution, shares all non-zero innovations, but does not ask much of his fellow players. In fact, in this simple world, a Type II player never switches into retaliatory mode. Type II players are thus categorized as *cooperative* but *undemanding*. Using the same reasoning, the Type III player is described as a *defector* (his X_r parameter is behaviorally unimportant; the degree to which this player is 'demanding' is never expressed).

For the purposes of computational experiments with this system, we need to define payoffs explicitly. Consider again the know-how sharing example. Here, payoffs can be written as

$$\pi(i,j) = \alpha(S_i) - \beta(S_j) \quad \alpha > \beta$$

The payoffs for each possible type interaction are summarized in the following table:

		Player j		
		I	II	III
	I	$2N(\alpha - \beta)/3$	$2N(\alpha - \beta)/3$	$\Sigma_i^N(1/2^{i+1})(4\alpha-8\beta)/3 + \Sigma_i^N(1/2^i)(N-i)(\alpha - \beta)/3$
Player i	II	$2N(\alpha - \beta)/3$	$2N(\alpha - \beta)/3$	$N(\alpha-2\beta)/3$
	III	$\Sigma_i^N(1/2^{i+1})(8\alpha-4\beta)/3 + \Sigma_i^N(1/2^i)(N-i)(\alpha - \beta)/3$	$N(2\alpha - \beta)/3$	$N(\alpha - \beta)/3$

[7]Note that a player's type is a decision variable, which is not the case in the signaling literature. There, agents know their type, but for whatever reason will not reveal this to others. This is important in the literature on insurance markets, etc. In this paper, however, a player's type changes over time, depending on his relative fitness.

where payoffs are for player i. N is the number of interactions between players each time they meet[8]. As N falls, the effectiveness of punishment falls, since there is a shorter time horizon over which players revert to Nash play. Note that N is not known by any player, so that this game remains an indefinitely repeated SIPD.

In the absence of any type III players, types I and II do equally well on average. This observation will be important in subsequent discussion, as it has implications for the stability of equilibria in this game.

Using this type simplification, we can represent any possible population of players in a simplex, as in figure 2. The top vertex represents an all Type-I population. The other vertices have analogous interpretations. The choice of behavioral dynamics, which determines the aggregate population dynamic, will dictate the way in which the system evolves through this space.

Note that, using this simple distribution/type space, there are two equilibria which are consistent with the notion of subgame perfection: One equilibrium is at the top vertex, where all players are Type I (called a Type I equilibrium); the other has all players choosing to play Type III, which is at the bottom left vertex (called a Type III equilibrium). The next section attempts to use the techniques discussed above to resolve this multiplicity problem.

III. Boundedly Rational Agents Play the Simple SIPD

A. Myopic Best Response Agents

i. Updating strategies.

Suppose the action chosen by a player is a best reply given the actions chosen by the other players in the previous period. This will be referred to as a myopic best-response learning model (c.f. Cournot conjecture). The idea is that, in a complicated world, agents cannot calculate best responses to their stage environment. The myopia assumption amounts to saying that at the same time that players are learning, they are not taking into account the long run implications of their strategy choice.[9]

It isn't hard to see that the number of agents playing each strategy in a given period completely determines the future evolution of the strategies chosen here. Hence, we can represent this evolutionary dynamic as a Markov process with a state space given by the number of agents playing any strategy i.

ii. Basins of Attraction and Drift.

We start simply by assuming that there are only three players. If there are three possible types for each player, this leaves us with ten possible states in the simplex.

[8]Play takes place in a tournament fashion, so that each player *meets* every other once, and *interacts* N times with each meeting.

[9]See for example Bergin and Lipman (1994) for another example of this learning dynamic.

State		
1	(3, 0 ,0)	This means that 3 players are of Type I, 0 of Type II and 0 of Type III.
2	(2, 0, 1)	
3	(2, 1, 0)	
4	(1, 0, 2)	
5	(1, 1, 1)	
6	(1, 2, 0)	
7	(0, 0, 3)	
8	(0, 1, 2)	
9	(0, 2, 1)	
10	(0, 3, 0)	

where each state can be located on the Simplex, as illustrated in Figure 2.

Using the myopic best response assumption, I calculated the transition matrix P, which looks as follows:

	1 **(.125)**	0	0 **(.375)**	0	0	0 **(.375)**	0	0	0	0 **(.125)**
	0.5	0	0.5	0	0	0	0	0	0	0
	0 **(.125)**	0	1 **(.375)**	0	0	0 **(.375)**	0	0	0	0 **(.125)**
	0	1	0	0	0	0	0	0	0	0
P =	0	.5	0	0	.5	0	0	0	0	0
	0	0 **(.25)**	0	0	0 **(.5)**	0	0	0	1 **(.25)**	0
	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	0	1	0	0	0

Each element represents the probability of transiting from state[row] to state[column]. So, to find the probability of moving from state 5 to state 2, look at element (5,2), and find $P_{5,2} = 0.5$.

There are actually two transition matrices here. The main matrix has embedded in it an inertia assumption which is not present in the modified matrix, represented in bold. Consider for example state 3 (2,1,0). There are no uncooperative players in the population, so that Types I and II do equally well. Moreover, there is no incentive for any player to change strategies to type III. Under these circumstances then, type I and II strategies are equally attractive; there is no unique best response for either type I or type II players. There is then the possibility that the population will *drift* through some region of the strategy space. Drift can be illustrated by reference to figure 2. Here, the non-uniqueness problem means that the population moves back and forth along the I-II locus.

We can make two polar assumptions at this point. One is an *inertia assumption*. which essentially assumes away drift. Faced with an equally attractive alternative strategy, a player will choose to continue playing his current strategy (type) will probability 1. Alternatively, we could make the assumption that players like to *experiment*, in which case the player faced with an equivalent-payoff strategy will choose to adopt the alternative strategy with probability 1. Note that experimentation here is distinct from mutation. It is important to keep in mind that these are completely distinct notions.

