# Contagion and State Dependent Mutations*

**In Ho Lee**
University of Southampton

**Ádam Szeidl**
Harvard University

**Ákos Valentinyi**
University of Southampton,
Institute of Economics, Hungarian Academy of Sciences
CEPR

March 12, 2001

*We thank Luca Anderlini, Glenn Ellison, Drew Fudenberg, Robin Mason, Thomas Mountford, Michele Piccione, Juuso Välimäki and seminar participants at ESSET99, Gerzensee, at the 2000 EEA Meeting in Bolzano for comments. Remaining errors are ours.

**Abstract**

Early results of evolutionary game theory showed that the risk dominant equilibrium is uniquely selected in the long run under the best-response dynamics with mutation. Bergin and Lipman (1996) qualified this result by showing that for a given population size the evolutionary process can select any strict Nash equilibrium if the probability of choosing a nonbest response is state-dependent. This paper shows that the unique selection of the risk dominant equilibrium is robust with respect to state dependent mutation in local interaction games. More precisely, for a given mutation structure there exists a minimum population size beyond which the risk dominant equilibrium is uniquely selected. Our result is driven by contagion and cohesion among players, which exist only in local interaction settings and favor the risk dominant strategy. Our result strengthens the equilibrium selection result of evolutionary game theory.

*Keywords:* Contagion, state dependent mutations, risk dominance, local interaction games.

*JEL classification:* C72; D83

# 1 Introduction

Kandori, Mailath and Rob (1993) and Young (1993) showed that the risk dominant equilibrium is uniquely selected for global interaction games with mutation under the best-response dynamics. When a group of agents play a given coordination game among themselves and there is a small mutation probability that each agent may choose a strategy that is not a best response, the risk dominant equilibrium is selected uniquely as the probability of mutation vanishes to zero. While the strength of the result follows from the parsimony of the model, Bergin and Lipman (1996) showed that the result depends crucially on the mutation parameters of the model. In particular the equilibrium selection result may be overturned in favor of the payoff dominant equilibrium if state dependent mutation is allowed and the ratio of the mutation probabilities becomes unbounded in the limit. If agents are more cautious when playing the payoff dominant action than the risk dominant action and make fewer mistakes or experiment less, then playing payoff dominant action remains the long-run equilibrium.

This paper reexamines the argument of Bergin and Lipman for equilibrium selection in the context of local interaction games i.e., in games where each player interacts only with a subset of the population called neighbors. We show that for a given mutation structure there exists a minimum finite population size such that for population size above that level, the risk dominant equilibrium is uniquely selected. This result is driven by the specific feature of local interaction games that the equilibrium selection depends on the size of the population. This is also the reason why we obtain a different result from that of Bergin and Lipman (1996). They take the population size as given, and then pick a mutation structure that selects some long run equilibrium. In contrast, we fix the mutation structure first, then we set the finite population size sufficiently so that the risk dominant equilibrium is selected. Since the equilibrium selection is independent of the population size in global interaction games [see the motivating example in Bergin and Lipman (1996)], in such an environment it does not make a difference which parameter is regarded as fixed relative to the other one. In contrast, it does matter for local interaction games. The crucial requirement is that the ratio of the mutation probabilities remains bounded as the population size changes. For example, this requirement is satisfied if the mutation structure depends only on the local configuration, which seems coherent with the approach of the local interaction games.

The nature of our result is easiest to understand in the model where agents play against the nearest neighbors on a circle of size $N$ as in Ellison's (1993) example. Assume that the probability of nonbest reply is

smaller when playing the payoff dominant action than when playing the risk dominant action. Note that if there are two adjacent players playing the risk dominant action, they continue to play it independent of the strategy profile played in the remainder of the circle. Moreover local interaction environment allows that one particular strategy can spread gradually through the population from one neighborhood to the next i.e., it allows for *contagion*. This works determininistically in one dimension because the neighbors of the two players mentioned above simply switch to the risk dominant action in the next period in the absence of mutation. Thus, transition from the configuration of playing the payoff dominant strategy everywhere to that of playing risk dominant strategy everywhere takes just 2 mutations. In contrast, transition from the configuration of playing the risk dominant strategy everywhere to that of playing the payoff dominant strategy everywhere requires $N$ simultaneous mutations because of *cohesion* among players playing the risk dominant strategy. The number of mutations required to make a transition to the state of playing the risk dominant strategy everywhere is independent of the population size while the transition to the state of playing the payoff dominant strategy everywhere depends on the population size. Therefore, if the population is big enough, the latter transition probability can be made much smaller than the former if the mutation structure across the two strategies is taken as fixed.

More generally, we demonstrate the above result in an environment where the players are located on a two dimensional torus and they interact with their nearest neighbors. We have chosen this structure for two reasons. First, this relatively simple setup shares some fundamental features of more general local interaction environments: it allows for cohesive groups formed by players located close to each other, and, in contrast to the one dimensional case, it makes contagion a stochastic phenomenon. Thus, it requires a non-trivial generalization of the one dimensional argument. Secondly, this environment is the simplest one that allows us to show how to apply some useful tools to study local interaction environments. We explore the general idea of renormalization based on cohesion and we characterize the evolution in terms of connected islands of players playing the same strategy.

The basic mechanism that characterizes local interaction environments is the presence of contagion. This mechanism is fairly trivial and deterministic in the one dimensional setup. In contrast, it is fundamentally stochastic in two dimensions. To see this point, consider a two dimensional large torus and suppose that all agents on a square of side length 2 play the risk dominant strategy. It is easy to see that these agent continue

to play the risk dominant strategy regardless of the strategy profile in the remainder of the torus without mutation. Hence the square plays a similar role to the 2 adjacent agents playing the risk dominant strategy on the circle. However the risk dominant strategy does not spread beyond the square in the absence of an additional agent playing it next to the square. We need further mutations to induce contagion, which is a stochastic event. Hence we consider the analysis as an exercise on the interaction of contagion dynamics and mutation dynamics. Without mutation (or initial randomness as in Lee and Valentinyi (2000)), there is no seed for contagion to work on; however if there is a little mutation or randomness that contagion can work on, the rest is driven by the contagion, and not mutation.

To simplify the analysis we rely on two ideas. The first one is renormalization which makes use of the following observations. If all members of a group play the risk dominant strategy and these players never abandon it without mutation, then we say that the environment allows for cohesive groups. The smallest such a group is called a team, and a team is said to play the risk dominant strategy if and only if all of its team members play it. This construction ensures that if a team plays the risk dominant strategy, then so does all of its members in the original process. However, if a player plays the risk dominant strategy in the original process, it may belong to a team that plays the payoff dominant strategy. It follows that if we want to show that the whole population eventually adopts the risk dominant strategy, then it is sufficient to show that the population of teams adopts it. The switch from players to teams is called renormalization. The advantage of the renormalized process is that teams playing the risk dominant strategy never disappear unless their members mutate. Thus, the set of teams playing the risk dominant strategy can only grow over time without mutation, but cannot shrink.

Our second idea deals with the problem of having a large and complicated state space. Indeed, for a torus of size $N$, the number of states is $2^N$. The following argument enables us to simplify the analysis of this large state space. Note that each configuration contains a (possibly empty) connected island of teams playing the risk dominant strategy. We show that the transition from one state to another depends crucially on the size of the largest such island contained in the configuration. More precisely, in a two dimensional environment we need only to know the size of the largest square of teams playing the risk dominant strategy contained in a configuration to estimate the transition probabilities to the other states. The intuitive reason is the following: small squares of teams playing the risk dominant strategy disappear with few mutations. In contrast, large squares are more difficult to break because the required number of muta-

tions is proportional to their size. Therefore, the size of the largest square is pivotal in determining the transition to another state. In particular, we show that once the process reached a state which contains a square of critical size where all teams in the square play the risk dominant strategy, that square exhibits a strong growth property. Thus, contagion makes it very likely that it grows further ensuring convergence to the risk dominant equilibrium.

