Robust Conventions and the Structure of Social Networks*

Colin Stewart[†]

May 30, 2007

Abstract

This paper considers the equilibrium selection problem in coordination games when players interact on an arbitrary social network. We examine the impact of the network structure on the robustness of the usual risk dominance prediction as mutation rates vary. For any given network, a sufficiently large bias in mutation probabilities favoring the non-risk dominant action overturns the risk dominance prediction; bounds are obtained on the size of this bias depending on the network structure. As the size of the population grows large, the risk dominant equilibrium is highly robust in some networks. This is true in particular if the risk dominant action spreads contagiously in the network and there does not exist a sufficiently cohesive finite group of players. Examples demonstrate that robustness does not coincide with fast convergence.

KEYWORDS: Conventions, Networks, Equilibrium Selection, Risk Dominance, Coordination, Local Interaction

---

[*] I am grateful to Stephen Morris for many valuable discussions. I also thank Dirk Bergemann and seminar participants at Yale University and the VI Trento Summer School in Adaptive Economic Dynamics for their comments.

[†] Department of Economics, Yale University; email: colin.stewart@yale.edu

1 Introduction

In large population coordination games, social conventions are often thought to provide common expectations of behavior, thereby allowing coordination on a particular equilibrium. A fundamental problem in the literature on conventions has been to understand which properties lead to the selection of a particular convention in the presence of multiple equilibria. Unique outcomes generally emerge in the long-run under evolutionary dynamics in which agents play myopic best responses except for a small probability of mutation. These outcomes are called *stochastically stable.* The basic question posed in this paper is the following: how robust is the stochastically stable equilibrium selection to changes in mutation rates?

Foster and Young (1990) and Kandori, Mailath, and Rob (1993) introduced the criterion of stochastic stability, and showed that when players are randomly matched to play a 2×2 coordination game, coordination on the risk dominant action is the unique stochastically stable outcome. Bergin and Lipman (1996) criticize this approach on the grounds that the results are sensitive to the formulation of mutation rates that are freely chosen by the modeller. They show, in particular, that any equilibrium of the dynamic process without mutations is selected for some specification of mutation probabilities. When the matching process selects from neighbors in a social network, the risk dominance prediction holds for the specific formulations of mutation rates that have appeared in the literature, regardless of the structure of the network (Pęski (2004), Young (1998)). However, the robustness of this prediction as the mutation rates vary depends heavily on this structure. The main results of this paper provide sufficient conditions on the structure of the interaction network to guarantee robustness.

First we consider a fixed population interacting on a given network. The risk dominance prediction is found to be robust when all players interact with roughly the same number of other players, and there do not exist small, highly cohesive clusters in the network. If there exist some players who interact with many others, then it is possible that mutation by these players alone could influence enough of the population to move away from the risk dominant action to a different equilibrium. If the number of required mutations is small, then the risk dominant equilibrium will tend not to be robust. Similarly, if there exists a small, highly cohesive set of players in the network, then mutation by these players alone suffices to move away from the equilibrium coordinated on the risk dominant action.

In general, a sufficiently large bias in mutation probabilities always suffices to overturn the risk dominance prediction in a population of a given size. However, as the size of the population grows, there exist networks for which this prediction is robust to arbitrarily large mutation biases. As in the case of a fixed population, the nonexistence of a finite highly-cohesive cluster in the network is a necessary condition for the risk dominance prediction to be robust. When the risk dominant action spreads contagiously in the network, this nonexistence condition is also sufficient. Intuitively, contagion makes it possible to reach coordination on the risk dominant action from any initial strategy profile with a relatively small number of mutations. As the size of the population grows, a much larger number of mutations is required to move away from this coordinated equilibrium. When mutation probabilities are small, this implies that transitions to the risk dominant equilibrium occur much more frequently than transitions away from it, regardless of the specification of mutation rates.

Ellison (1993) argues that when mutation probabilities are small, the expected time to convergence to a stochastically stable outcome may be unreasonably long for practical applications. Furthermore, he demonstrates that the stucture of the interaction network has a strong influence on the speed of convergence. Intuitively, fast convergence occurs if "not many" simultaneous mutations are required for best response dynamics to lead play to the predicted outcome from any initial strategy profile. It may therefore be tempting to believe that the structural properties generating fast convergence should coincide with those leading to robustness to varying mutation rates. Section 7 presents examples demonstrating that this intuition is false: neither of these properties is sufficient to guarantee the other. This suggests that in order to evaluate the relevance of stochastically stable outcomes in a particular game, it is necessary to examine both the speed of convergence and the robustness to varying mutation rates.

2 Literature review

The model studied in this paper is based on that of Kandori, Mailath, and Rob (1993) (henceforth KMR), in which a large population of agents are matched in each of an infinite sequence of periods to play a 2×2 coordination game. Matches are drawn according to a uniform distribution over the entire population. Each player chooses a best response to the distribution of actions in the preceding period, except for a small probability of mutation, in which the action is chosen randomly according to a uniform distribution. Mutation probabilities are independent across players and periods, and constant across players and strategy profiles. If there is a strictly risk dominant action, A, in the 2×2 game, KMR show that as the mutation probabilities tend to zero, the probability that the population will be coordinated on A in a given period tends to one in the long-run. In other words, coordination on A is the unique stochastically stable state.

Young (1998) and Pęski (2004) consider variants of the KMR model in which each player is matched according to a uniform distribution over a particular subset of the population, namely that player's neighbors in a fixed social network. Under the uniform mutation rates considered by Pęski and the payoff-dependent mutation rates used by Young, the risk dominance result continues to hold subject to mild regularity conditions. Goyal and Vega-Redondo (2005) and Hojman and Szeidl (2006) find similar results when players interact in endogenously formed networks as long as the cost of forming links is small.

A more general model of evolutionary processes is employed by Bergin and Lipman (1996), who show that any distribution over states that is stable in the process without mutations can be obtained as the (unique) long-run distribution as mutation probabilities tend to zero for some specification of mutation rates. Blume (2003) addresses this critique of stochastic stability results by considering a class of payoff-dependent mutation probabilities, and identifying conditions on these probabilities under which the usual results are preserved.

Lee, Szeidl, and Valentinyi (2003) study the robustness of the risk dominance prediction to varying mutation rates when players lie on a 2-dimensional torus, and each interacts with her four nearest neighbors. They show that, for given mutation rates, as the size of the torus grows large, the risk dominant equilibrium will eventually be stochastically stable. In contrast, with the general interaction structures considered here, the analogous result need not hold.

3 The model

A population of N agents forms the nodes of a (social) network $\Gamma = (V, L)$, where V is the set of nodes, and L is a set of unordered pairs of distinct elements of V. The elements of L are called the *links* of the network, and nodes $i, j \in V$ are said to be *neighbors* in Γ if $\{i, j\} \in L$. We will say that i *interacts with* j if i and j are neighbors in Γ. Note that the neighbor relation is symmetric, so that i interacts with j whenever j interacts with i.

Each agent plays one of two actions, A or B. Payoffs from each interaction are given by a function $u(\cdot, \cdot)$ corresponding to the matrix

$$\begin{array}{c|c|c|} & A & B \\ \hline A & a, a & c, d \\ \hline B & d, c & b, b \\ \hline \end{array}.$$

The 2×2 game with these payoffs will be referred to as the *underlying game*.

The following restrictions are imposed on the payoffs:

1. (A, A) and (B, B) are Nash equilibria; that is, $a > d$ and $b > c$.