These two polar assumptions lead to very different population dynamics. The transition probabilities given in the primary table correspond to the inertia assumption, while those bracketed in bold correspond to the experimentation assumption. Under the inertia assumption (which is sometimes referred to as 'stickiness'), states 1, 3, and 7 are stable i.e. (3,0,0), (2,1,0), and (0,0,3). See figure 3a. When we relax the inertia assumption, and allow players to experiment with equally attractive strategies, neither state 1 nor 3 is an attractor. All states fall in the basin of attraction of 7 i.e. (0,0,3). Why does 7 become the only stable steady state? The blame rests squarely on the shoulders of drift. Over a long enough time horizon, the system moves with probability 1 into the basin of attraction of state 7 (0,0,3). Once here, there is nothing to break this equilibrium, and it stays here forever. This is illustrated in figure 3b.

iii. Asymptotic Equilibria: Adding Mutation.

We are interested in looking at the asymptotic properties of this system. This means that we want to allow for a small rate of mutation, and then ask: after a long period of time, at which equilibrium (equilibria) does the system spend most of its time? We also want to know how much time is spent in transition between equilibria, vs. at an equilibrium[10].

Assume that the probability of a mutation is some fixed ε, independent of time, the current state, and the agent. If an agent does not mutate, he changes strategy or not according to the best response dynamics described above. If he does mutate, he changes strategy with some fixed probability. So, with probability $(1-\varepsilon)$ agents choose their type according to the best response dynamic previously described. With probability ε the agent randomizes over the remaining two types.[11]. The transition probabilities are shown in Appendix A.

Using the *inertia assumption*, the following fixed state vector summarizes the asymptotic equilibrium of this simple ten-state system:

$$\mathbf{q}^* = [0.313, 0, 0.402, 0, 0, 0, 0.268, 0, 0, 0].$$

This fixed state vector indicates that the system spends about 30 percent of its time in state 1 [3,0,0], 40 percent of its time in state 3 [2,1,0], and just under 30 percent of its time in state 7 [0,0,3]. The rest of the time (1.7 percent) is spent in transition from one equilibrium to another. These results are initially quite interesting, since the system appears to spend a lot of time is state 3. Upon reflection, however, it is not so surprising. By invoking the inertia assumption, drift is essentially removed, so that state 3 is no longer a precarious place to be.

When this same system is run over a long horizon under the *experimentation assumption*, the results are markedly different. In particular, for a sufficiently small mutation rate, the system

[10] Kandori, Mailath and Rob (1993) and Bergin and Lipman (1994) provide excellent discussions this notion of asymptotic equilibrium.

[11] If the agent does not face a unique best response, then with probability ε he chooses to play the remaining (payoff inferior) type.

spends almost all of its time in state 7, where sharing is almost non-existent. Notice that these results, where $\varepsilon \to 0$, coincide with those found using the deterministic case, $\varepsilon = 0$.

C. Imitative Players

While defensible, the assumption that players are myopic best-response decision makers is not altogether satisfying. In particular, it still assumes a level of computational sophistication on the part of players that is not entirely realistic. We might want then to explore the dynamics of the simple 3-player population under the assumption that decisions are made *imitatively.* Under an imitative (or adaptive) learning assumption, players look not at potential payoffs, but rather at the current fitness (or payoffs) of existing players.

Under this learning assumption, players are less sophisticated, in that they do not know how to calculate best replies and are using other players' successful strategies as guides for their own choices. These naive players sometimes observe the current performance of other players, and simply mimic the most successful strategy. This captures an aspect of learning that is considered important, namely imitation, or emulation. People learn what are good strategies by observing what has worked well for others[12].

i.. Updating.

The updating of strategies takes place in the following way. After each round of play, payoffs are calculated, and each player in the population is polled. They are given the opportunity to update their strategy by drawing randomly one other player, whose type and corresponding fitness they may then know. If this reference player's fitness is higher, then with probability 1 his strategy is adopted by the polled player. If his fitness is lower, then the polled player does not change strategies.

The issue of drift becomes important if a reference player's fitness is equal to that of the polled player. As in the myopic best response case, we need to think about how ties are to be broken. Are agents inclined toward experimentation, or do they require a strict increase in payoffs to induce them to change strategies? Again as in the previous case, we make two polar assumptions. In the first instance, ties are broken in favor of inertia - that is, a player adopts a new strategy only if there is an unambiguous increase in his payoff. In the second instance, we adopt the alternative assumption, that players are experimenters by nature, and thus try a new strategy as long as it provides at least the same payoff as the current strategy.

ii. Basins of Attraction.

We can again construct a transition table, and find the stationary points of this system. Figures 4a and 4b summarize the results under these two assumptions. In case 1, under the inertia assumption, there are 5 equilibria, indicated by the open circles. Note that (1,1,1) lies in the basin of attraction of both I (S1) and III. (S7). In case two, where we assume that agents prefer to experiment, the number of equilibria is reduced to three. Again, there are initial points, such as (1,1,1) which lie in the basin of more than one equilibrium. In both these scenarios, it is important to note the probabilistic element that is driving these results. While each agent is polled (i.e. asked to update their strategy) with probability 1, he selects a reference player

[12] See for instance Routledge (1995).

probabilistically. Thus we find that, starting from state 3 [2,1,0], with probability 15/16 the system ends up at the steady state [3,0,0], while with probability 1/16 this same initial point falls in the basin of attraction of [0,3,0].

It may seem odd that states 6 [1,2,0] and 10 [0,3,0] do not lie in the basin of attraction of the Type III steady state (state 7). This is because of the nature of the imitative dynamic. In order for a player to adopt a new strategy when polled, he must choose a player who is currently playing a strategy different from his own. In state 10, there are only Type II players; there are no Type III players (or Type I players for that matter) whom the Type II player can imitate. Similarly, in state 6, the only possibility is for players to change from Type I to Type II and visa versa. Since there are no Type III's, there is no way to introduce this strategy to the population without mutation This is true under both the inertia and the experimentation assumptions.

iii. Adding Mutation.
Where does the system, on average, spends most of its time when it is subject to small, random perturbations? The interesting result here is that inertia plays no role in determining the long run behavior of a system. When agents are modeled as adaptive, the long run equilibrium is in the neighbourhood of Type III, independent of the rate at which agents experiment. Different rates of experimentation influence the amount of time it takes to arrive in the type III equilibrium. However, once there, it is exceeding difficult to leave this equilibrium. Why? Under the adaptive learning assumption, agents look only at current fitness. Type I players do well only if they can find other cooperative players. With only three players, this requires that two-thirds of the population mutate from Type III to Type I, a zero probability event as $\varepsilon \to 0$. Thus, once in the neighbourhood of the Type III equilibrium, the system will stay there indefinitely. On the other hand, mutation will move the population along the I-II locus[13], even in the absence of experimentation, leading eventually to the Type-III equilibrium.