Recently a number of papers have analyzed local interaction games. For instance Anderlini and Ianni (1996), Bala and Goyal (1998), Blume (1993, 1994, 1995), Lee and Valentinyi (2000) and Morris (2000) examined the issue of equilibrium selection in local interaction games with and without mutation. However, none of these papers addresses the issue for the analysis of robustness of the equilibrium selection to state dependent mutations. Our idea about how to analyze the complicated state space is related to the recent work of Ellison (2000). He develops a powerful technique for the determination of the long-run distribution of a stochastic process when there are many limit states in the system. This also can be applied in local interacting settings. He focuses on how the system evolves step-by-step from one limit set to the next. In a similar spirit, we focus on how the system evolves from one configuration containing a large connected island of players playing the risk dominant strategy to the next similar configuration. However, it is not clear from Ellison's argument why step-by-step evolution should favor the selection of the risk dominant equilibrium. Our paper demonstrates why the dynamics of the local interaction model favors the selection of risk-dominant equilibrium.

Since Bergin and Lipman (1996) suggested a careful examination of implication of the state dependent mutation, more attention has been paid to the issue. For instance van Damme and Weibull (2000) examine the decision making process associated with the cost of mistake. They conclude that the introduction of state dependent mutation does not overturn the standard equilibrium selection result since the concern for the cost of mistake favors the mutation toward the risk dominant strategy. Similarly, Young (1998) investigates whether one can relax the uniformity of mutation in a plausible way without changing the stochastically stable outcome. Although refining the decision making process is an important issue, we believe that it is at least as important to prove the robustness of the equilibrium selection result for a certain environment. In particular, since any refinement is susceptible to a perturbation of how the refinement has been made, it is important to establish the robustness of mutation for an important environment.

The rest of the paper organized as follows. Section 2 describes a 2-

dimensional environment of local interaction. Section 3 contains the main result of the paper. This section starts with a renormalization argument, then characterizes the stochastically stable outcome using a series of lemmas and propositions. Section 4 concludes the paper.

# 2 Local Interaction Game

## 2.1 A framework of local interaction

There is a population of $N^2$ players located on a 2-dimensional torus $\Lambda(N) = \mathbb{Z}^2 \cap [0, N)^2 \pmod N$ for $N \geq 1$. A player with address $x \in \Lambda(N)$ interacts with her nearest neighbors. The set of neighbors for the origin is defined by $\mathcal{N} \equiv \{y : \|y\| = 1\}$ where $\|y\| \equiv (|y_1| + |y_2|)$, and the set of neighbors for player $x$ is given by $x + \mathcal{N} \equiv \{y : \|x - y\| = 1\}$, namely the translation of $\mathcal{N}$ by $x$.

There are two actions $\{A, B\}$; and pure strategies $s_t : \Lambda(N) \longrightarrow \{A, B\}$. We characterize the population at time $t$ in terms of the set of players playing $A$, thus

$$S_t = \{x : s_t(x) = A, x \in \Lambda(N)\}. \tag{1}$$

## 2.2 Coordination game

Consider the $2 \times 2$ coordination game given in Table 1. We require that $a > c$, $d > b$ and $(a - c) > (d - b)$ so that both $(A, A)$ and $(B, B)$ are Nash equilibria and $(A, A)$ is the risk dominant one.

| | **A** | **B** |
|---|---|---|
| **A** | $a, a$ | $b, c$ |
| **B** | $c, b$ | $d, d$ |

Table 1: Coordination Game

All players play the game simultaneously over discrete periods and infinite horizon. The feature of local interaction is reflected in that the payoff of each player depends on the strategy played by herself and everyone in the neighborhood. The payoff of player $x$ playing strategy $A$ in period $t$ is given by

$$u_t(x, A) = b + (a - b)\frac{|S_t \cap (x + \mathcal{N})|}{|x + \mathcal{N}|}, \tag{2a}$$

where $|\cdot|$ denotes the cardinality of a set. Similarly, the payoff of player $x$ playing strategy $B$ in period $t$ is given by

$$u_t(x, B) = d + (c - d)\frac{|S_t \cap (x + \mathcal{N})|}{|x + \mathcal{N}|}. \tag{2b}$$

In the absence of mutation, players are assumed to play the myopic best-response: player $x$ in period $t + 1$ chooses

$$s_{t+1}(x) = \arg\max_{\{A,B\}}\{u_t(x, A), u_t(x, B)\}. \tag{3}$$

Thus player $x$ plays $A$ in period $t + 1$ if

$$\frac{|S_t \cap (x + \mathcal{N})|}{|x + \mathcal{N}|} \geq \frac{d - b}{(a - c) + (d - b)} \equiv \theta. \tag{4}$$

Since $(A, A)$ is the risk dominant equilibrium, $\theta < 1/2$. This describes the dynamics of the model without mutations.

## 2.3 State Dependent Mutation

We introduce mutation into the model: the agent may make a mistake or an experiment and thus chooses a strategy at random with a small probability. In particular we consider a stochastic process $\{S_t^\varepsilon\}_{t\geq 0}$ which is derived from $\{S_t\}_{t\geq 0}$, allowing for randomness due to mutation. In the configuration of the noise structure, we recognize explicitly the possibility of state dependent mutation probabilities.

State dependent mutations are allowed for in the following framework. Let

$$p : \Lambda(N) \times [0, 1] \times \{A, B\}^{\Lambda(N)} \times \{A, B\} \longrightarrow [0, 1]$$

be a map assigning two non-negative numbers, $p(x, \varepsilon, S, A)$ and $p(x, \varepsilon, S, B)$, to player $x$ that represent the probability of mutation to $A$ and $B$, respectively. We require that $p(x, \varepsilon, S, A) + p(x, \varepsilon, S, B) \in [0, 1]$. The probability depends on the location of the player, on the overall state of the economy, on the state of the player and on a scaling parameter $\varepsilon$.

Define $\{\lambda_t(x)\}_{x\in\Lambda(N)}$ for all $t$ as a collection of independent random variables uniformly distributed on $[0, 1]$, and define $X_t$ as:

$$X_{t+1}(x, S_t, \varepsilon) = \begin{cases} A & \text{if } \lambda_t(x) \leq p(x, \varepsilon, S_t^\varepsilon, A) \\ B & \text{if } \lambda_t(x) \geq 1 - p(x, \varepsilon, S_t^\varepsilon, B) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Thus, agent $x$ in state $S_t$ mutates to $A$ in period $t+1$ if $X_t(x, S_t, \varepsilon) = A$; mutates to $B$ if $X_t(x, S_t, \varepsilon) = B$; and does not mutate otherwise. Also note that $\{X_t(x, S_t, \varepsilon)\}_{x\in\Lambda(N)}$ is a collection of random variables for each $t$ that are independent across players and time. The specification ensures that the mutation probability depends on the state of the population, the location of the agent, and so on. In particular, the state of the population includes the state of agent $x$ so that the mutation probability may be different depending on whether the agent plays $A$ or $B$ and one strategy may be chosen with a higher mutation probability than the other one. To simplify notation we drop the arguments $S_t, \varepsilon$ from $X_t(x, S_t, \varepsilon)$ and use $X_t(x)$ if it does not lead to confusion.

We make the following assumption about the mutation structure.