2. A is strictly risk dominant; that is, $a + c > b + d$.

The first of these conditions ensures that the equilibrium selection problem is nontrivial. The case that has received the most attention in the literature is when the risk dominant and payoff dominant equilibria differ; under the best response dynamics considered here, however, payoff dominance plays no role, so it is not necessary to identify the payoff dominant equilibrium. As will become clear below, the analysis is trivial when risk dominance is not strict (i.e. when $a + c = b + d$), and the choice of A as the strictly risk dominant action is therefore without loss of generality.

Player i's payoff $U_i\left(s_i, s_{-i}\right)$ from playing s_i when the remaining agents play the profile s_{-i} is given by adding the payoffs $u\left(s_i, s_j\right)$ over all neighbors j of i. Formally, payoffs are given by

$$U_i\left(s_i, s_{-i}\right) = \sum_{\{i,j\} \in L} u\left(s_i, s_j\right).$$

In order to distinguish it from the underlying game, the N player game with these payoffs will be referred to as the *population game*. Assume for simplicity that each agent's best response correspondence is single-valued. In other words, letting $\delta(i) = \#\left\{j \in V | \{i,j\} \in L\right\}$, assume that $\frac{b-c}{a-d+b-c}\delta(i)$ is not an integer for any i.

In the *unperturbed best response dynamics*, the population game is played over infinitely many periods $t = 0, 1, \ldots$. Starting from some strategy profile in period 0, in each period $t \geq 1$, each player updates her action with independent probability $\pi \in (0, 1)$, and otherwise plays the same action as in period $t-1$. When updating, players myopically choose the best response to the strategy profile played in the preceding period. The reason for introducing randomness into the updating process is technical: it guarantees that the stable states of the unperturbed process are precisely the Nash equilibria of the game. This is formalized in Lemma 2.

Note that action A is a best response for player i in the population game if and only if the fraction of her neighbors choosing action A is at least

$$p = \frac{b-c}{a-d+b-c}.$$

Therefore, for any initial strategy profile, the unperturbed dynamics depend only on the values of π and p.

The *perturbed best response dynamics* agree with the unperturbed dynamics except that players may "mutate" by switching to an action that is not a best response. Fix $\alpha > 0$ and $\varepsilon \in (0, 1)$. In each period in which a given player i is called upon to update her strategy, i mutates to B with probability ε^{α} if A is the best response to the strategy profile of the previous period, and mutates to A with probability ε if B is the best response. In both cases, player i plays her best response otherwise. Note that ε and α depend neither on the player nor the state. Random draws are independent across players and time.

The parameter α captures the bias in mutations toward action A. When α is small, players are much more likely to mutate to action B when A is a best response than they are to mutate to A when B is a best response. If mutations are interpreted as experimentation by boundedly rational players, such a bias may result, for example, from a tendency to try to attain the payoff-dominant outcome. As usual in models of this type, only the orders of magnitude of the mutation probabilities are relevant in determining the stochastically stable outcomes. If the probability ε^{α} were to be replaced by $\alpha\varepsilon$ in the above formulation, then α may affect the stationary distribution over outcomes, but not the set of stochastically stable states.

Both the unperturbed and the perturbed best response dynamics define finite Markov chains whose states are the strategy profiles of the population game. Recall that two states σ, σ' are said to *communicate* in a Markov chain if, beginning from σ, there is a positive probability that σ' will occur within a finite number of periods, and vice versa. A *recurrent class* is a set of states within which each pair of states communicate, and from which no other state occurs with positive probability in finite time. A Markov chain is *irreducible* if the entire state space forms a recurrent class; otherwise, it is *reducible.* A finite Markov chain possesses a unique stationary distribution if and only if it is irreducible (see, e.g., Young 1998).

Whereas the unperturbed best response dynamics form a reducible Markov chain, the Markov chain defined by the perturbed dynamics is irreducible. For each ε and α, let $\mu^{\varepsilon,\alpha}(\cdot)$ denote the stationary distribution of the perturbed process.

Definition 1 *Given α, the state σ is* stochastically stable *if* $\lim_{\varepsilon\to 0}\mu^{\varepsilon,\alpha}(\sigma) > 0.$

The main question to be addressed here concerns the extent to which, depending on the structure of the interaction network, mutations must be biased in favor of action B in order to overturn the risk dominance prediction. Accordingly, define the *mutation robustness threshold* $\overline{\alpha}$ to be

$$\overline{\alpha} := \inf\left\{\alpha \mid \sigma_A \text{ is stochastically stable}\right\},$$

where σ_A denotes the state in which all players play action A. If $\alpha < \overline{\alpha}$, then the stochastically stable states may contain the equilibrium σ_B coordinated on B, or may contain only *coexistent conventions*, equilibria in which the population is not coordinated on a single action.

4 Ellison's radius and coradius

This section introduces some Markov chain terminology and describes Ellison's radius-coradius method (Ellison 2000), which will be used to compute bounds on the threshold value $\overline{\alpha}$. The reader who is familiar with this material may wish to jump ahead to the next section.

Fix $\alpha > 0$, and let $P_\varepsilon(\sigma,\sigma')$ denote the transition probability from σ to σ' in the Markov chain describing the perturbed dynamics. For any states σ,σ', define the *transition cost*

$c(\sigma, \sigma')$ to be the unique real number satisfying

$$\lim_{\varepsilon \to 0} \frac{P_\varepsilon(\sigma, \sigma')}{\varepsilon^{c(\sigma,\sigma')}} \in (0, \infty).$$

Note that since best response updating is random, there may be different ways to transition from σ to σ' in a single period depending on which players update and which mutate. The cost $c(\sigma, \sigma')$ is the minimum value of the sum $n_A + \alpha n_B$ over all such single-period transitions, where n_A is the number of mutations required from B to A, and n_B is the number required from A to B.

A *path* from σ to σ' is a finite sequence $(\sigma_0, \sigma_1, \ldots, \sigma_n)$ of distinct states such that $\sigma_0 = \sigma$ and $\sigma_n = \sigma'$. Let $\Pi(\sigma, \sigma')$ denote the set of all such paths. Define the *cost* $c(\overline{\sigma})$ of the path $\overline{\sigma} = (\sigma_0, \sigma_1, \ldots, \sigma_n)$ to be

$$c(\overline{\sigma}) = c(\sigma_0, \sigma_1) + c(\sigma_1, \sigma_2) + \cdots + c(\sigma_{n-1}, \sigma_n).$$

Paths that minimize the transition cost between states play a special role in identifying the stochastically stable states, as these are the transitions that occur most frequently in the limit as mutation probabilities vanish. Accordingly, for each pair of states σ, σ', define the *minimal cost* $m(\sigma, \sigma')$ by

$$m(\sigma, \sigma') = \min_{\overline{\sigma} \in \Pi(\sigma, \sigma')} c(\overline{\sigma}).$$

Let Ω be a union of recurrent classes of the unperturbed dynamics. The *basin of attraction* $B(\Omega)$ of Ω is the set of states from which some state in Ω is almost surely reached in finite time under the unperturbed dynamics. Equivalently, $B(\Omega)$ consists of those states from which there exists a zero-cost path to some state in Ω, but there exists no such path to any recurrent class not in Ω.