IV. Summary

The results discussed above are summarized in Figure 5. What has this told us about the SIPD game? The starting point of this investigation was the multiplicity of equilibria with which we are confronted when the standard, game theoretic notion of (subgame perfect) equilibrium is used. Does the use of boundedly rational agents in this model help to resolve the multiplicity problem? Not really. By judicious choice of parameter values (for mutation and experimentation), combined with the 'appropriate' operationalization of bounded rationality, we can influence the amount of time spent in either of the two equilibria.

This leads to a more general question, which goes beyond the simple SIPD game. Is it possible that the notion of asymptotic equilibria can be used as an equilibrium selection criterion? There is a growing literature which seems to answer 'yes' to this question, stemming largely from the work of Kandori, Mailath and Rob (1993) and Young (1993). The suggestion from this literature

[13] Notice that mutation is modeled slightly differently here. Under the Myopic Best Response dynamic, mutation was to strictly payoff inferior strategies. If a player adopted an equal-payoff strategy, it was due solely to experimentation. Here, using the imitative dynamic, drift can occur via both experimentation and mutation. Future work will address this modeling problem.

is that, as the mutation rate approaches zero, the limit distribution of a system selects the salient equilibrium.

In this regard, the lessons from this paper are two. First, I show that, while the notion of asymptotic stability is useful, it cannot be applied without regard to the nature of the game's equilibria. In the simple model outlined in this paper, one of the equilibria is neutral because of the possibility of drift. We are thus able to choose the asymptotic equilibrium by judicious choice of the experimentation parameter. This indicates that one must proceed with care in applying the results of the equilibrium selection literature. This caveat is consistent with the recent work by Bergin and Lipman (1995).

Second, the results from this work indicate that the way in which learning is modeled plays an important role in determining the aggregate dynamics of a system, and thus in determining the long run, or asymptotic equilibrium of any game. The question we must ask here is then whether or not there is any criterion which can guide us in the modeling of learning, and thus the evolution of behavioral rules.

A number of issues, specific to the SIPD game, should be discussed briefly in closing. First is the question of population size. Do these results depend on the small size of the population considered here? The answer is a qualified 'no'. The small population provided a useful heuristic device for exploring the notion of drift, which is crucial in the SIPD game, and remains so *independent of the size of the population* or the complexity of the payoff distribution. However, the use of a small population makes the equilibrium results 'hyper-sensitive' to the way in which bounded rationality is operationalized. Both of these issues are addressed in Bennet (1996), forthcoming.

Another important issue to consider is the nature of drift and the underlying rate of experimentation which drives it. If drift is responsible for the breakdown of cooperation, then one might expect agents to learn the optimal rate of drift. This too is investigated in Bennet (1996). Finally, it is a worthwhile exercise to consider some of the equilibrium notions from the evolutionary game theory literature, an exercise which is undertaken as well in Bennet (1996).

Appendix A.

The epsilon transition matrix under the myopic best response assumption **(randomization assumption indicated in bold)**

$P_{1,1} = (1-\varepsilon)^3$	$\mathbf{(1-\varepsilon)^3/8}$
$P_{1,2} = 3\varepsilon(1-\varepsilon)^2/2$	$\mathbf{3\varepsilon(1-\varepsilon)^2/4}$
$P_{1,3} = 3\varepsilon(1-\varepsilon)^2/2$	$\mathbf{3(1-\varepsilon)^3/8}$
$P_{1,4} = 3\varepsilon^2(1-\varepsilon)/4$	$\mathbf{3\varepsilon^2(1-\varepsilon)/2}$
$P_{1,5} = 6\varepsilon^2(1-\varepsilon)/4$	$\mathbf{3\varepsilon(1-\varepsilon)^2/2}$
$P_{1,6} = 3\varepsilon^2(1-\varepsilon)/4$	$\mathbf{3(1-\varepsilon)^3/8}$
$P_{1,7} = \varepsilon^3/8$	$\mathbf{\varepsilon^3}$
$P_{1,8} = 3\varepsilon^3/8$	$\mathbf{3\varepsilon^2(1-\varepsilon)/2}$

$P_{1,9} = 3\varepsilon^3/8$ **$3\varepsilon(1-\varepsilon)^2/4$**
$P_{1,10} = \varepsilon^3/8$ **$(1-\varepsilon)^3/8$**
$\Sigma_j P_{1,j} = 1.0$

$P_{2,1} = (1-\varepsilon)^3/2$
$P_{2,2} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{2,3} = (1-\varepsilon)^3/2 + \varepsilon(1-\varepsilon)^2/2$
$P_{2,4} = 9\varepsilon^2(1-\varepsilon)/8$
$P_{2,5} = 5\varepsilon^2(1-\varepsilon)/4 + \varepsilon(1-\varepsilon)^2/2$
$P_{2,6} = \varepsilon^2(1-\varepsilon)/8 + \varepsilon(1-\varepsilon)^2/2$
$P_{2,7} = \varepsilon^3/4$
$P_{2,8} = \varepsilon^3 + \varepsilon^2(1-\varepsilon)/8$
$P_{2,9} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/4$
$P_{2,10} = \varepsilon^2(1-\varepsilon)/8$
$\Sigma_j P_{2,j} = 1.0$