**Assumption 1** *The mutation probability satisfies:*

1. $\lim_{\varepsilon\to 0} p(x, \varepsilon, S, \cdot) = 0$ *for all* $x \in \Lambda(N)$ *and* $S \subset \Lambda(N)$*, and*

2. *there exist* $\alpha(A) > 0$ *and* $\alpha(B) > 0$ *such that*

$$\inf_N \min_{x\in\Lambda(N), S\subset\Lambda(N)} p(x, \varepsilon, S, A) = \varepsilon^{\alpha(A)} \tag{6a}$$

$$\sup_N \max_{x\in\Lambda(N), S\subset\Lambda(N)} p(x, \varepsilon, S, B) = \varepsilon^{\alpha(B)}. \tag{6b}$$

The first assumption is standard. The second one requires that the mutation probabilities are uniformly bounded across agents, states and populations sizes i.e., $\varepsilon^{\alpha(A)}$ and $\varepsilon^{\alpha(B)}$ are the lower bound and upper bound on the probability that an agent chooses strategy $A$ and $B$ as a consequence of mutation, respectively. This assumption is not more restrictive than other similar assumptions used in the literature (see Ellison (2000) and Young (1993) for example). Note that $\varepsilon^{\alpha(B)}$ can be of a lower order than $\varepsilon^{\alpha(A)}$, so that the ratio of the mutation probabilities may well be unbounded as $\varepsilon \downarrow 0$, and also $B$ can be adopted increasingly more frequently than $A$ as $\varepsilon \downarrow 0$. Hence this is compatible with the environment for which Bergin-Lipman's result holds for a given population size.

Using the construction, the transition rule for the process $\{S_t^\varepsilon\}_{t\geq 0}$ can be formally written as

$$\begin{aligned} S_{t+1}^\varepsilon &= \Phi(S_t^\varepsilon) \\ &\equiv \left\{x \in \Lambda(N) : X_{t+1}(x) = 0, \frac{|S_t^\varepsilon \cap (x+\mathcal{N})|}{|x+\mathcal{N}|} \geq \theta\right\} \cup \left\{x : X_{t+1}(x) = A\right\}. \end{aligned} \tag{7}$$

This construction defines a Markov-chain $S_t^\varepsilon$.

Denoting the event where everybody plays $A$ by $\vec{A}$, we are interested in the long-run probability of the event $\vec{A}$:

$$\mu_\varepsilon(\vec{A}) = \lim_{t\to\infty} \Pr(S_t^\varepsilon = \Lambda(N)). \tag{8}$$

which is the same as the probability of $\vec{A}$ under the invariant distribution of $S_t^\varepsilon$. In particular our main goal is to characterize $\mu_\varepsilon(\vec{A})$ as $\varepsilon \downarrow 0$.

# 3 Best Response Dynamics with State Dependent Mutations

Our analysis of the long-run distribution of the population consists of two steps. First, we construct a new process which satisfies a sample path inequality with respect to the original process; the sample path inequality then implies a certain distributional inequality.[1] Secondly, we characterize the long-run distribution for the new process and deduce the properties of the limiting distribution of the original process using the distributional inequality. The main benefit of the stepwise approach is the analytical convenience the new process provides.

## 3.1 Renormalization and Coupling Argument

Suppose that $N$ is even.[2] Let $\Lambda(M)$ be a torus where $M \equiv N/2$, and assign to each $y \in \Lambda(M)$ a set of four players from the original torus $\Lambda(N)$ by

$$H(y) = \Big\{x \in \Lambda(N) : x_i \in \{2y_i, 2y_i + 1\}, i = 1, 2, y \in \Lambda(M)\Big\}. \tag{9}$$

We call $y$ and $x \in H(y)$ a team and a team member, respectively. We shall also refer to the population of teams as the renormalized population. Let $z_t : \Lambda(M) \longrightarrow \{A, B\}$ be a map which represents the state of team $y$ at time $t$, which will be defined shortly. As before, we characterize the dynamics of the population in terms of the set of teams playing $A$, thus $Z_t^\varepsilon = \{y : z_t(y) = A\}$. We construct an initial configuration $Z_0^\varepsilon$, and a transition rule such that that if the state of a team is $A$, then all of its team members play $A$ in the original population.[3]

First, we define the *initial state* for the renormalized process. At time zero, a team is said to play $A$ if all of its members in the original population

[1] This is called the coupling technique. See Aldous and Fill (1999, Chapter 14).

[2] We discuss later what happens if $N$ is odd.

[3] On the other hand we do not require if a team plays $B$, then all members play $B$.

play $A$; otherwise it plays $B$. Formally, $Z_0^\varepsilon$ is defined by

$$Z_0^\varepsilon = \{y : z_0(y) = A\}. \tag{10}$$

where

$$z_0(y) = \begin{cases} A & \text{if} \quad s_0(x) = A \quad \forall x \in H(y) \\ B & \text{otherwise.} \end{cases} \tag{11}$$

Note that

$$\{x \in H(y) : y \in Z_0\} \subset S_0. \tag{12}$$

Thus, the set of agents belonging to teams playing $A$ in the initial period is a subset of the set of agents playing $A$ in the original population.

Secondly, we construct the transition rule for the renormalized process *without mutation* such that the condition $\{x \in H(y) : y \in Z_t\} \subset S_t$ (if a team plays $A$, then all of its members in the original population play $A$) is satisfied for all $t \geq 0$. Observe that if all members of a team play $A$ in the original population, then the best response for all team members is $A$. Therefore, if a team plays $A$, it will never adopt $B$ in the absence of mutation. Moreover, let

$$G_i(y) = \{y - e_i, y + e_i\}, \quad \text{for } i = 1, 2 \qquad \text{where} \quad e_1 = (1, 0) \quad e_2 = (0, 1).$$

That is, $G_i(y)$ represents the two neighbors of team $y$ in coordinate direction $i$. It is easy to see that if a team playing $B$ encounters two other teams all of whose members are playing $A$ in two different coordinate directions, then all members of the team adopt $A$ in at most 3 periods under the original process (see Figure 1 where the reference to time indicates the period when the agent adopts $A$).

For future reference we summarize the transition rule for the renormalized process in the absence of mutation by the following lemma.

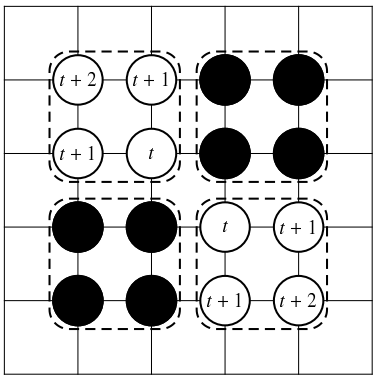
**Lemma 1 (Contagion)** *In the absence of mutations*

1. *if a team y plays A, then never adopts B,*
2. *if a team y has at least one member playing B, then all of its members adopt A if $G_i(y) \cap Z_t^\varepsilon \neq \varnothing$ for both $i = 1$ and $i = 2$.*

Thirdly, we construct a random variable representing *mutations* for the renormalized population using the random variable representing mutations for the original population. We do so in two steps. Let

$$\xi_t(x) = \begin{cases} A & \text{if } \lambda_t(x) \leq \varepsilon^{\alpha(A)} \\ B & \text{if } \lambda_t(x) \geq 1 - \varepsilon^{\alpha(B)} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Figure 1: Propagation mechanism for the teams with no mutation




● agents playing $A$ in period $t$

The construction ensures that the event $\{\xi_t(x) = A\}$ implies $\{X_t(x) = A\}$, and $\{X_t(x) = B\}$ implies $\{\xi_t(x) = B\}$. Next, let $Y_t(y)$ be a random variable with values

$$Y_t(y) = \begin{cases} A & \text{if } \xi_{3t}(x) = A \ \forall\, x \in H(y) \\ 0 & \text{if } \xi_n(x) \neq B \ \forall\, x \in H(y) \text{ and } n = 3t-2, 3t-1, 3t, \\ & \text{and } \xi_{3t}(x) \neq A \\ B & \text{otherwise.} \end{cases} \tag{14}$$

In words $\{Y_t(y) = A\}$ is the event in which all team members mutate to A in period $3t$, $\{Y_t(y) = 0\}$ is the event that no mutation to $B$ occurred during the periods $3t-2, 3t-1, 3t$, and we are not in the previous case. Finally, $\{Y_t(y) = B\}$ is the event that at least one team mutates to $B$ during the periods $3t-2, 3t-1, 3t$, and all team members do not mutate to $A$ in period $3t$.