Ellison (2000) defines the *radius* $R(\Omega)$ of the basin of attraction of Ω by

$$R(\Omega) := \min_{\sigma \in \Omega, \sigma' \notin B(\Omega)} m(\sigma, \sigma').$$

Thus the radius $R(\Omega)$ is the lowest cost associated with any transition that does not almost surely return to Ω under the unperturbed dynamics. Similarly, the *coradius* $C(\Omega)$ of the basin of attraction of Ω is defined by

$$C(\Omega) := \max_{\sigma' \notin B(\Omega), \sigma \in \Omega} m(\sigma', \sigma).$$

Thus, starting from any initial state, the coradius of Ω is the greatest cost that could be necessary in order to reach Ω. Ellison shows that if $R(\Omega) > C(\Omega)$ then Ω contains the set of stochastically stable states. Intuitively, when this is the case, transitions to Ω occur more frequently than transitions away from it, so as the mutation probabilities approach zero, much more time is spent at states in Ω than at any other state.

Since the radius-coradius condition is sufficient but not necessary, it can be used to identify upper bounds on the threshold $\overline{\alpha}$ by taking $\Omega = \{\sigma_A\}$. If, for some β, $R(\{\sigma_A\}) > C(\{\sigma_A\})$ whenever $\alpha > \beta$, then we have $\overline{\alpha} \leq \beta$. Similarly, by taking Ω to be the union over all recurrent classes except $\{\sigma_A\}$, then we obtain a lower bound on $\overline{\alpha}$ by identifying β such that $R(\Omega) > C(\Omega)$ whenever $\alpha < \beta$.

5 Fixed populations

Given a fixed network, it is possible to derive bounds on the mutation robustness threshold $\overline{\alpha}$ based on the network structure and the value of the payoff parameter p. Since both of the methods described in the previous section involve transition costs among recurrent classes of the unperturbed dynamics, we begin by identifying these classes.

Lemma 2 *The recurrent classes of the unperturbed best response dynamics are precisely the singleton sets containing the Nash equilibria of the population game.*

Proof. Clearly each Nash equilibrium forms a recurrent class.

For the converse, we must show that, beginning from any state σ_0, a Nash equilibrium will be reached with positive probability in finite time. It suffices to construct a finite sequence of states $\sigma_0, \sigma_1, \ldots, \sigma_m$ such that for each $k = 1, \ldots, m$, σ_k differs from σ_{k-1} only through best response updating by a single player. Without loss of generality, suppose that action A is a best response for some player who plays action B under σ_0. Choose any such player i, and define σ_1 to be equal to σ_0 except that player i plays action A. Repeat this step until a state σ_r is reached at which no such player remains. Now repeat this process beginning from σ_r, except with actions A and B reversed.

I claim that the final state σ_m attained under this process is a Nash equilibrium. Suppose for contradiction that player i plays an action σ_m^i that is not a best response under σ_m. It is clear by construction that $\sigma_m^i = B$ and player i's best response under σ_m is A. Since the number of i's neighbors playing action A is nonincreasing along the path $\sigma_r, \ldots, \sigma_m$, A must also be a best response for player i under each σ_k for $k = r, \ldots, m$. Therefore, player i must choose action B under σ_r, contradicting the construction of σ_r. ■

Pęski (2004) considers the special case of the present model in which $\alpha = 1$. He shows that the state σ_A is stochastically stable regardless of the structure of the network, which, in our terminology, immediately implies the upper bound $\overline{\alpha} \leq 1$ on the mutation robustness threshold. To obtain tighter bounds, some definitions are required concerning structural properties of networks.

Definition 3 *The* degree $\delta(i)$ *of node* i *is the number* $\#\{j \in V \mid \{i,j\} \in L\}$ *of its neighbors in the network.*

Let $\delta_{\min} = \min_{i \in V} \delta(i)$ and $\delta_{\max} = \max_{i \in V} \delta(i)$.

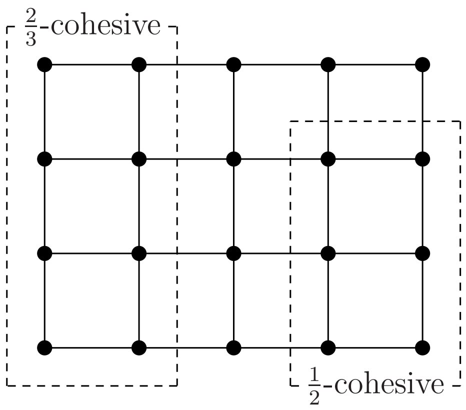

Figure 1: $\frac{2}{3}$-cohesive and $\frac{1}{2}$-cohesive sets of nodes in a finite lattice.

Definition 4 *Given $r \in [0,1]$, a subset S of the set of nodes V of the network $\Gamma = (V, L)$ is* r-cohesive *in Γ if for every $i \in S$,*

$$\#\{j \in S \mid \{i,j\} \in L\} \geq r\#\{j \in V \mid \{i,j\} \in L\}.$$

In words, each node in S has a fraction of at least r of its neighbors in S.

Figure 4 exhibits two sets of r-cohesive nodes in a finite lattice. For each set, the given value of r is the largest for which the set is r-cohesive.

The cohesiveness of sets of nodes in the network is directly related to the best response dynamics of the interaction game. If i lies in a p-cohesive set of nodes S, then A is a best response for i whenever all other players in S play A. Similarly, if S is $(1-p)$-cohesive, then B is a best response for i whenever all other players in S play B. Identifying each strategy profile with the set S of agents playing A, S is a Nash equilibrium if and only if it is p-cohesive and its complement $V \setminus S$ in V is $(1-p)$-cohesive.

Consider the $(1-p)$-cohesive sets in Γ. These are partially ordered by inclusion and include the sets $\emptyset$ and V. A *chain* of $(1-p)$-cohesive sets of length l is an increasing sequence of $l+1$ distinct sets $V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_l$. Such a chain is *maximal* if $V_0 = \emptyset$, $V_l = V$, and for each $k = 1, \ldots, l$, there does not exist any $(1-p)$-cohesive set U such that

$V_{k-1} \subsetneq U \subsetneq V_k$.

Lemma 5 *When $\alpha = 1$, the coradius of the basin of attraction of σ_A is at most the length l of the shortest maximal chain of $(1-p)$-cohesive sets.*

Proof. Let σ_B denote the state in which all players choose action B, and identify each state with the set of agents choosing action B. Thus, in particular, $\sigma_A = \emptyset$ and $\sigma_B = V$. First we show that $C(\sigma_A) = m(\sigma_B, \sigma_A)$. For each state σ let $B(\sigma)$ denote the set of agents for which B is a best response to σ in the population game. Given any path $\overline{\sigma} = (\sigma_0, \ldots, \sigma_k) \in \Pi(\sigma_B, \sigma_A)$, define the set

$$M_A(\overline{\sigma}) := \{v \in V | \exists j \in \{1, \ldots, k\} \text{ such that } v \in \sigma_{j-1} \setminus \sigma_j \text{ and } v \in B(\sigma_{j-1})\}.$$

Thus the set $M_A(\overline{\sigma})$ consists of all nodes that mutate to action A at some point along the path $(\overline{\sigma})$.

I claim that for any state σ, there exists a zero-cost path from $\sigma \setminus M_A(\overline{\sigma})$ to σ_A; that is, beginning from σ, if all agents in $M_A(\overline{\sigma})$ switch to (or remain at) action A, then it is possible to reach σ_A through best responses alone. If not, then let $j < k$ be the largest index for which there exists a state $\sigma' \subset (\sigma_j \setminus M_A(\overline{\sigma}))$ such that $m(\sigma \setminus M_A(\overline{\sigma}), \sigma') = 0$. Then there is some $v \in \sigma' \setminus \sigma_{j+1}$ for which $v \in B(\sigma')$. Since $\sigma' \subset \sigma_j$, $v \in B(\sigma')$ implies that $v \in B(\sigma_j)$, and hence $v \in M_A(\overline{\sigma})$, contradicting that $v \in \sigma'$ and $\sigma' \cap M_A(\overline{\sigma}) = \emptyset$.