$P_{3,1} = \varepsilon(1-\varepsilon)^2/2$ **$(1-\varepsilon)^3/8$**
$P_{3,2} = \varepsilon(1-\varepsilon)^2/2 + \varepsilon^2(1-\varepsilon)/2$ **$3\varepsilon(1-\varepsilon)^2/4$**
$P_{3,3} = (1-\varepsilon)^3 + \varepsilon^2(1-\varepsilon)/2$ **$3(1-\varepsilon)^3/8$**
$P_{3,4} = \varepsilon^3/8 + \varepsilon^2(1-\varepsilon)/2$ **$3\varepsilon^2(1-\varepsilon)/2$**
$P_{3,5} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/2 + \varepsilon(1-\varepsilon)^2$ **$3\varepsilon(1-\varepsilon)^2/2$**
$P_{3,6} = \varepsilon^3/8 + \varepsilon(1-\varepsilon)^2$ **$3(1-\varepsilon)^3/8$**
$P_{3,7} = \varepsilon^3/8$ **ε^3**
$P_{3,8} = \varepsilon^2(1-\varepsilon)/4 + \varepsilon^3/4$ **$3\varepsilon^2(1-\varepsilon)/2$**
$P_{3,9} = \varepsilon^2(1-\varepsilon)/2 + \varepsilon^3/8$ **$3\varepsilon(1-\varepsilon)^2/4$**
$P_{3,10} = \varepsilon^2(1-\varepsilon)/4$ **$(1-\varepsilon)^3/8$**
$\Sigma_j P_{3,j} = 1.0$

$P_{4,1} = \varepsilon(1-\varepsilon)^2/2$
$P_{4,2} = (1-\varepsilon)^3 + \varepsilon^2(1-\varepsilon)/2$
$P_{4,3} = \varepsilon(1-\varepsilon)^2/2 + \varepsilon^2(1-\varepsilon)/2$
$P_{4,4} = \varepsilon^3/8 + \varepsilon(1-\varepsilon)^2$
$P_{4,5} = \varepsilon^2(1-\varepsilon)/2 + \varepsilon^3/4 + \varepsilon(1-\varepsilon)^2$
$P_{4,6} = \varepsilon^3/8 + \varepsilon^2(1-\varepsilon)/2$
$P_{4,7} = \varepsilon^2(1-\varepsilon)/4$
$P_{4,8} = \varepsilon^3/8 + \varepsilon^2(1-\varepsilon)/2$
$P_{4,9} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/4$
$P_{4,10} = \varepsilon^3/8$
$\Sigma_j P_{4,j} = 1.0$

$P_{5,1} = \varepsilon(1-\varepsilon)^2/4$
$P_{5,2} = (1-\varepsilon)^3/2 + 5\varepsilon^2(1-\varepsilon)/8$

$P_{5,3} = \varepsilon^2(1-\varepsilon)/8 + \varepsilon(1-\varepsilon)^2/2$
$P_{5,4} = 5\varepsilon(1-\varepsilon)^2/4 + \varepsilon^3/4$
$P_{5,5} = \varepsilon^3/4 + 5\varepsilon^2(1-\varepsilon)/8 + \varepsilon(1-\varepsilon)^2/4 + (1-\varepsilon)^3/2$
$P_{5,6} = \varepsilon(1-\varepsilon)^2/4 + \varepsilon^2(1-\varepsilon)/4$
$P_{5,7} = \varepsilon^2(1-\varepsilon)/2$
$P_{5,8} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/2 + \varepsilon(1-\varepsilon)^2/4$
$P_{5,9} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/8 + \varepsilon(1-\varepsilon)^2/4$
$P_{5,10} = \varepsilon^2(1-\varepsilon)/8$
$\boldsymbol{\Sigma_j P_{5,j} = 1.0}$

$P_{6,1} = \varepsilon^3/8$	$\boldsymbol{\varepsilon(1-\varepsilon)^2/8}$
$P_{6,2} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/4$	$\boldsymbol{(1-\varepsilon)^3/4 + \varepsilon^2(1-\varepsilon)/2}$
$P_{6,3} = \varepsilon^3/8 + \varepsilon^2(1-\varepsilon)/2$	$\boldsymbol{3\varepsilon(1-\varepsilon)^2/8}$
$P_{6,4} = \varepsilon^3/8 + \varepsilon^2(1-\varepsilon)/2$	$\boldsymbol{\varepsilon^3/2 + \varepsilon(1-\varepsilon)^2}$
$P_{6,5} = \varepsilon^3/4 + \varepsilon^2(1-\varepsilon)/2 + \varepsilon(1-\varepsilon)^2$	$\boldsymbol{(1-\varepsilon)^3/2 + \varepsilon^2(1-\varepsilon)}$
$P_{6,6} = \varepsilon^2(1-\varepsilon)/2 + \varepsilon(1-\varepsilon)^2/2$	$\boldsymbol{3\varepsilon(1-\varepsilon)^2/8}$
$P_{6,7} = \varepsilon^2(1-\varepsilon)/4$	$\boldsymbol{\varepsilon^2(1-\varepsilon)}$
$P_{6,8} = \varepsilon^3/8 + \varepsilon(1-\varepsilon)^2$	$\boldsymbol{\varepsilon(1-\varepsilon)^2 + \varepsilon^3/2}$
$P_{6,9} = \varepsilon^2(1-\varepsilon)/2 + (1-\varepsilon)^3$	$\boldsymbol{(1-\varepsilon)^3/4 + \varepsilon^2(1-\varepsilon)/2}$
$P_{6,10} = \varepsilon(1-\varepsilon)^2/2$	$\boldsymbol{\varepsilon(1-\varepsilon)^2/8}$
$\boldsymbol{\Sigma_j P_{6,j} = 1.0}$	

$P_{7,1} = \varepsilon^3/8$
$P_{7,2} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{7,3} = 3\varepsilon^3/8$
$P_{7,4} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{7,5} = 3\varepsilon^2(1-\varepsilon)/2$
$P_{7,6} = 3\varepsilon^3/8$
$P_{7,7} = (1-\varepsilon)^3$
$P_{7,8} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{7,9} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{7,10} = \varepsilon^3/8$
$\boldsymbol{\Sigma_j P_{7,j} = 1.0}$