There are two important facts about $Y_t(x)$. First, the construction of $Y_t(x)$ and Lemma 1 ensure that the stochastic process for the renormalized population is a Markov process. Secondly, the timing for the renormalized population is different from that of the original population. As we can see on Figure 1, it takes three periods under the original process for all team members to adopt $A$. Therefore the clock for the renormalized population ticks slower: three periods under the original process correspond to one period under the renormalized one. This is reflected by the timing convention used in the definition of the random variable $Y_t(x)$.

We can use Assumption 1 to provide bounds on the events that a team mutates to $A$ or to $B$. Setting $\bar{\alpha}(A) = 4\alpha(A)$, we obtain

$$\Pr(Y_t(y) = A) = \varepsilon^{\bar{\alpha}(A)}. \tag{15a}$$

Moreover, a team mutates to $B$ with probability at most $12\varepsilon^{\alpha(B)}$. For small $\varepsilon$ there is a $\bar{\alpha}(B)$ such that this probability is at most $\varepsilon^{\bar{\alpha}(B)}$ i.e.,

$$\Pr(Y_t(y) = B) \leq \varepsilon^{\bar{\alpha}(B)} \tag{15b}$$

Finally, after the construction of the transition rule without mutations, and defining the random variable representing mutations for the teams, we can construct the *transition rule* for the renormalized process with mutation

$$\begin{aligned} Z^{\varepsilon}_{t+1} &= \Psi(Z^{\varepsilon}_t) \\ &\equiv \left\{ y \notin Z^{\varepsilon}_t : Y_{t+1}(y) = 0, G_i(y) \cap Z^{\varepsilon}_t \neq \varnothing, \exists u Y_{t+1}(u) = 0, u \in G_i(y) \ \forall i \right\} \\ &\quad \cup \left\{ y \in Z^{\varepsilon}_t \, : \, Y_{t+1}(y) = 0 \right\} \cup \left\{ y \in \Lambda(M) \, : \, Y_{t+1}(y) = A \right\}. \end{aligned} \tag{16}$$

Equation (16) can be decomposed to three transition mechanisms: contagion, cohesion, and mutation. *Contagion* is represented by the first bracket. If team $y$ is playing $B$ at time $t$, and meets two other teams playing $A$ in two different coordinate directions, it chooses $A$ at time $t + 1$. Since we require that neither team $y$ nor any of the two other teams mutate at time $t + 1$, the construction in (14) implies that no members of any of these three teams mutated in the original process at $3t + 1, 3t + 2, 3t + 3$. Therefore, all members of team $y$ make a transition from $B$ to $A$ during $3t + 1, 3t + 2, 3t + 3$. Consequently, all players in these three teams plays $A$ at time $3t + 3$ in the original process. *Cohesion* is captured by the second bracket. It indicates that a team playing $A$ continues with the same choice in the absence of mutation. Since $Y_{t+1}(y) = 0$ implies that no member of team $y$ mutates during $3t + 1, 3t + 2, 3t + 3$, all members of team $y$ play $A$ at time $3t + 3$ in the original process. *Mutation* is represented by the last bracket. Again, the construction of mutation for teams ensures that if team $y$ mutates to $A$ at time $t + 1$, so do all team members at time $3t + 3$ in the original process.

All three mechanisms are inherited from the original population; renormalization makes cohesion explicit and simplifies the way contagion works. The above decomposition also highlights the analytical advantage of working with the renormalized process rather then the original one. Transition from $B$ to $A$ can take place both in the presence and the absence of mutation in the original process. In contrast, transition from $B$ to $A$ takes place only as a consequence of mutation in the renormalized process. This simplifies the derivation of the limit distribution of the renormalized process.

It follows from our discussion above that our construction ensures the desired relationship between the original and renormalized process as stated in the next proposition.

**Proposition 1** *The process* $\{Z_t^\varepsilon\}_{t\geq 0}$ *governed by the transition rule* (16) *and starting from the initial condition* (10) *satisfies*

$$\{x \in H(y) : y \in Z_t^\varepsilon\} \subset S_{3t}^\varepsilon. \tag{17}$$

*for all t.*

Proposition 1 in turn has an important consequence for our analysis captured by the following statement. Let $\nu_\varepsilon(\vec{A})$ be the long run probability of the event where all teams play $A$.

**Corollary 1** *If* $\lim_{\varepsilon\to 0} \nu_\varepsilon(\vec{A}) = 1$, *then* $\lim_{\varepsilon\to 0} \mu_\varepsilon(\vec{A}) = 1$.

**Proof.** It suffices to prove that $\mu_\varepsilon$ dominates $\nu_\varepsilon$, i.e. $\nu_\varepsilon(\vec{A}) \leq \mu_\varepsilon(\vec{A})$. This follows from equation (17) i.e., the event that $Z_\infty^\varepsilon = \Lambda(M)$ implies the event that $S_\infty^\varepsilon = \Lambda(N)$. □

## 3.2 Contagion versus State Dependent Mutation

This subsection presents the main result of the paper. Our main goal is to demonstrate that for any state dependent mutation structure satisfying the conditions in Assumption 1, contagion dominates mutation: given the mutation structure, the risk dominant equilibrium will be uniquely selected in the long run for a suffuciently large but finite population.

The proof is built around two important concepts.

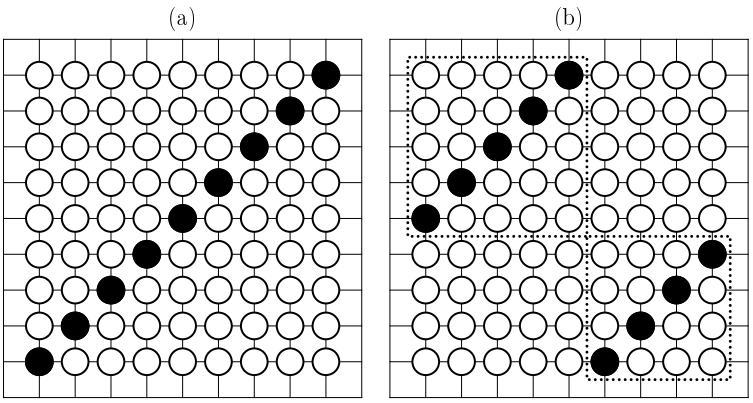
**Definition 1 (Lock-in)** *A state Z is called a lock-in if no team changes strategy in the absence of mutation. The set of these states is denoted by* $\mathcal{L}$.

If strategy $A$ does not spread further through contagion, then the state is a lock-in.[4] The next concept is new.

**Definition 2 (k-restricted lock-in)** *Draw an* $M \times M$ *square representation of the torus. Denote the "left-upper"* $k \times k$ *square by* $R(k)$. *For state Z, define the restriction of Z to* $R(k)$ *to be a state where all teams inside* $R(k)$ *play the strategy according to Z while all teams outside* $R(k)$ *play B. If the restriction of Z to* $R(k)$ *is a lock-in, then Z is called a k-restricted lock-in. The set of k-restricted lock-ins is denoted by* $\mathcal{L}_k$.