The claim implies that, for any state σ and any path $\overline{\sigma} \in \Pi(\sigma_B, \sigma_A)$, the minimal cost $m(\sigma, \sigma_A)$ is at most $|M_A(\overline{\sigma})|$. If $\overline{\sigma}$ is a cost-minimizing path from σ_B to σ_A, then it cannot involve any mutations to action B. Thus we have

$$m(\sigma_B, \sigma_A) = |M_A(\overline{\sigma})| = \max_{\sigma} m(\sigma, \sigma_A),$$

and therefore, $C(\sigma_A) = m(\sigma_B, \sigma_A)$.

Let $\emptyset = V_0 \subset V_1 \subset \cdots \subset V_l = V$ be a shortest maximal chain of $(1-p)$-cohesive sets. For $i \in \{1, \ldots, l\}$, consider the initial state $\sigma_0 = V_i$ in which precisely the agents in V_i choose action B. Suppose some agent $v_0 \in V_i \setminus V_{i-1}$ mutates to action A, so that the state becomes $\sigma_1 = V_i \setminus \{v_0\}$. If $V_i \setminus \{v_0\} \neq V_{i-1}$, then $V_i \setminus \{v_0\}$ cannot be $(1-p)$-cohesive by the maximality of the chain $V_0 \subset \cdots \subset V_l$. Since V_{i-1} is $(1-p)$-cohesive, there must exist some $v_1 \in (V_i \setminus \{v_0\}) \setminus V_{i-1}$ such that $v_1 \notin B\left(V_i \setminus \{v_0\}\right)$. Repeating this argument with $\sigma_2 = V_i \setminus \{v_0, v_1\}$ in place of $V_i \setminus \{v_0\}$, and continuing recursively in this fashion gives rise to a path $\overline{\sigma} = (\sigma_0, \sigma_1, \ldots, \sigma_k)$ of states such that $\sigma_k = V_{i-1}$ and $c(\overline{\sigma}) = 1$. Connecting these paths in sequence over all $i \in \{1, \ldots, l\}$ gives a path in $\Pi(\sigma_B, \sigma_A)$ having cost l, as needed. ■

Definition 6 *An r-cohesive set of nodes S is a* minimal r-cohesive set *if it contains no nonempty r-cohesive proper subset.*

Theorem 7 *Let k be the size of the smallest nonempty $(1-p)$-cohesive set whose complement in V is p-cohesive (or empty), and let K be the size of the largest minimal $(1-p)$-cohesive set in V. We have the following upper bound:*

$$\overline{\alpha} \leq \frac{\left(\delta_{\max} + \delta_{\min} - 2\left\lfloor p\delta_{\min}\right\rfloor\right)(N - K + 1)}{\left(\delta_{\min} - 2\left\lfloor p\delta_{\min}\right\rfloor\right) k}.$$

Proof. Let R and C respectively denote the radius and coradius of σ_A when $\alpha = 1$. For general α, the coradius of σ_A is the same, whereas its radius is equal to αR. The state σ_A is therefore stochastically stable if $\alpha R > C$, indicating that the ratio $\frac{C}{R}$ is an upper bound on the threshold $\overline{\alpha}$. The proof proceeds in two steps. The first is to demonstrate that $R \geq \frac{\delta_{\min} - 2\lfloor p\delta_{\min}\rfloor}{\delta_{\max} + \delta_{\min} - 2\lfloor p\delta_{\min}\rfloor} k$. Showing that $C \leq N - K + 1$ then gives the result.

Suppose that the initial state is σ_A. If the radius is R, then there exists a set S_0 of R agents and a sequence $i_1, ..., i_m$ of agents such that B is a best response for i_l if every agent in $S_0 \cup \{i_1, ..., i_{l-1}\}$ chooses B (that is, if a fraction of at least $(1-p)$ of i_l's neighbors lie in $S_0 \cup \{i_1, ..., i_{l-1}\}$), and there is a coexistent convention in which at least one agent in $S_0 \cup \{i_1, ..., i_r\}$ chooses B.

Recall that any coexistent convention corresponds to a partition of the nodes into a p-cohesive set and a $(1-p)$-cohesive set. Let S_B denote the set of agents who play B in the final equilibrium. Note that we may restrict ourselves to the subgraph containing only those links involving at least one node in S_B. Let $\delta_R\left(S_B\right)$ be the sum of the degrees of the R nodes in this subgraph having the largest degrees. That is, denoting the subgraph by $\Gamma\left(S_B\right)$, and the degree of node j in $\Gamma\left(S_B\right)$ by $\delta_{\Gamma(S_B)}\left(j\right)$, define

$$\delta_R\left(S_B\right) := \max_{\{j_1,\ldots,j_R\}} \sum_{l=1}^{R} \delta_{\Gamma(S_B)}\left(j_l\right).$$

In order for the sequential best response condition mentioned above to hold, it must be the case that for each i_l, the number of edges connecting i_l to nodes in $S_0 \cup \{i_1, ..., i_{l-1}\}$ is at least $\lceil (1-p)\,\delta\left(i_l\right)\rceil$. After adding node i_1, that leaves at most

$$\delta_R\left(S_B\right) - \lceil (1-p)\,\delta\left(i_1\right)\rceil + \delta\left(i_1\right) - \lceil (1-p)\,\delta\left(i_1\right)\rceil = \delta_R\left(S_B\right) - \left(\delta\left(i_1\right) - 2\left\lfloor p\delta\left(i_1\right)\right\rfloor\right)$$

edges to be connected to other i_l's. Continuing recursively in this fashion, we obtain the following inequality:

$$\lceil (1-p)\,\delta\left(i_r\right)\rceil \leq \delta_R\left(S_B\right) - \sum_{j=1}^{r-1}\left(\delta\left(i_j\right) - 2\left\lfloor p\delta\left(i_j\right)\right\rfloor\right).$$

This implies the weaker condition that

$$\delta_R\left(S_B\right) \geq \sum_{j=1}^{r}\left(\delta\left(i_j\right) - 2\left\lfloor p\delta\left(i_j\right)\right\rfloor\right).$$

Since $\delta(i_j) - 2\lfloor p\delta(i_j)\rfloor \geq \delta_{\min} - 2\lfloor p\delta_{\min}\rfloor$ for each i_j, this implies that

$$R\delta_{\max} \geq r\left(\delta_{\min} - 2\lfloor p\delta_{\min}\rfloor\right).$$

By definition of k, we must have $R + r \geq k$, which, when substituted for r in the last expression, gives

$$R \geq \frac{\delta_{\min} - 2\lfloor p\delta_{\min}\rfloor}{\delta_{\max} + \delta_{\min} - 2\lfloor p\delta_{\min}\rfloor} k,$$

as desired.

For the coradius, C, note that the existence of a minimal $(1-p)$-cohesive set U of size K implies that there is a maximal chain of $(1-p)$-cohesive sets $\emptyset = V_0 \subset V_1 \subset \cdots \subset V_l = V$ such that $V_1 = U$. By Lemma 5, the coradius of σ_A is at most l, which cannot exceed $N - K + 1$. ■

In the special case in which there are no coexistent conventions in the population game, the structural conditions of Theorem 7 follow from a simpler property of the network, namely, the existence of a small p-cohesive set. This observation is formalized in the following result.