$P_{8,1} = \varepsilon^3/8$
$P_{8,2} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{8,3} = 3\varepsilon^3/8$
$P_{8,4} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{8,5} = 3\varepsilon^2(1-\varepsilon)/2$
$P_{8,6} = 3\varepsilon^3/8$
$P_{8,7} = (1-\varepsilon)^3$
$P_{8,8} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{8,9} = 3\varepsilon^2(1-\varepsilon)/4$

$P_{8,10} = \varepsilon^3/8$
$\Sigma_j P_{8,j} = 1.0$

$P_{9,1} = \varepsilon^3/8$
$P_{9,2} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{9,3} = 3\varepsilon^3/8$
$P_{9,4} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{9,5} = 3\varepsilon^2(1-\varepsilon)/2$
$P_{9,6} = 3\varepsilon^3/8$
$P_{9,7} = (1-\varepsilon)^3$
$P_{9,8} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{9,9} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{9,10} = \varepsilon^3/8$
$\Sigma_j P_{9,j} = 1.0$

$P_{10,1} = \varepsilon^3/8$
$P_{10,2} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{10,3} = 3\varepsilon^3/8$
$P_{10,4} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{10,5} = 3\varepsilon^2(1-\varepsilon)/2$
$P_{10,6} = 3\varepsilon^3/8$
$P_{10,7} = (1-\varepsilon)^3$
$P_{10,8} = 3\varepsilon(1-\varepsilon)^2/2$
$P_{10,9} = 3\varepsilon^2(1-\varepsilon)/4$
$P_{10,10} = \varepsilon^3/8$
$\Sigma_j P_{10,j} = 1.0$

References

Axelrod, R. (1984) *The Evolution of Cooperation* (New York, NY: Basic Books)

Axelrod, R. (1987) "The Evolution of Strategies in the Iterated Prisoner's Dilemma," in *Genetic Algorithms and Simulated Annealings*

Bergin, J. and B.L. Lipman (1994) "Evolution with State-Dependent Mutation," Draft, Queen's University Economics Dept.

Eaton, B. C. and M. Eswaran (1996) "Know-How Sharing as a Stochastic Prisoners' Dilemma," Department of Economics Discussion Paper, SFU.

Friedman, J (1971), "A Non-Cooperative Equilibrium for Supergames," *Review of Economic Studies* 38:1-12.

Kandori, M, G.J. Mailath, and R. Rob (1993) "Learning, Mutation and Long Run Equilibrium in Games," *Econometrica* 61(1): 29-56.

Marks (1989) "Breeding Hybrid Strategies: Optimal Behavior for Oligopolists," *ICGA Proceedings.*

Routledge, B. (1995) "Simple Spatial Evolutionary Dynamics in the Prisoner's Dilemma," UBC PhD Dissertation.

Young, H.P. (1993) "The Evolution of Conventions," *Econometrica* 61(1): 57-84.

Young, H.P and D. Foster (1991) "Cooperation in the Short and the Long Run," *Games and Economic Behavior* 3:145-156.