[4]This state is also called a limit set of the mutationless process in the literature.

Figure 2: The sufficient number of teams ensuring coordination on $A$ without mutation




● agents playing $A$ in period $t$

Put differently, we treat all teams outside $R(k)$ as if they played $B$. If no team ever adopts $A$ inside $R(k)$ due to contagion, then $Z \in \mathcal{L}_k$. The definition describes the set of states where the configuration outside $R(k)$ does not generate contagion[5] i.e., there are no contagious effects coming from outside $R(k)$. It is easy to see that a $k$-restricted lock-in is also an $l$-restricted lock-in for all $l < k$. Although a restricted lock-in is not necessarily a lock-in, any lock-in Z is a $k$-restricted lock-in for all $k < M$.

We start by proving a preparatory lemma. Let $\mathcal{L}_k^A$ be the set of $k$-restricted lock-ins with all teams in $R(k)$ playing $A$. Moreover, let $\mathcal{Z}_k$ be the set of configurations where at least $k$ teams in $R(k)$ play $B$.

**Lemma 2** *If $Z \in \mathcal{L} - \vec{A}$ i.e, $\{Z \in \mathcal{L} : Z \neq \vec{A}\}$, then $Z \in \mathcal{Z}_M$ i.e., it contains at least $M$ teams playing $B$. If $Z \in \mathcal{L}_k - \mathcal{L}_k^A$, then $Z \in \mathcal{Z}_k$ i.e., it contains at least $k$ teams playing $B$.*

**Proof.** Consider the statement for lock-ins. Observe that it is sufficient for the whole population to adopt $A$ in the absence of mutation if at least $M$ teams play $A$ at $t = 0$ and these teams are located on a diagonal [see the illustration in Figure 2a]. It is easy to see that all teams next to the diagonal have two neighbors in two different coordinate directions playing $A$. Therefore these teams adopt $A$ by lemma 1. Applying induction, we obtain that all teams adopt $A$ at $t = M - 1$.

Next, suppose that at most $M - 1$ teams play $B$ initially. To see that such a configuration cannot be a lock-in, note that a torus of size $M \times M$

[5]To ensure that the teams at the boundary of the $M \times M$ square have a regular neighborhood structure, we assume that there is an $(M + 1)$st row and column of teams all playing $B$ for all $t \geq 0$.

can be viewed as an object consisting of $M$ parallel diagonals of length $M$. If at most $M-1$ teams play $B$ at $t=0$, there must be at least one diagonal on the torus on which all teams play $A$. Hence our previous argument implies that the the whole population adopts $A$ which proves the first part of our lemma.

Now consider the statement for $k$-restricted lock-ins. Since a $k$-restricted lock-in is not a torus but a square which cannot be viewed as an object consisting of parallel diagonals, the previous argument cannot be applied directly. However, a redefinition of the concept "diagonal" for squares solves our problem. View $R(k)$ as consisting of an upper and a lower triangle plus the main diagonal. Pick up a diagonal of length $n_1$ from the upper triangle and the corresponding diagonal of length $n_2 = k - n_1$ from the lower triangle. Let a "diagonal" be defined by these two pieces. The $k-1$ such "diagonals" plus the main diagonal give us $k$ "diagonals". Elementary argument shows that if any of the constructed diagonals contains $k$ teams playing $A$, all teams in $R(k)$ eventually adopt $A$ in the absence of mutation [see the illustration in Figure 2b]. Therefore our previous argument applies proving the second part of the lemma. □

These two results are a very useful when we need to know the minimum number of mutations required to leave the basin of attraction of $\vec{A}$ or a $k$-restricted lock-in with all teams in $R(k)$ playing $A$.

After this preparation, we shall characterize the limit distribution of the process $\{Z_t^\varepsilon\}$ using a result on first passage times. First, we introduce some notation. Let $Z$ and $Z'$ be two states and $T_Z = \inf\{t > 0 : Z_t^\varepsilon = Z\}$ be the first passage time to state $Z$. Let

$$\Pr{}_Z(Z_t^\varepsilon = Z') \equiv \Pr(Z_t^\varepsilon = Z' | Z_0^\varepsilon = Z).$$

and

$$\Pr{}_{\mathcal{Z}}(E) = \inf_{Z \in \mathcal{Z}} \Pr{}_Z(E)$$

for an event $E$ and $\mathcal{Z} \subset \Lambda(M)$. We use the following well-known identity for ergodic Markov-processes to characterize the long run distribution [see Durrett (1996, Chapter 5)]:

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\vec{A})} = \frac{\Pr_{\vec{A}}\left(T_Z < T_{\vec{A}}\right)}{\Pr_Z\left(T_{\vec{A}} < T_Z\right)} \tag{18}$$

To show that for a given mutation structure, $\vec{A}$ is played most of the time in a sufficiently large but finite population as $\varepsilon \downarrow 0$, we look for an appropriate upper bound in terms of $\varepsilon$ on the above ratio. We estimate the numerator and denominator separately. The key is to estimate the latter.

Now we turn to the proof for the main theorem which is constructed via three lemmas. Although they are somewhat technical, we believe that there is a clear intuition behind each of these lemmas. The first lemma establishes the order of the probability of getting from one lock-in to another one which has "fewer teams" playing $A$.

**Lemma 3** *Let $Z$ and $Z'$ be two lock-ins with $Z \not\subseteq Z'$. The probability of getting from lock-in $Z$ to lock-in $Z'$ without hitting any other lock-in is given by*

$$\Pr_Z(T_{Z'} < T_{\mathcal{L}-Z'}) = O(\varepsilon^{\bar{\alpha}(A)|Z\setminus Z'|}). \tag{19}$$

**Proof.** We estimate the following sum

$$\Pr_Z(T_{Z'} < T_{\mathcal{L}-Z'}) = \sum_{t=1}^{\infty} \Pr_Z(T_{Z'} = t, T_{\mathcal{L}-Z'} > t).$$

First, observe that for any $t$, the event $\{T_{Z'} = t, T_{\mathcal{L}-Z'} > t\}$ requires that the process leaves $Z$ in the first period since $Z \in \mathcal{L} - Z'$. Next, if the process leaves $Z$, at least one team adopts $A$ in every period due to contagion until the next lock-in is reached. Since the process may hit $\vec{A}$ at time $t = M^2$, we need at least $\max\{t - M^2, 0\}$ mutations to $B$ not to hit a lock-in by time $t$. Moreover, since $Z \not\subseteq Z'$ and transition from $A$ to $B$ is not possible without mutation, we need at least $|Z\setminus Z'|$ more mutations to $B$ to reach $Z'$. Thus, reaching $Z'$ by time $t$ without hitting any other lock-in along a given trajectory has probability at most $\varepsilon^{\bar{\alpha}(B)(|Z|-|Z'|+\max\{t-M^2,0\})}$. We have to multiply this value by the number of possible trajectories. Since each trajectory is $t$ long, by choosing a sufficiently large $K$ depending only on $M$, we can estimate this number by $K^t$.[6] For $\varepsilon$ sufficiently small, the sum in the inequality

$$\Pr_Z(T_{Z'} < T_{\mathcal{L}-Z'}) \leq \varepsilon^{\bar{\alpha}(B)(|Z\setminus Z'|)} \sum_{t=1}^{\infty} K^t \varepsilon^{\bar{\alpha}(B)\max\{t-M^2,0\}}.$$

is convergent. Indeed the sum is $O(1)$ which proves our claim. □

Note that any configuration with fewer then $M$ teams playing $B$ (other then $\vec{A}$) is not a lock-in by Lemma 2. Therefore this lemma also implies that if the process starts from $\vec{A}$, the probability that it hits any configuration with at least $M$ teams playing $B$ before getting back to $\vec{A}$ is bounded from above by $K\varepsilon^{\bar{\alpha}(A)M}$.