Corollary 8 *Suppose that there are no coexistent conventions in the population game. If the network Γ contains a p-cohesive set of size m, then we have the bound*

$$\overline{\alpha} \leq \frac{\left(\delta_{\max} + \delta_{\min} - 2\lfloor p\delta_{\min}\rfloor\right)(m-1)}{\left(\delta_{\min} - 2\lfloor p\delta_{\min}\rfloor\right) N}.$$

Proof. If we show that in the absence of any coexistent convention, the existence of a p-cohesive group S of size m implies the existence of a minimal $(1-p)$-cohesive group of size at least $N - m + 2$, then we are done. To see this, it suffices to show that there is no $(1-p)$-cohesive group of nodes containing at most one element of S. Note first that since S is p-cohesive, any set S' of nodes containing exactly one element of S cannot be $(1-p)$-cohesive since the node in both sets cannot have enough neighbors in S'. Thus we

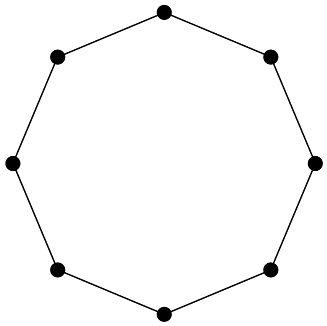

Figure 2: Nearest neighbor interaction on the circle.

need only consider sets S' disjoint from S.

Suppose for contradiction that S' is a $(1-p)$-cohesive set disjoint from S. Now apply the following recursive procedure to the remaining nodes in $V \setminus (S \cup S')$: (i) take all nodes having a fraction of at least $(1-p)$ of their neighbors in S' and assign them to S'; (ii) repeat step (i) until there are no more such nodes, and assign all remaining nodes to the set S. I claim that the resulting partition of the nodes describes a coexistent convention. By abuse of notation, let S and S' denote the resulting sets after all other nodes have been assigned, so that $S \cup S' = V$. It is clear by construction that S' is $(1-p)$-cohesive, and that each element of S has a fraction of at most $(1-p)$ of its neighbors in S'. But then since $S \cup S' = V$, each element of S must have a fraction of at least p of its neighbors in S, proving the claim. ■

The p-cohesiveness condition of the corollary may be interpreted as the existence of a small clique. In a regular network that does not support any coexistent convention for the given value of p, the existence of a single clique that is small relative to the size of the population is sufficient to guarantee the robustness of the risk dominance prediction. The regularity assumption precludes the existence of a leader who, by single-handedly changing her action, could affect the incentives of a large number of players.

For the given structural properties, Theorem 7 is tight, as the following example demonstrates.

Example 9 (Nearest-neighbor interaction on the circle) *An even number, N, of players lie at distinct points on a circle. Each agent interacts with the immediate neighbors lying to each side (see Figure 2). Formally, let $\{1, \ldots, N\}$ be the set of agents. The nodes i and j are neighbors in the network if and only if $i - j \equiv \pm 1 \pmod{N}$. For any $p \in \left(0, \frac{1}{2}\right)$, it suffices for one agent to choose action A in order for A to be a best response for both of her neighbors. Therefore, there are no coexistent conventions in the population game, as any such convention must involve at least two neighboring agents who play different actions. In order to apply Corollary 8, note that the set $\{1, 2\}$ is p-cohesive, so we may take $m = 2$. Further, we have $\delta_{\max} = \delta_{\min} = 2$, and $\lfloor p\delta_{\min} \rfloor = 0$. Corollary 8 therefore gives the bound*

$$\overline{\alpha} \leq \frac{2}{N}.$$

As N grows large, the threshold $\overline{\alpha}$ tends to zero. The risk-dominance prediction is therefore strongly robust for this network when the population is large.

To check that this bound is tight, we may compute the precise value of the threshold $\overline{\alpha}$. Since there are only two recurrent classes, σ_A and σ_B, Ellison's radius-coradius method provides both necessary and sufficient conditions for stochastic stability whenever the inequality $R(\sigma) > C(\sigma)$ is strict. Note that, beginning from σ_B, it suffices for a single agent to mutate to action A in order for there to exist a zero-cost path to σ_A. Thus we have $C(\sigma_A) = r(\sigma_B, \sigma_A) = 1$. To compute the radius of σ_A, note that if two neighboring players choose action A, then A will be a best response for both no matter what actions the other players take. Thus any path from σ_A to σ_B must involve a mutation to action B by at least one from every pair of neighboring agents, which implies that $R(\sigma_A) \geq \frac{N}{2}\alpha$. Conversely, there exists a zero-cost path to σ_B from the state σ in which precisely the even-numbered players choose action A. Since σ can be reached from σ_A by $\frac{N}{2}$ mutations to action B, we have $R(\sigma_A) = \frac{N}{2}\alpha$. Combining these results, σ_A is stochastically stable precisely when

$\frac{N}{2}\alpha \geq 1$, and therefore

$$\overline{\alpha} = \frac{2}{N},$$

demonstrating that the bound of Theorem 7 is tight.

The preceding upper bounds exploit Ellison's radius-coradius theorem to identify conditions under which coordination on the risk dominant action is stochastically stable. Applying the same theorem to the collection of all recurrent classes except for σ_A gives rise to a lower bound.

Theorem 10 *Suppose that there exists a $(1-p)$-cohesive set of size r in Γ. Then the we have the lower bound $\overline{\alpha} \geq \frac{1}{r-1}$.*

Proof. Let Ω be the set of all Nash equilibria of the population game except for σ_A. Beginning from any state outside of Ω, either there exists a zero-cost path to some state in Ω, or there exists a zero-cost path to σ_A. Thus for computing the radius and coradius of Ω it suffices to consider paths to and from σ_A.

Let S be a $(1-p)$-cohesive set of size r, and let σ denote the state in which all agents in S play action B, and all other agents play action A. Let σ' be identical to σ except that one of the agents x in S plays A. Since S is $(1-p)$-cohesive, B is the best response for x to the profile σ'. Hence we have $c(\sigma', \sigma) = 0$, and $m(\sigma_A, \sigma) \leq \alpha(r-1)$. In order to reach σ_A from σ, at least one agent in S must mutate to action A. Therefore, σ lies in the basin of attraction of Ω, and we have $C(\Omega) \leq \alpha(r-1)$. Since σ_A cannot be reached from any state in Ω without at least one mutation to action A, the radius $R(\Omega)$ is at least one. Therefore, the condition $R(\Omega) > C(\Omega)$ holds whenever $1 > \alpha(r-1)$, in which case every stochastically stable state lies in Ω. ■

The lower bound of Theorem 10 is also tight. In the trivial example of two interacting agents, the unperturbed dynamics are symmetric with respect to the two actions. The state σ_B is therefore the unique stochastically stable state whenever $\alpha < 1$, and the threshold $\overline{\alpha}$ is equal to one.

6 Large population games

One difficulty in interpreting the preceding finite network results is that, without a compelling justification for any particular mutation probabilities, the question of how small the value of the threshold $\overline{\alpha}$ must be in order to accept the risk dominance prediction remains open. In general, this value is bounded below by the reciprocal of the population size. Thus it is natural to consider conditions under which $\overline{\alpha}$ approaches zero as the population grows large, ensuring that coordination on A is stochastically stable for a wide range of mutation probabilities.