Figure 1a : Player Representation

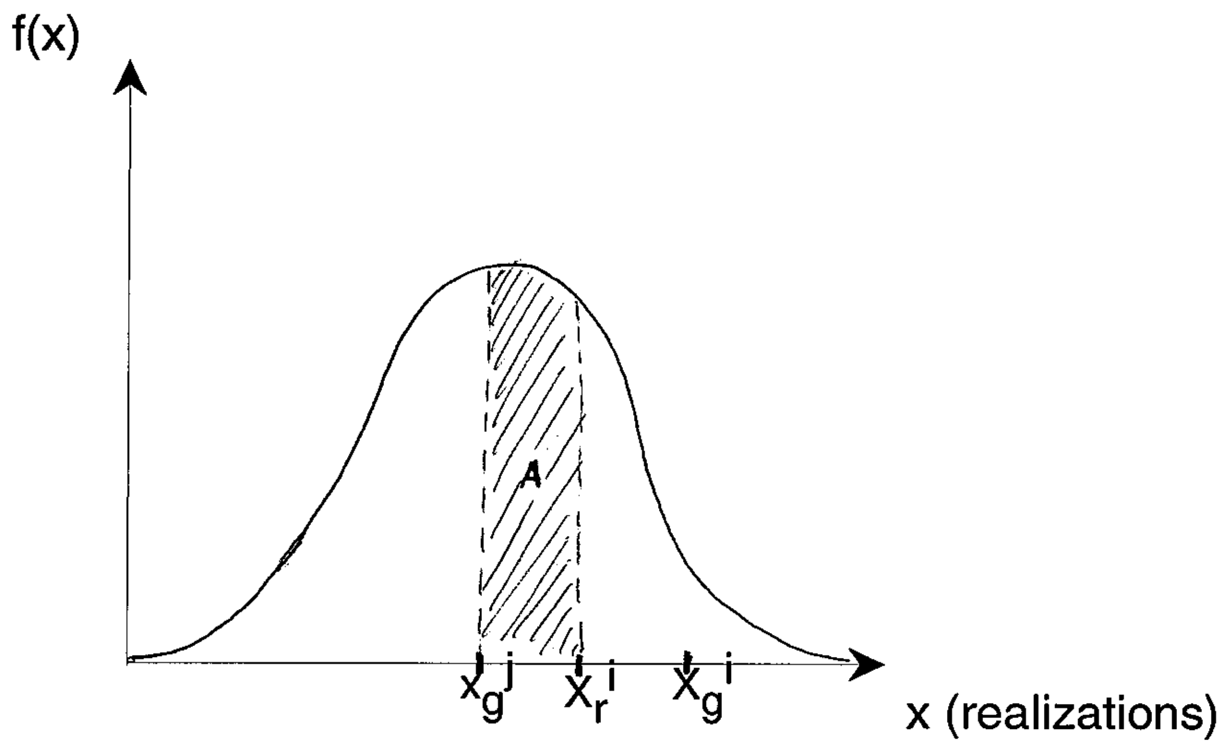


Figure 1b : A Simple Discrete Distribution

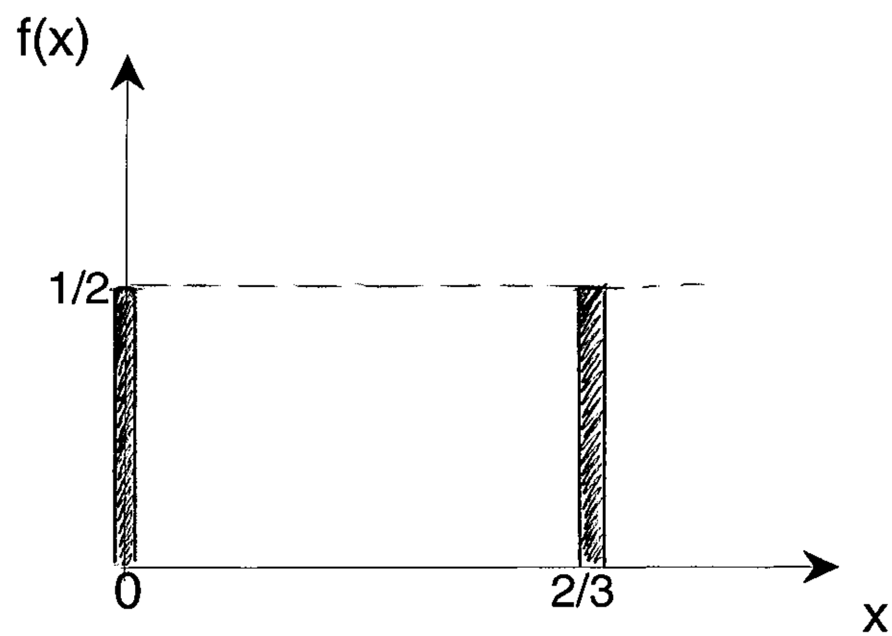


Figure 2: Population States in the SImplex

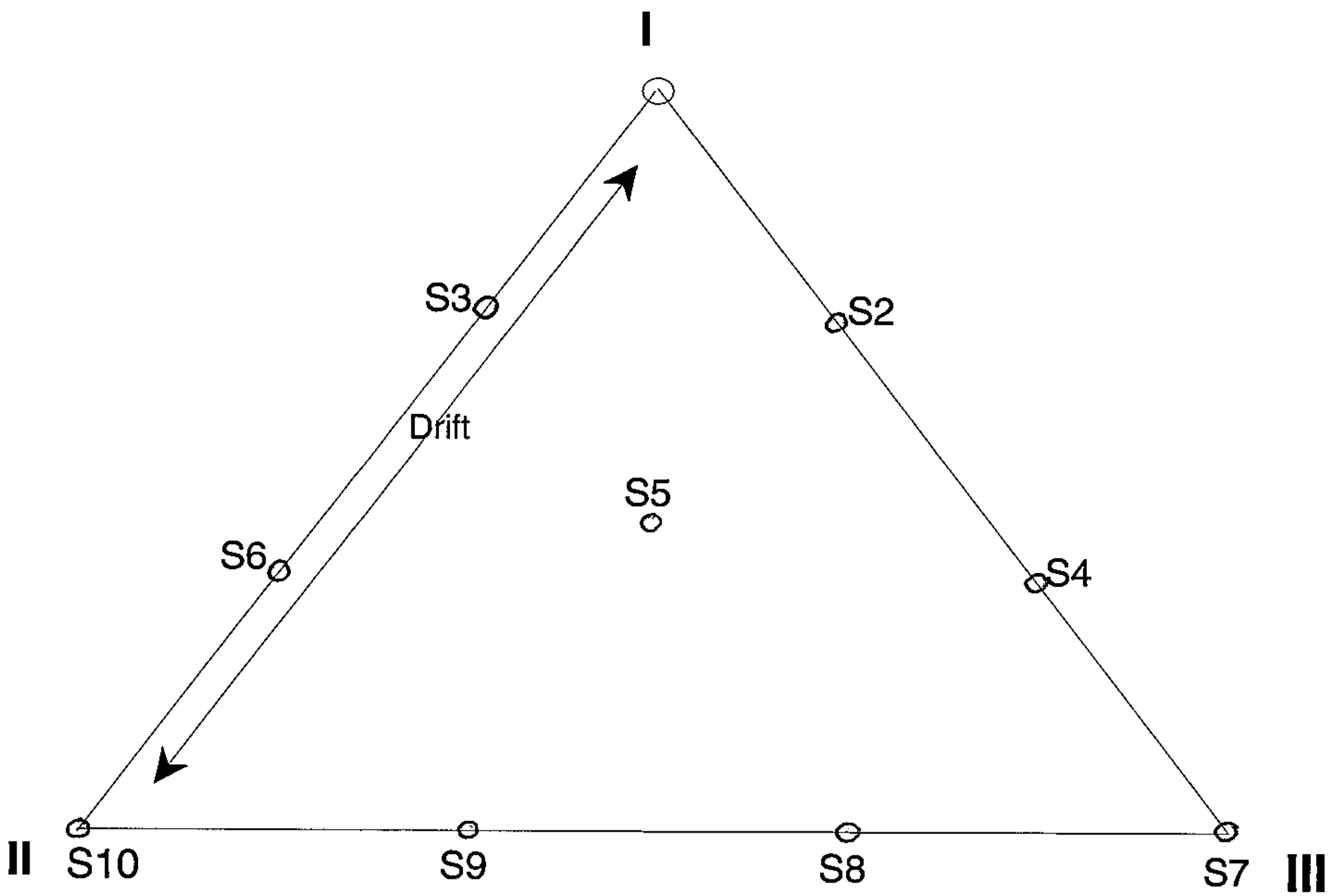
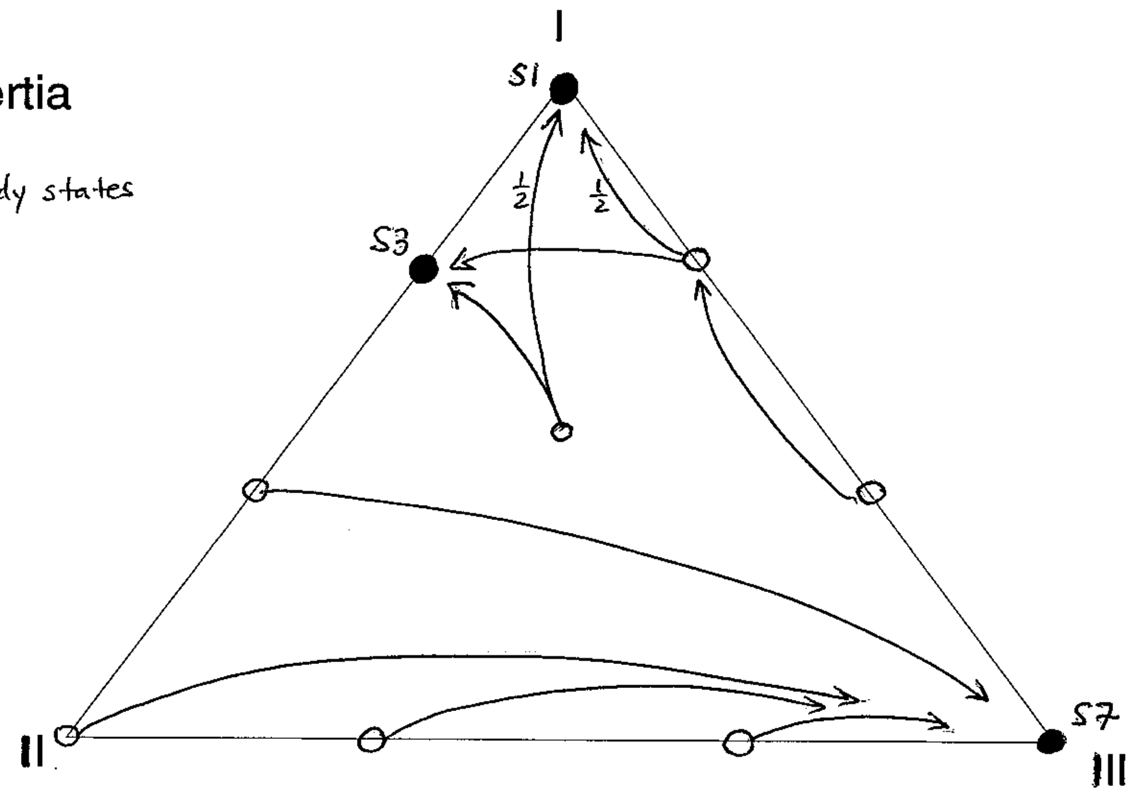