[6] Different $K$'s appearing in different formulas may be different.

The next lemma formulates a "local" version of this observation. In words, the lemma says that the probability of leaving a configuration which contains a $k \times k$ square of teams playing $A$ and returning to a similar configuration before hitting a state where the $k \times k$ square has at least a row or column of players playing $B$ is bounded by some constant multiple of $\varepsilon^{\bar{\alpha}(B)k}$. This local version of the previous lemma will make it possible to track the evolution of the process more closely and in the end prove our main claim.

**Lemma 4** *For some $K > 0$ independent of $\varepsilon$*

$$\Pr_{\mathcal{L}_k^A}\left(T_{\mathcal{Z}_k} < T_{\mathcal{L}_k^A}\right) \leq K\varepsilon^{\bar{\alpha}(B)k}. \tag{20}$$

**Proof.** The proof is based on the construction of a shadow process $Z_t^{k,\varepsilon}$, which is intended to keep track of only those events taking place inside of $R(k)$. We prove the claim for the shadow process making use of the previous lemma and then we transform the result to $Z_t^{\varepsilon}$ by using a coupling argument.

First we construct the shadow process as follows. At the initial state, teams inside $R(k)$ in $Z_0^{k,\varepsilon}$ play the same strategy they play in $Z_0^{\varepsilon}$. However, teams outside $R(k)$ in $Z_0^{k,\varepsilon}$ play $B$. This is to ensure that we focus only on the "local problem" i.e., what happens inside $R(k)$. We define the transition rule as follows. Teams inside $R(k)$ in $Z_t^{k,\varepsilon}$ mutate to a given strategy if they do so inside $R(k)$ in $Z_t^{\varepsilon}$. Contagion within $R(k)$ in $Z_t^{k,\varepsilon}$ takes place according to the standard rule for teams. Finally, teams outside $R(k)$ in $Z_t^{k,\varepsilon}$ play $B$ forever.

It is easy to see that $Z_t^{k,\varepsilon} \subset Z_t^{\varepsilon}$ for all $t \geq 0$. It is true for $t = 0$ by construction. Since no team ever plays $A$ outside $R(k)$ in $Z_t^{k,\varepsilon}$, we have to show only that our claim is true inside $R(k)$. First, mutation generates teams playing $A$ under $Z_t^{k,\varepsilon}$ whenever it does so under $Z_t^{\varepsilon}$. Secondly, since contagion is induced only by $A$ and teams outside $R(k)$ play $B$ forever, teams outside $R(k)$ have no contagious effects on teams inside $R(k)$ in $Z_t^{k,\varepsilon}$. Since this is not true for $Z_t^{\varepsilon}$, we have $Z_t^{k,\varepsilon} \subset Z_t^{\varepsilon}$.

Consider now the shadow process $Z_t^{k,\varepsilon}$. Lemmas 2 and 3 imply for the shadow process that the probability of getting to $\mathcal{Z}_k$ before returning to $\mathcal{L}_k^A$ is bounded from above by $K\varepsilon^{\bar{\alpha}(B)k}$.[7] The condition $Z_t^{k,\varepsilon} \subset Z_t^{\varepsilon}$ implies that if $Z_t^{\varepsilon}$ hits $\mathcal{Z}_k$, then so does $Z_t^{k,\varepsilon}$, and conversely if $Z_t^{k,\varepsilon}$ returns to $\mathcal{L}_k^A$, then so does $Z_t^{\varepsilon}$. It follows that the event that $Z_t^{\varepsilon}$ hits $\mathcal{Z}_k$ before returning to $\mathcal{L}_k^A$ implies the event that $Z_t^{k,\varepsilon}$ hits $\mathcal{Z}_k$ before returning to $\mathcal{L}_k^A$. Thus if

[7] Note that from the point of view of the shadow process the set $\mathcal{L}_k^A$ could be considered to be a singleton since teams outside $R(k)$ do not count anyway.

the required inequality holds for $Z_t^{k,\varepsilon}$, then it also holds for $Z_t^{\varepsilon}$. The lemma is proved. □

The next lemma builds on the previous two lemmas to provide the key to the proof of our main claim.

**Lemma 5** *Let*

$$m \equiv \left\lceil \frac{\bar{\alpha}(A)}{\bar{\alpha}(B)} \right\rceil, \tag{21}$$

*where* $\lceil . \rceil$ *is the ceiling function (the next integer up if* $\bar{\alpha}(A)/\bar{\alpha}(B)$ *is not an integer). For all* $M > m$*,*

$$\frac{1}{\Pr_Z\left(T_{\vec{A}} < T_Z\right)} = O(\varepsilon^{-\bar{\alpha}(A)m}) \tag{22}$$

*for any* $Z \in \mathcal{L} - \vec{A}$*.*

**Proof.** We construct an event which is a subset of the event $\{T_{\vec{A}} < T_Z\}$, and estimate its probability from below. The key step of the proof is to show that for all $k \geq m$ play $A$, there is a positive probability independent of $\varepsilon$ that all teams in $R(k+1)$ will play $A$ before returning to $Z$. Thus $m$ plays the role of a critical mass: once a sufficiently large set of teams (measured by $m$) play $A$, it is easy to proceed. So the difficulty lies in reaching this critical mass of $m$, and hence the bound $\varepsilon^{\bar{\alpha}(A)m}$.

Since $Z \in \mathcal{L} - \vec{A}$, it follows from lemma 2 that state $Z$ contains at least $M$ teams playing $B$. Draw an $M \times M$ square representation of the torus such that the uppermost and leftmost team plays $B$ in $Z$. Denote the "left upper" $k \times k$ square in this $M \times M$ torus by $R(k)$, and the set of diagonal elements by $D(k)$. Observe that this construction ensures that $Z \in \mathcal{L}_k - \mathcal{L}_k^A$ for all $k = m, \ldots, M$ which implies by lemma 2 that $Z \in \mathcal{Z}_k$ for all $k = m, \ldots, M$.

To estimate $\Pr_Z\left(T_{\vec{A}} < T_Z\right)$ from below, consider the following inequality:

$$\Pr_Z\left(T_{\vec{A}} < T_Z\right) \geq \Pr_Z\left(T_{\mathcal{L}_m^A} < T_Z\right) \prod_{k=m}^{M-1} \Pr_{\mathcal{L}_k^A}\left(T_{\mathcal{L}_{k+1}^A} < T_{\mathcal{Z}_k}\right). \tag{23}$$

To see why this inequality holds, first note that $\mathcal{L}_M^A = \vec{A}$. Next observe that the first term is the probability that the process visits an $m$-restricted lock-in where all teams in $R(m)$ play $A$ before getting back to $Z$. The second term is the product of probabilities. Each term in this product measures the probability that the process passes from a $k$-restricted lock-in to a $k+1$-restricted lock-in without hitting $\mathcal{Z}_k$. Along such a path the process does not hit $Z$ because $Z \in \mathcal{Z}_k$ for all $k = m, \ldots, M$. Since this chain of events

is sufficient but not necessary for $\left\{T_{\vec{A}} < T_Z\right\}$ to occur, its probability is a lower bound on $\Pr_Z\left(T_{\vec{A}} < T_Z\right)$. Thus the inequality indeed holds.