Consider a network $\Gamma = (V, L)$ on a countably infinite set of nodes V. Assume that there is a uniform upper bound $\Delta \in \mathbb{N}$ on the number of neighbors of any node; that is, assume that for all $i \in V$,

$$\#\left\{j \in V \mid \{i, j\} \in L\right\} \leq \Delta.$$

The approach taken here to understand stochastic stability in large population games will be to consider a increasing sequences of finite networks that approach the infinite network in the limit. Such a sequence may be obtained from a *labelling* of the set of nodes V, that is, from a bijection $\iota : \mathbb{N} \longrightarrow V$. Given any labelling ι, define for each $n \in \mathbb{N}$ the subnetwork $\Gamma_\iota(n)$ of Γ whose nodes are given by the set $V_\iota(n) := \iota\left(\{1, \dots, n\}\right)$, and whose links $L_\iota(n)$ consist of all links in L between any two nodes in $V_\iota(n)$; thus

$$L_\iota(n) := \left\{\{i, j\} \in L \mid \iota^{-1}\left(\{i, j\}\right) \subset \{1, \dots, n\}\right\}.$$

Fixing the payoffs in the underlying 2×2 game, define for each $n \in \mathbb{N}$ the mutation robustness threshold $\overline{\alpha}_\iota(n)$ to be the value of $\overline{\alpha}$ for the network $\Gamma(n)$.

Definition 11 Mutation robustness *holds in the infinite network* $\Gamma = (V, L)$ *if there exists a labelling* ι *of* V *such that* $\lim_{n \to \infty} \overline{\alpha}_\iota(n) = 0$.

If mutation robustness holds, then the range of mutation probabilities giving rise to σ_A

as a stochastically stable outcome can be made arbitrarily large by taking a sufficiently large population. Note that the set of limit points of the sequence $\overline{\alpha}_\iota(n)$ depends in general on the choice of labelling ι. For example, it is always possible to choose a labelling such that, for each n, the network $\Gamma(n)$ contains at least one isolated node that has no neighbors. In this case, since such an isolated node forms a $(1-p)$-cohesive set of size 1, it follows from Theorem 10 that $\overline{\alpha}_\iota(n) = 1$ for all n.

Morris (2000) characterizes conditions under which the risk dominant action spreads contagiously in a similar model, which differs only in that the dynamics are deterministic: in each period, every player chooses a best response to the strategy profile of the previous period. *Contagion* occurs in the infinite network Γ if, starting from some initial strategy profile in which only a finite number of agents play action A, every member of the population plays A in the limit as time tends to infinity. Morris shows in particular that contagion occurs in Γ if and only if there exists a labelling ι of the nodes of Γ such that for some sufficiently large $N \in \mathbb{N}$, A is a best response for $\iota(n)$ whenever each node $\iota(1), \ldots, \iota(n-1)$ plays A and $n \geq N$.

Although we have not defined analogues of Ellison's radius and coradius for games played on infinite networks, the occurrence of contagion corresponds intuitively to the coradius of σ_A being finite. Thus mutation robustness should hold as long as the radius of σ_A is infinite. Morris (2000) shows that it is impossible for action B to spread to an infinite set of agents from an initial strategy profile in which only a finite set of agents play B, suggesting that mutation robustness should hold as long as there is no coexistent convention in which a finite set of agents plays B. This intuition is formalized in the following theorem.

Theorem 12 *If there exists a finite $(1-p)$-cohesive set of nodes in Γ then mutation robustness does not hold. Conversely, if contagion occurs in Γ and there does not exist a $(1-p)$-cohesive set of nodes, then mutation robustness holds.*

Proof. For the first part, let S be a $(1-p)$-cohesive set of nodes in Γ of finite size m. It suffices to note that given any labelling ι, there exists some N such that ι assigns a label of at most N to every node in S. By Theorem 10, $\overline{\alpha}_\iota(n) \geq \frac{1}{m}$ for all $n > N$.

For the converse, first consider, for each n, the coradius of σ_A in the game played on the network $\Gamma_\iota(n)$. Since contagion occurs, there exists some labelling ι for which there is some absolute bound M_{CR} and some sufficiently large N such that whenever $n \geq N$, this coradius is at most M_{CR}. To prove this, choose a finite set S of nodes from which, if all of these choose A, best response dynamics lead to all agents choosing A. Consider best response dynamics where, in period 0, only members of S choose action A. For each node v, there is some earliest period $k(v)$ after which A is always a best response for v as the best response dynamics are iterated. The desired labelling is any for which $k \circ \iota$ is nondecreasing (that is, the lowest labels are assigned to the nodes that switch to A earliest). Let $M_{CR} = |S|$. It is clear by construction that, beginning from σ_B, mutation of all M_{CR} nodes in S is sufficient to lead to σ_A.

All that remains is to show that for some labelling ι satisfying the requirement of the preceding paragraph, the radius of σ_A in $\Gamma_\iota(n)$ tends to infinity as n grows large. Note that any $(1-p)$-cohesive group in $\Gamma_\iota(n)$ must contain some member of S by the way in which the labelling ι was chosen. Recall that the degrees of the nodes of Γ are uniformly bounded by some number Δ. For each n, and each $d \in \mathbb{N}$, let $g^d(n) \in \mathbb{N}$ be the smallest number for which all nodes within distance d of any node in $\iota(\{1, ..., n\})$ are in $\Gamma_\iota\left(g^d(n)\right)$.

Suppose that the radius of σ_A in $\Gamma_\iota(n)$ does not tend to infinity with n. Then there exists some $M_R \in \mathbb{N}$ such that for each N, there exists some $n > N$ for which the radius of σ_A in $\Gamma_\iota(n)$ is at most M_R. I claim that there exists a number K, depending only on p and Δ, such that beginning from σ_A, any number m of mutations to B can lead, through best response dynamics, to at most Km players choosing B. Assuming for now that the claim is true, let $N = g^{KM_R+1}(|S|)$. For some $n > N$, there exists a $(1-p)$-cohesive set S_{1-p} in $\Gamma_\iota(n)$ of size at most KM_R, for otherwise the radius of σ_A would be greater than M_R for all $n > N$. As noted above, this set S_{1-p} must contain some element of S. Since we may assume without loss of generality that this set is connected (otherwise take some component), it follows that every node in S_{1-p} lies within distance KM_R of some member of S, and therefore that every neighbor of every node in S_{1-p} lies in $\Gamma_\iota(n)$. But then S_{1-p} is $(1-p)$-cohesive in Γ, contradicting the assumption that Γ contains no finite $(1-p)$-cohesive group.