Figure 3 : Myopic Best Response

a. Inertia

● steady states



b. Experimentation

● steady state

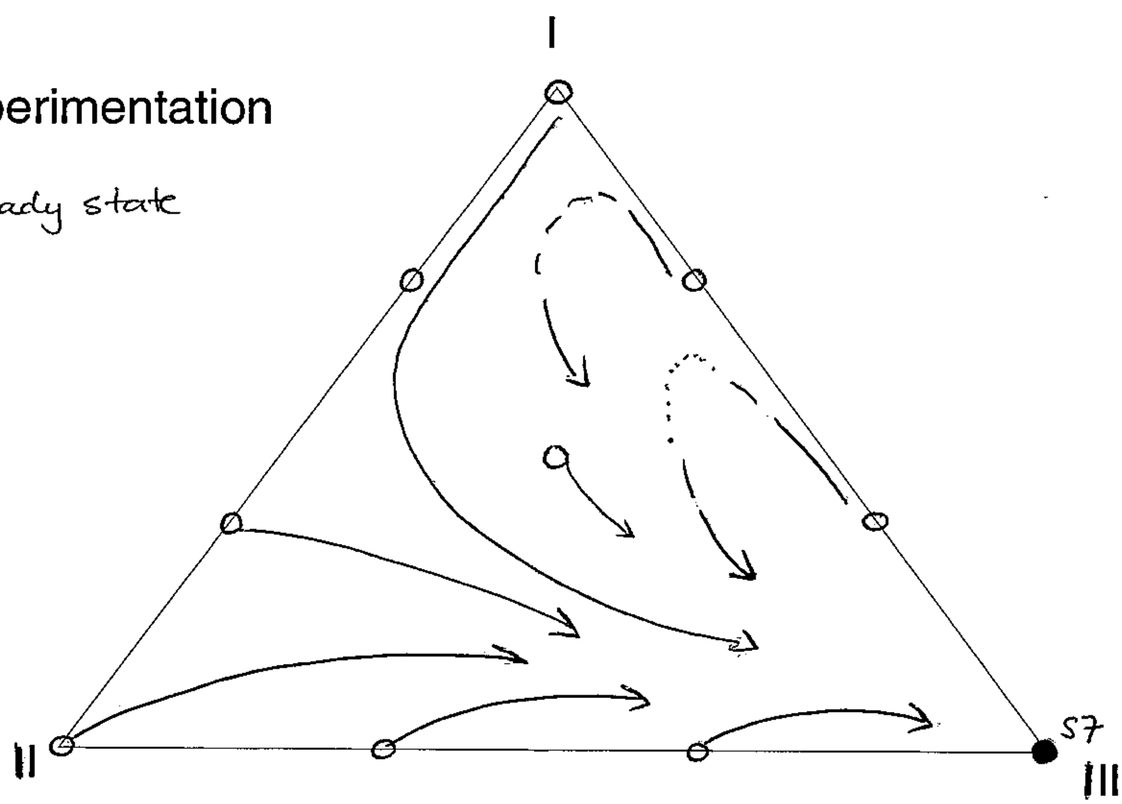
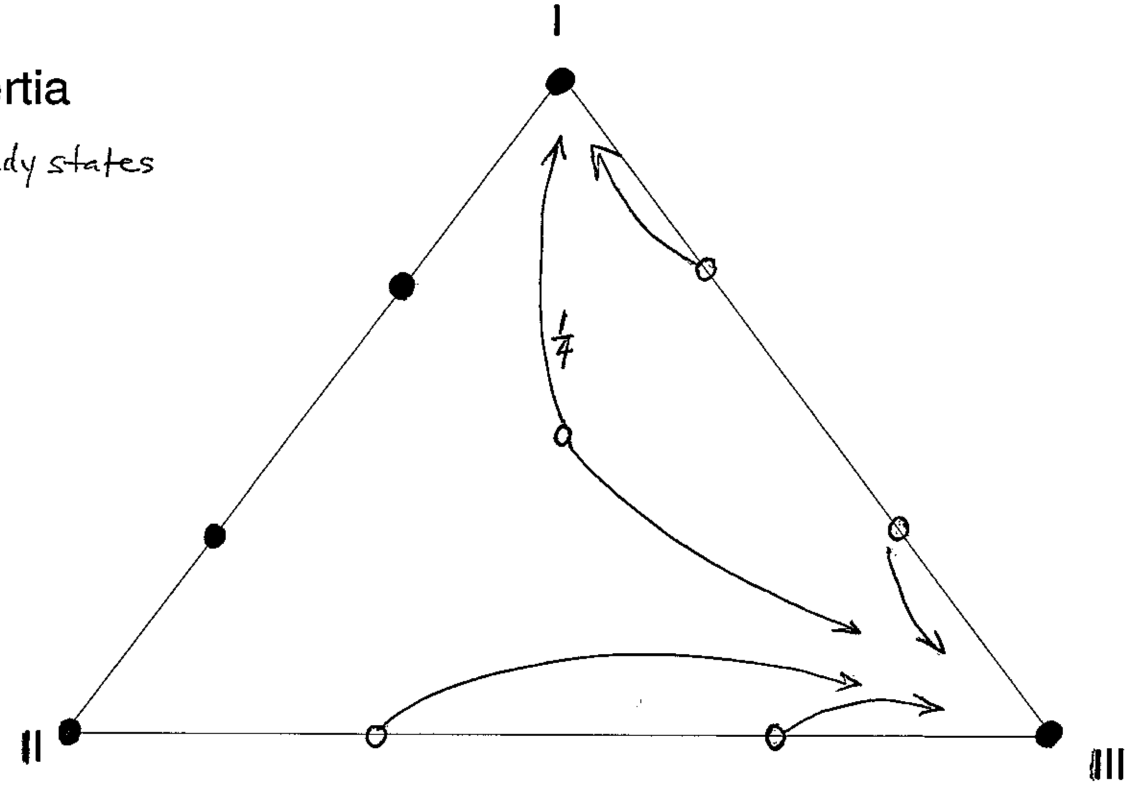