We now turn to estimate each term on the right hand side of the above inequality separately. We shall complete this estimation in two steps.

**Step 1.** Suppose that at least the teams in $D(m)$ mutate to $A$ in the initial period and no mutations to $B$ take place during the next $m$ periods. Lemma 1 implies that all teams in $R(m)$ play $A$ after at most $m$ periods, and we have not returned to $Z$. Thus

$$\Pr_Z\left(T_{\mathcal{L}_m^A} < T_Z\right) \geq K_1 \varepsilon^{\bar{\alpha}(A)m} \tag{24}$$

for some $K_1 > 0$ and $\varepsilon$ small.

**Step 2.** We prove now that the probability of getting from $\mathcal{L}_k^A$ to $\mathcal{L}_{k+1}^A$ without hitting $\mathcal{Z}_k$ is bounded from below by a constant independently of $\varepsilon$. To find this bound we recycle our construction of the shadow process from the previous lemma to construct an event that is a subset of $\left\{T_{\mathcal{L}_{k+1}^A} < T_{\mathcal{Z}_k}\right\}$.

Let $\tilde{T}_{\mathcal{Z}}^i$ be the $i$th hitting times of the shadow process $Z_t^{k,\varepsilon}$ on some subset $\mathcal{Z}$ of the $M \times M$ representation of the torus. Consider the following inequality:

$$\begin{aligned}\Pr_{\mathcal{L}_k^A}\left(T_{\mathcal{L}_{k+1}^A} < T_{\mathcal{Z}_k}\right) \geq \sum_{i=1}^{\infty} \Pr_{\mathcal{L}_k^A}\left(\tilde{T}^i_{\mathcal{L}_k^A} < \tilde{T}_{\mathcal{Z}_k}\right) \\ \times \Pr\left(\begin{array}{c}\text{The team in } D(k+1)/D(k) \text{ does not mutate} \\ \text{to } A \text{ in } Z_t^\varepsilon \text{ for the first } (i-1) \text{ times when } Z_t^{k,\varepsilon} \\ \text{hits } \mathcal{L}_k^A, \text{ but it does at the } i\text{th time.}\end{array}\right) \\ \times \Pr\left(\begin{array}{c}\text{No mutation to } B \text{ occurs in } \Lambda(M) \text{ for} \\ k+1 \text{ periods}\end{array}\right).\end{aligned}$$

The right hand side captures the following event. The shadow process starting from $\mathcal{L}_k^A$ returns to $\mathcal{L}_k^A$ several times without hitting $\mathcal{Z}_k$. One of these times the next team along the diagonal mutates to $A$. Finally, no mutation takes place for the next $k+1$ periods after that. We can write the probability of this joint event in a product form because of the Markov property and because the shadow process is independent of what mutations take place outside $R(k)$ in $Z_t^\varepsilon$ (and of course $D(k+1)/D(k)$ is outside $R(k)$).

If this joint event happens, then all teams in $R(k)$ play $A$ in $Z_t^\varepsilon$ and $Z_t^\varepsilon$ did not hit $\mathcal{Z}_k$ on the way because $Z_t^{k,\varepsilon} \subset Z_t^\varepsilon$ by construction. Since the next team along the diagonal mutated to $A$, and no mutations happened afterwards for $k+1$ periods, contagion drove all teams in $R(k+1)$ to play $A$ in the end. Therefore the probability of this joint event is indeed a lower bound for the left-hand side.

We estimate each term on the right hand side separately. The strong Markov property of the stopping times $\tilde{T}^i_{\mathcal{L}^A_k}$, $i = 1, 2, \ldots$ and Lemma 4 imply for the first event on the right hand side that

$$\Pr_{\mathcal{L}^A_k}\left(\tilde{T}^i_{\mathcal{L}^A_k} < \tilde{T}_{\mathcal{Z}_k}\right) \geq \left(1 - K\varepsilon^{\bar{\alpha}(B)k}\right)^i \geq \left(1 - K\varepsilon^{\bar{\alpha}(A)}\right)^i,$$

where we use the fact that $\bar{\alpha}(A) \leq \bar{\alpha}(B)k$ for all $k \geq m$ by construction. The second term is equal to $\left(1 - \varepsilon^{\bar{\alpha}(A)}\right)^{i-1} \varepsilon^{\bar{\alpha}(A)}$. Finally, the probability of the third term i.e., that no team mutates to $B$ for $k+1$ periods is bounded from below by $\left(1 - \varepsilon^{\bar{\alpha}(B)}\right)^{M^2(k+1)}$.

Putting these three pieces together yields

$$\begin{aligned}
\Pr_{\mathcal{L}^A_k}\left(T_{\mathcal{L}^A_{k+1}} < T_{\mathcal{Z}_k}\right) &\geq \sum_{i=1}^{\infty}\left(1 - K\varepsilon^{\bar{\alpha}(A)}\right)^i\left(1 - \varepsilon^{\bar{\alpha}(A)}\right)^{i-1}\varepsilon^{\bar{\alpha}(A)}\left(1 - \varepsilon^{\bar{\alpha}(B)}\right)^{M^2(k+1)} \\
&= \varepsilon^{\bar{\alpha}(A)}(1 - \varepsilon^{\bar{\alpha}(B)})^{M^2(k+1)}\frac{1 - K\varepsilon^{\bar{\alpha}(A)}}{K\varepsilon^{\bar{\alpha}(A)} + \varepsilon^{\bar{\alpha}(A)} - K\varepsilon^{2\bar{\alpha}(A)}} \\
&\geq \varepsilon^{\bar{\alpha}(A)}(1 - \varepsilon^{\bar{\alpha}(B)})^{M^2(k+1)}\frac{1 - K\varepsilon^{\bar{\alpha}(A)}}{(1+K)\varepsilon^{\bar{\alpha}(A)}} \geq K_2. \qquad (25)
\end{aligned}$$

for some $K_2 > 0$ independent of $\varepsilon$. The second step is complete.

Putting the results from equation (24) and (25) together proves the lemma. □

The heart of the previous lemma is step 2 where we assess the outcome of a "race": the next diagonal team needs to mutates to $A$ before too many teams at wrong locations inside $R(k)$ mutate to $B$. The probability of the event that the next diagonal team mutates to $A$ at each occasion when all teams play $A$ inside $R(k)$ is $\varepsilon^{\bar{\alpha}(A)}$. The probability of the event that the process hits a $k$-restricted lock-in with positive number of teams playing $B$ in $R(k)$ is at most $K\varepsilon^{\bar{\alpha}(B)k}$. The fact that the ratio of these two probabilities is bounded from below by a constant for $k \geq m$ implies that the set of teams playing $A$ has a strong growth property for small $\varepsilon$. This is due to two factors. One is cohesion: once a large group of teams plays $A$, it is very hard to destroy them. Secondly, contagion ensures that only one additional team is required for the large group playing $A$ to grow beyond its boundaries. Intuitively, $m$ acts as a "critical mass" in the argument; once the teams in $R(m)$ play $A$, the critical mass has been reached and it is easy to proceed.

The next theorem states our main result.

**Theorem 1** *For $M \geq m^2$,*

$$\lim_{\varepsilon \to 0} \nu_\varepsilon(\vec{A}) = 1,$$

*where m is defined in equation* (21).

**Proof.** We prove our claim in two steps. The first step focuses on $Z \in \mathcal{L}-\vec{A}$ whereas the second step considers $Z \notin \mathcal{L}$.