All that remains is to prove the claim of the preceding paragraph. Accordingly, let S_{1-p} be any (finite) $(1-p)$-cohesive group of size M in an arbitrary network Ω, and suppose that m mutations suffice for the members of S_{1-p} to switch to playing B. Then there exists a labelling $\kappa : \{1, ..., M\} \twoheadrightarrow S_{1-p}$ such that for each $n > m$, a fraction of at least $(1-p)$ of $\kappa(n)$'s neighbors in Ω lie in the set $\{1, ..., n-1\}$. For each $l = 1, ..., M$, let δ_l be the degree of $\kappa(l)$. For each $l > m$, there must be at least $(1-p)\,\delta_l$ links connecting $\kappa(l)$ to nodes with smaller labels, and hence at most $p\delta_l$ links connecting $\kappa(l)$ to nodes with higher labels. Thus we have

$$\sum_{l=1}^{m} \delta_l + p \sum_{j=m+1}^{M} \delta_j \geq (1-p) \sum_{k=m+1}^{M} \delta_k.$$

Assuming a uniform upper bound of Δ on the degrees of the nodes in Ω, this implies that

$$m\Delta \geq (1-2p) \sum_{k=m+1}^{M} \delta_k.$$

Assuming that Ω contains no solitary nodes, so that $\delta_k \geq 1$ for all k, this gives

$$m\Delta \geq (1-2p)\,(M-m)\,,$$

and therefore,

$$m \geq \frac{(1-2p)}{\Delta + (1-2p)} M.$$

Taking $K = \frac{\Delta+(1-2p)}{(1-2p)}$ therefore gives the desired result. ■

In a similar model, Lee et al. (2003) consider the interaction structure formed by a 2-dimensional torus, and find that an analogue of mutation robustness holds for all values of p, that is, given any model of the mutation probabilities, the risk dominant equilibrium is stochastically stable when the population on the torus is sufficiently large. As they explain, this result is driven by the existence of small p-cohesive sets of nodes that cover the entire network combined with a stochastic form of contagion. Significantly, however, the result depends on the fact that, for each $p \in \left(0, \frac{1}{2}\right)$, the size of the smallest $(1-p)$-cohesive set of nodes on the torus grows without bound as the size of the network grows large. On the other hand, the stochastic contagion underlying their robustness result is a weaker property than the deterministic contagion of Theorem 12, suggesting that it may be possible to generalize this result.

In networks possessing enough symmetry, contagion cannot occur if there exists a finite $(1-p)$ cohesive set of nodes. To be precise about the relevant notion of symmetry, we require the following definition:

Definition 13 *an* automorphism ϕ *of* Γ *is a bijection* $\phi : V \longrightarrow V$ *such that* x *and* y *are neighbors in* Γ *if and only if* $\phi(x)$ *and* $\phi(y)$ *are neighbors in* Γ.

Thus an automorphism of a network is a permutation of its nodes that preserves the link structure, and each nontrivial automorphism corresponds to a symmetry of the network.

Proposition 14 *Suppose that for each* $x \in V$ *there exist infinitely many* $y \in V$ *such that there is some automorphism* ϕ *of* Γ *satisfying* $\phi(x) = y$. *Then contagion cannot occur if there exists a finite* $(1-p)$*-cohesive set in* Γ.

Proof. Suppose that contagion occurs in Γ, and that there exists a finite $(1-p)$-cohesive set of nodes C. Then there exists a finite set of nodes S whose complement does not contain a $(1-p)$-cohesive set (see Morris (2000)). Let d_C be the diameter of the set C; that is, d_C is the greatest distance between any two nodes in C. Since S is finite and the degrees of the nodes of Γ are uniformly bounded, given any $d \in \mathbb{N}$, there are only finitely many nodes y for which there exists a node in S within distance d of y. Therefore there exists some y lying at a distance strictly greater than d from all nodes of S such that y is the image of some $x \in C$ under some automorphism ϕ of Γ. By construction, the set $\phi(C)$ is disjoint from S, and $\phi(C)$ is $(1-p)$-cohesive since C is, contradicting the choice of S. ■

The following corollary is immediate from Theorem 12 and Proposition 14.

Corollary 15 *Under the symmetry assumption of Proposition 14, the occurrence of contagion is a sufficient condition for mutation robustness.*

7 Waiting times

Ellison (1993) argues that the relevance of stochastically stable outcomes depends on the expected waiting time to convergence, which in turn depends on the interaction structure. Young (1998) extends an argument due to Ellison (1993) to bound the expected waiting time in local interaction games when each node in the network lies in a sufficiently close-knit group. Close-knittedness is a clustering property similar to, but stronger than, the r-cohesiveness used above. Fast convergence occurs under Young's conditions because the required mutations can take place in small steps, each of which is much more likely to occur than are many simultaneous mutations.

In a more general setting, Ellison (2000) bounds the expected waiting time using only the coradius of the set of stochastically stable states, showing that a small coradius is sufficient to ensure fast convergence. Since a small coradius of σ_A tends to favor mutation robustness, one might expect fast convergence and mutation robustness to be closely related. The examples below demonstrate that this intuition is false in general. First, however, we must give a more precise definition of fast convergence for large networks.

Consider a sequence $\overline{\Gamma} = (\Gamma_1, \Gamma_2, \ldots)$ of networks such that $|V(\Gamma_{N'})| > |V(\Gamma_N)|$ whenever $N' > N$; that is, the size of the population is strictly increasing along the sequence.

Any labelling ι of the nodes of an infinite network Γ naturally gives rise to such a sequence $\overline{\Gamma}_\iota$ by taking $\Gamma_N = \Gamma_\iota(N)$ for all N.

Definition 16 Fast convergence *occurs in* $\overline{\Gamma}$ *if there exists some* T *not depending on* N *such that, for each* N*, from any initial state, the expected time until* σ_A *is reached in* Γ_N *is* $O\left(\varepsilon^{-T}\right)$ *when* $\alpha = 1$.

This definition captures the idea that, for convergence to be fast, the expected waiting time should not grow by orders of magnitude as the network becomes large. Since best response dynamics typically require more periods to adjust following mutations in a large network compared to a small one, the constant implied by the big-O will generally depend on N. However, the order of magnitude, as measured by the exponent T, must remain bounded as the network grows large.

We say that fast convergence occurs in an infinite network Γ if there exists a labelling ι such that fast convergence occurs in the sequence $\overline{\Gamma}_\iota$. If contagion occurs in Γ, then there exists a labelling ι and a set of nodes S of size K such that, for large enough N, any state in which all members of S choose action A lies in the basin of attraction of σ_A in the network $\Gamma_\iota(N)$. In particular, fast convergence holds in Γ with $T = K$ since, from any initial state, mutation of all members of S to action A is sufficient for the unperturbed dynamics to lead to σ_A.

Example 17 (Nearest-neighbor interaction on the circle) *In this case, mutation robustness and fast convergence both hold. For each* N*, let* Γ_N *be the network corresponding to nearest-neighbor interaction on a circle of size* N*, as in Example 9. Since* σ_A *can be reached through the unperturbed dynamics whenever a single agent chooses action* A*, fast convergence holds with* $T = 1$*. From Example 9, the threshold* $\overline{\alpha}$ *tends to zero as* N *grows large.*

The preceding example captures the intuition that mutation robustness and fast convergence coincide if σ_A can be reached from any initial state by a small number of muta-

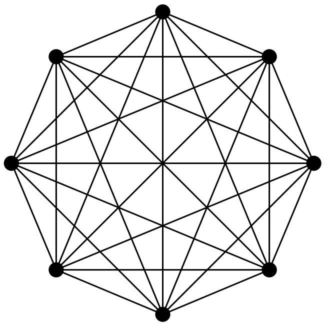

Figure 3: Uniform interaction.

tions. This coincidence, however, does not extend more generally, as the following examples demonstrate.