Figure 4 : Imitative Dynamics

a. Inertia

● steady states



b. Experimentation

● steady states

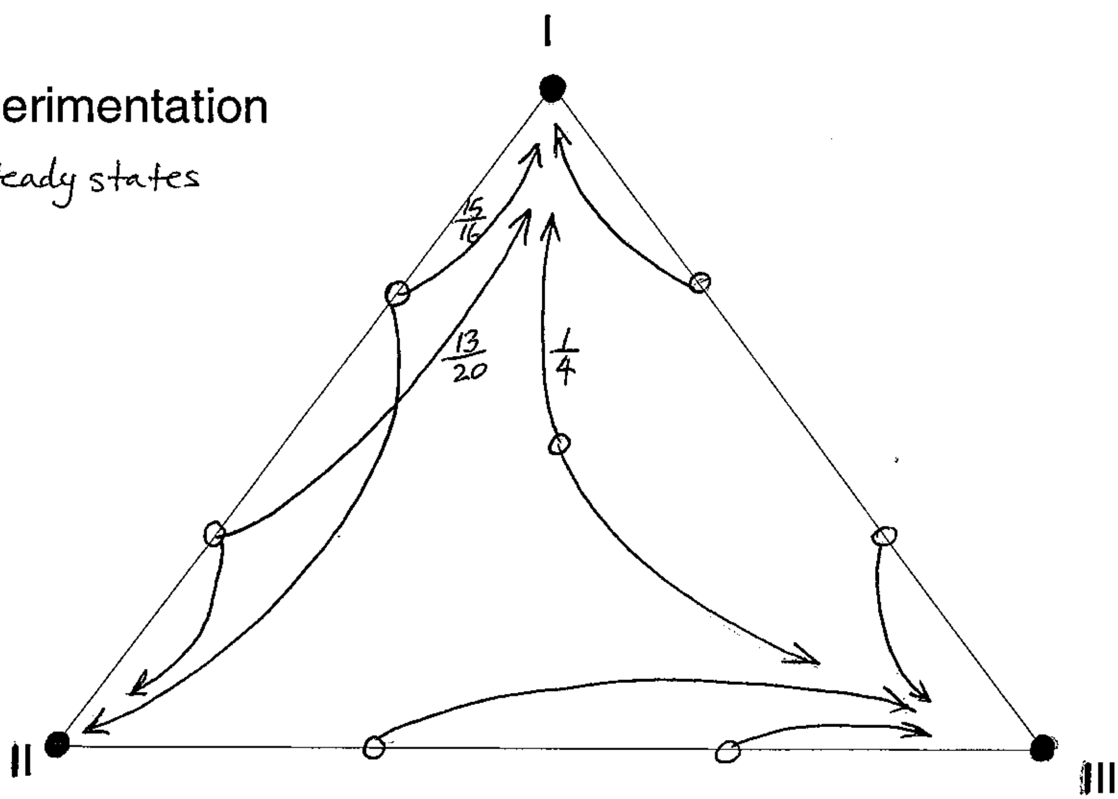


Figure 5: Table of Results.

	Deterministic Equilibria ($\varepsilon=0$)		Asymptotic Equilibria ($\varepsilon \to 0$)	
	Inertia	Experimentation	Inertia	Experimentation
Myopic Best Response	[S1, S3, S7] are steady states	[S7] is the unique steady state	[S1,S3,S7] are asymptotic equilibria	[S7] is the unique asymptotic equilibrium
	Inertia	Experimentation	Inertia	Experimentation
Imitative Behavior	[S1,S3, S6,S7, S10] are steady states	[S1, S7, S10] are steady states	[S7] is the unique asymptotic equilibrium	[S7] is the unique asymptotic equilibrium