**Step 1.** Consider the states $Z \in \mathcal{L} - \vec{A}$. We have that

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\vec{A})} = \frac{P_{\vec{A}}(T_Z < T_{\vec{A}})}{P_Z(T_{\vec{A}} < T_Z)}.$$

Since $Z$ is a lock-in and $Z \neq \vec{A}$, $Z$ has at least $M$ teams playing $B$. It follows from Lemma 3 that

$$P_A(T_Z < T_A) = O\left(\varepsilon^{\bar{\alpha}(B)M}\right). \tag{26a}$$

Indeed, starting from $\vec{A}$, the first lock-in on the way requires at least $M$ mutations to $B$.

Moreover, the previous lemma implies

$$\frac{1}{P_Z(T_{\vec{A}} < T_Z)} = O\left(\varepsilon^{-\bar{\alpha}(A)m}\right) \tag{26b}$$

for $M \geq m$.

Putting these together yields

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\vec{A})} = O\left(\varepsilon^{\bar{\alpha}(B)M-\bar{\alpha}(A)m}\right).$$

In particular, for $M \geq m^2$,

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\vec{A})} \to 0$$

as $\varepsilon \to 0$.

**Step 2.** Consider the states $Z \notin \mathcal{L}$. Let $\hat{Z}$ be the state to which $Z$ converges in the absence of mutations. Then

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\hat{Z})} = \frac{P_{\hat{Z}}(T_Z < T_{\hat{Z}})}{P_Z(T_{\hat{Z}} < T_Z)}.$$

The denominator is of order 1, because if no mutation occurs for the next $M^2$ periods (which happens with probability approaching 1) then we hit $\hat{Z}$ without hitting $Z$ for sure. The numerator goes to zero as $\varepsilon \to 0$ because of Assumption 1. Indeed, a mutation is required in the first step, otherwise we would stay at $\hat{Z}$. The aforementioned condition guarantees that all mutation probabilities tend to zero.

Therefore

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\hat{Z})} \to 0$$

as $\varepsilon \to 0$. This holds true for $\hat{Z} = \vec{A}$ as well as other lock-ins. All in all we have that

$$\frac{\nu_\varepsilon(Z)}{\nu_\varepsilon(\vec{A})} \to 0$$

for any $Z$ such that $Z \neq \vec{A}$. The proof is complete. □

Putting together this result and Proposition 1 yields that for $N \geq 2M$ even, the only long-run stochastically stable set of the original $\{S_t^\varepsilon\}_{t\geq 0}$ process is $\vec{A}$.

It remains to consider the situation when $N$ is odd. Let $N = 2M + 1$. Consider a a $2M \times 2M$ rectangle on the torus. Note that throughout the proofs so far we have not used the fact that the environment is a torus, and not an $M \times M$ square. Thus the theorem is applicable for the $2M \times 2M$ square on the torus. Note that if all teams in the $2M \times 2M$ square are playing $A$, then contagion in the absence of mutation implies that the $(2M + 1)$st column and row will also play $A$ in a finite time period. Since mutation probabilities go to zero, we can conclude that $\vec{A}$ is the only long-run stochastically stable set of the game for $N = 2M + 1$ too. Thus the following corollary holds.

**Corollary 2** *For $N \geq 2m^2$ we have*

$$\lim_{\varepsilon\to 0} \mu_\varepsilon(\vec{A}) = 1$$

*where $m$ is defined in* (21)*, and $\mu_\varepsilon$ is the unique invariant distribution of $S_t^\varepsilon$.*

# 4 Conclusion

In contrast to the crucial role of state dependent mutation in the equilibrium selection for global interaction games, the details of the mutation structure do not matter much for local interaction games. The result strengthens the equilibrium selection result of the evolutionary game theory. Moreover it justifies the approach that deliberately avoids a detailed formulation of the decision making process. Finally considering the fact that the situations captured by the approach are mainly social phenomena involving many loosely-related agents, the local interaction framework could be a better description of the reality than the global interaction framework.

Our proof uses two basic features of the local interaction environment. First, it allows for cohesive groups based on the risk dominant strategy which make the the risk dominant strategy more resistant to mutations relative to the payoff dominant strategy. Secondly, the environment is sufficiently connected that contagion can propagate the risk dominant strategy through the whole population.

It is likely that our argument can be extended to more general settings with a higher dimension, a larger interaction range and many strategies. Using results from Blume (1995), one can construct cohesive teams for more *general interaction ranges*. Since contagion also works in a similar fashion as in our case, the main idea of our proof applicable. Furthermore, *higher dimensional environment* also allows the construction of teams (hypercubes of the corresponding dimension), and the arguments of Schonmann (1992) imply that the contagion mechanism works in the same way, hence we believe that our proof can be applied. Finally, if we consider *many strategies*, but there is a unique strategy $A$ that it is a best response to any configuration where at least half of the opponents play $A$, then stable teams can be constructed, and our proof can be applied.

We can also draw some conclusions about the relationship between mutation and contagion in a more *general local interaction environment* than the torus; see Morris (2000) and Young (1998, Chapter 6). The existence of cohesive groups and the contagion mechanism is essential for our argument to work. If both cohesion and contagion favors the risks dominant strategy, then the intuition behind our result implies that the number of mutations required to eliminate a large group playing the risk dominant strategy will depend on the size of the group. In contrast, to propagate the risk dominant strategy through contagion requires few mutations, independent of the population size. This asymmetry should ensure that results on state-dependent mutations similar to ours are obtained in a more general spatial environment.

# References

**Aldous, David and James Allen Fill**, *Reversible Markov Chains and Random Walks on Graphs*, Berkeley: Statistics Department, University of California, book in preparation 1999. http://www.stat.berkeley.edu/users/aldous/.

**Anderlini, Luca and Antonella Ianni**, "Path Dependence and Learning from Neighbors," *Games and Economic Behavior*, 1996, *13*, 141–177.

**Bala, Venkatesh and Sanjeev Goyal**, "Learning from Neighbors," *Review of Economic Studies*, 1998, *65*, 595–621.

**Bergin, James and Barton L. Lipman**, "Evolution with State Dependent Mutation," *Econometrica*, 1996, *64*, 943–956.

**Blume, Lawrence E.**, "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 1993, *5*, 387–424.

____ , "How Noise Matter," mimeo, Cornell University 1994.

____ , "The Statistical Mechanics of Best-Response Strategy Revision," *Games and Economic Behavior*, 1995, *11*, 111–145.

**Durrett, Richard**, *Probability: Theory and Examples*, second ed., Belmont, CA: Duxbury Press, 1996.

**Ellison, Glenn**, "Learning, Local Interaction and Coordination," *Econometrica*, 1993, *61*, 1047–1071.

____ , "Basins of Attraction, Long Run Stochastic Stability, and the Speed of Step-by-step Evolution," *Review of Economic Studies*, 2000, *67*, 17–45.

**Kandori, Michihiro, George J. Mailath, and Rafael Rob**, "Learning, Mutation, and Long-run Equilibria in Games," *Econometrica*, 1993, *61*, 29–56.

**Lee, In Ho and Ákos Valentinyi**, "Noisy Contagion without Mutation," *Review of Economic Studies*, 2000, *67*, 47–56.

**Morris, Stephen**, "Contagion," *Review of Economic Studies*, 2000, *67*, 57–78.

**Schonmann, Roberto H.**, "On the Behavior of Some Cellular Automata Related to Bootstrap Percolation," *Annals of Probability*, 1992, *20*, 174–193.

**van Damme, Eric and Jörgen Weibull**, "Evolution and Refinement with Endogenous Mistake Probabilities," mimeo, Center for Economic Research, Tilburg University 2000.

**Young, Peyton**, "The Evolution of Conventions," *Econometrica*, 1993, *61*, 57–84.

____ , *Individual Strategy and Social Structure*, Princeton, New Jersey: Princeton University Press, 1998.