Example 18 (Uniform interaction) *In this case, fast convergence fails for any* p*, but the threshold* $\overline{\alpha}$ *is small when* p *is small. Let* Γ_N *be the complete network on* N *nodes; that is, every player interacts with every other player (see Figure 3). Ellison (1993) shows that fast convergence fails in this network. The mutation robustness threshold, however, is approximately constant as the network grows large. Since there are only two equilibria of the population game, the radius and coradius of* σ_A *may be used to compute the precise value of* $\overline{\alpha}$*. Accordingly, we have* $C(\sigma_A) = \lceil p(N-1) \rceil$ *and* $R(\sigma_A) = \lceil (1-p)(N-1) \rceil$*, and hence*

$$\overline{\alpha} = \frac{\lceil p(N-1) \rceil}{\lceil (1-p)(N-1) \rceil}.$$

In particular, the threshold $\overline{\alpha}$ *is small for large* N *when the payoff parameter* p *is small.*

Example 19 (Regions of size m**)** *In this case, for some values of* p*, fast convergence holds but mutation robustness fails. Consider an infinite network* Γ *in which the nodes correspond to elements of* $\mathbb{Z} \times \{1, \ldots, m\}$*. Each node* (i, j) *interacts with all* $m-1$ *other nodes having coordinates* $(i, \cdot)$*, as well as to the two nodes* $(i-1, j)$ *and* $(i+1, j)$ *(see Figure 4). The sets* $\{\{i, 1\}, \ldots, \{i, m\}\}$ *are the regions of the network, within which interaction is uniform, and between which links are relatively rare. For each* m*, contagion occurs*

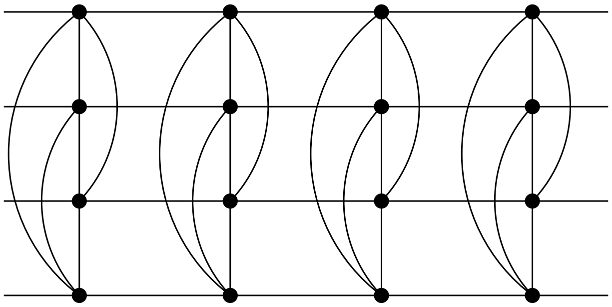

Figure 4: Regions of size m for $m = 4$ (Morris, 2000).

if and only if $p < \frac{1}{m+1}$. *If* $p > \frac{1}{m+1}$, *then any two adjacent regions together form a* $(1-p)$*-cohesive set, so by Theorem 12, mutation robustness holds in* Γ *if and only if* $p < \frac{1}{m+1}$. *On the other hand, fast convergence holds for all* $p < \frac{1}{2}$.[1] *Intuitively, when* $p > \frac{1}{m+1}$, *convergence is fast because transitions to* σ_A *can occur in many small steps through a sequence of coexistent conventions. This rich structure of conventions, however, also ensures that mutation robustness fails because, no matter how large the network, some coexistent convention may be reached from* σ_A *by only a fixed number of mutations.*

8 Discussion and conclusion

The structure of social networks has been widely studied in the sociology literature, and a number of regularities have been empirically observed in a variety of settings (see Newman (2003) for a survey). We may consider, then, how these properties relate to the structural conditions described above that are relevant for mutation robustness in order to assess the relevance of the risk dominance prediction in real-world networks when mutation probabilities are unmodelled. This discussion must, however, necessarily remain vague since definitions of observed network properties vary, and quantification is difficult in general.

---

[1] The proof of this result relies on a strengthening, also due to Ellison, of the radius-coradius method, in which the coradius is replaced by the (smaller) *modified coradius* (Ellison 2000). One can show that, for an appropriate labelling of Γ, the modified coradius of σ_A in $\Gamma_\iota(N)$ is at most $2m$ regardless of N. Theorem 2 of Ellison (2000) then implies that fast convergence holds for $T = 2m$. The details of the modified coradius calculation are somewhat involved, and are omitted here.

- *Small-world networks.* A number of real-world networks have been found to have significant local clustering, but at the same time a small global diameter relative to certain highly structured networks; that is, the distance between any two nodes is "small" given the size of the population. Such networks are said to possess the *small-world property.* Clustering alone, if sufficiently dense, can correspond to the existence of small highly cohesive groups of nodes, thereby placing a lower bound on the threshold parameter $\overline{\alpha}$ that depends on the size of the smallest such group. In large populations, the existence of such a group may preclude mutation robustness. Furthermore, Morris (2000) provides sufficient conditions for contagion that include "low neighbor growth," which is inconsistent with a small diameter. Thus insofar as contagion may contribute to mutation robustness, the small diameter property of small-world networks also appears to be contrary to the structural properties required for a robust risk dominance prediction, although one must be cautious here since this is based only on sufficient conditions for mutation robustness.

- *Community structure.* The nodes in networks sometimes form identifiable groups or communities in such a way that the density of links is much higher within groups than between them. As in the regions example of the preceding section, if these communities are sufficiently strong in the sense that a sufficiently large proportion of links in the network are within groups, then highly cohesive sets of nodes will exist, some of which will be small relative to the size of the population if many communities exist in the network. Thus community structure may also prevent mutation robustness from occurring.

- *Scale-free networks.* The distribution of degrees of nodes in a purely random network, in which there is a fixed independent probability that a link exists between any two nodes of the network, is binomial, approaching a Poisson distribution as the population grows large (Newman 2003). In real-world networks, the degree distribution typically features a heavier tail than that for random networks, corresponding to a greater number of high-degree nodes. Recall that the upper bound on the threshold value $\overline{\alpha}$ given in Theorem 7 is strongest when all nodes have the same degree, and becomes weaker as the distribution becomes more dispersed. The presence of high degree nodes can reduce the radius of coordination on the risk dominant action, lessening the bias in mutations necessary to overturn the stability of this equilibrium.

- *Tie strength.* The model employed here assumes for simplicity that all links are given equal weight in each player's payoffs. In general, however, these weights may differ, for example because of non-uniform probabilities of matching. The analysis extends naturally to this more general setting, with the role of p-cohesiveness replaced by *weighted* p-cohesiveness: action A is a best response for player i if the share of weights associated with those of i's neighbors who play action A is at least p. Strong links, corresponding to those which are assigned higher payoff weights, tend to exhibit greater clustering than weak links (Granovetter 1973). Weighting links will therefore increase the likelihood that a small highly (weighted) cohesive set will exist in the network, which again limits the size of the bias necessary to overturn the risk dominance prediction.

To summarize, for each of the network properties that have been most prominent in the empirical literature, none contributes to mutation robustness. This suggests the need to be careful when modelling mutation rates in local interaction environments, as large biases in mutation probabilities may not be necessary to alter the set of stochastically stable outcomes.

References

Bergin, J. and B. L. Lipman (1996). Evolution with state-dependent mutations. *Econometrica 64*, 943–956.

Blume, L. E. (2003). How noise matters. *Games and Economic Behavior 44*, 251–271.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*, 1047–1071.

Ellison, G. (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *Review of Economic Studies 67*, 17–45.

Foster, D. P. and H. P. Young (1990). Stochastic evolutionary game dynamics. *Theoretical Population Biology 38*, 219–232.

Goyal, S. and F. Vega-Redondo (2005). Network formation and social coordination. *Games and Economic Behavior 50*, 178–207.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology 78*, 1360–1380.

Hojman, D. and A. Szeidl (2006). Endogenous networks, social games and evolution. *Games and Economic Behavior 55*, 112–130.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica 61*, 29–56.

Lee, I. H., A. Szeidl, and A. Valentinyi (2003). Contagion and state dependent mutations. *Advances in Theoretical Economics 3*(1).

Morris, S. (2000). Contagion. *Review of Economic Studies 67*, 57–78.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review 45*, 167–256.

Pęski, M. (2004). State-independent evolution on networks. Working Paper available at http://home.uchicago.edu/~mpeski/si.pdf.

Young, H. P. (1998). *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press.