STOCHASTIC STABILITY AND DISAGREEMENTS BETWEEN DYNAMICS

AYDIN MOHSENI

Abstract. The replicator dynamics and Moran process are the main deterministic and stochastic models of evolutionary game theory. These models are connected by a mean-field relationship—the former describes the expected behavior of the latter. However, there are conditions under which their predictions diverge. I demonstrate that the divergence between their predictions is a function of standard techniques used in their analysis, and of differences in the idealizations involved in each. My analysis reveals problems for stochastic stability analysis in a broad class of games. I also demonstrate a novel domain of agreement between the dynamics, and draw a broader methodological moral for evolutionary modeling.

1. Introduction

The replicator dynamics (Taylor and Jonker 1978) and frequency-dependent Moran process (Moran 1962) are the main deterministic and stochastic dynamics of evolutionary game theory (Cressman and Tao 2014; García and Traulsen 2012). Both dynamics capture the basic idea that phenotypes that are more fit than the population average tend to grow in proportion, while phenotypes less fit than average tend to shrink in proportion. The replicator dynamics gives us a deterministic description of the behavior of evolution, assumes infinite populations, and isolates the influence of selection. The Moran process gives us a stochastic description of evolution, assumes finite populations, and introduces the effects of drift.

Importantly, the two dynamics are connected by a mean-field relationship (Benaïm and Weibull 2003, 2009). Intuitively, the replicator dynamics describes the expected behavior of the Moran process for large populations over finite stretches of time.[1]

Yet, there exists a striking puzzle: employing standard methods of analysis, the two dynamics can make contradictory predictions (Sandholm 2009, Ch.12). When the interactions within a population are modeled by an anti-coordination game, or game containing an anti-coordination subgame, the replicator dynamics may predict that selection will favor polymorphism,[2] but it is said that the Moran process shows that

Date: September 30, 2018.

[1]The replicator dynamics also provides a mean field for other dynamics, such as reinforcement learning (Benaïm and Weibull 2003), and emerges from distinct revision protocols, including pairwise proportional imitation, imitation driven by dissatisfaction, and imitation of success (Sandholm 2009, Ch.5.4).

[2]*Polymorphisms*, here, are population states in which multiple phenotypes are present. They are contrasted with *monomorphisms*, in which only a single type is present.

	A	B
A	1	2
B	2	1

TABLE 1. A 2×2 symmetric anti-coordination game.

such polymorphisms cannot persist in the long run (Taylor et al. 2004; Novak 2007). I examine this puzzle, and show that its standard explanation is not quite right. I demonstrate that, even in the long run, there are a range of conditions under which the Moran process sustains polymorphism. Under conditions I characterize the long run behavior of the Moran process will realign with the predictions of the replicator dynamics.

The misunderstanding of the behavior of the Moran process stems from a shortcoming in a standard technique of analysis: stochastic stability analysis. And the shortcoming of stochastic stability results from its assumption of vanishing mutation rates. My results indicate problems for stochastic stability in a broad class of games, reveal a novel domain of agreement between the two dynamics, and suggest a methodological moral for evolutionary modeling.

To understand our motivating puzzle, we can consider the simple anti-coordination game given in TABLE 1, and examine the predictions as to the evolutionary outcomes of its corresponding population game under each dynamics. For both dynamics let us assume: large populations, random pair-wise interactions, true breeding, the absence of mutation, and infinite-horizon play. Under the replicator dynamics, the prediction is that, from most all initial conditions, evolution will deliver the population to the *polymorphic state* $x = 1/2$,[3] where A-types and B-types coexist in equal proportions. In contrast, for the same anti-coordination game, the Moran process predicts that evolution will deliver the population, with equal probability, to one of the two *monomorphic states* $x = 0$ or $x = 1$,[4] where the population is composed entirely of either A-types or B-types. The evolutionary outcome that is a moral certainty in one model is an impossibility in the other.

Such a divergence in the predictions of the two dynamics leads naturally to the following questions: How are we deriving the predictions of each dynamics? And what is the cause of their divergence?[5]

The standard explanation for divergence in such cases is that the dynamics differ in the time-horizons of their predictions: the replicator dynamics approximates the

[3]For the replicator dynamics, these will be *asymptotically stable* states. These will be explained in §2.1.

[4]For the Moran process, these will be either *absorbing* states or *stochastically stable* states, depending on the presence of mutation. These will be explained in §2.2 and §2.3.

[5]Another important—and open—question is: how, generally, should we meaningfully compare the predictions of stochastic and deterministic dynamics?

short-to-medium run behavior of evolution, while the the Moran process can capture its long run behavior (Taylor et al. 2004; Novak 2007). The prediction of the replicator dynamics is polymorphism, and this correct for the short-to-medium run. The prediction of the Moran process is monorphism, and this is correct for the long run. Young (1998, 47) states this clearly: "While [the replicator dynamics] may be a reasonable approximation of the short run (or even medium run) behavior of the process, however, it may be a very poor indicator of the long run behavior of the process."

The dynamics differ with respect to the time-horizons of their predictions. This is true, but in the case of interest, this is not the cause of the divergence in their predictions, and it is not the answer to our puzzle. The cause of the divergence lies in the standard technique employed to derive predictions from the Moran process, *stochastic stability analysis*, introduced to game theory by Foster & Young (1993). Under conditions I will characterize, stochastic stability leads to the mis-prediction of homogeneity where long run diversity is to be expected.

Why does this matter? In brief, because the technique of stochastic stability analysis is ubiquitous. Among those having deployed stochastic stability in the analysis of the Moran process, and related processes,[6] include: Binmore & Samuelson (1995; 1997), Fudenberg & Imhof (2004; 2006), Fudenberg et al (2006), Imhof et al (2006), Nowak et al (2004; 2007), Ohtsuki et al (2007), Sandholm (2007; 2009; 2010; 2012), Taylor et al (2004), Trauelsen & Hauert (2010), and Young (1993; 1998; 2005; 2015). Evolutionary game theorists use stochastic stability analysis to explore and explain various phenomena in the domains of cultural and biological evolution, ranging from the diffusion of innovations to the emergence of conventions. Stochastic stability is a standard tool in both theoretical and applied work. Given this, understanding its limitations is important.

The structure of this paper is as follows. In §2, I will introduce the replicator dynamics and Moran process models along with the concepts of asymptotic stability, replacement probabilities, and stochastic stability needed to understand our results. In §3, I will demonstrate the conditions under which stochastic stability will mis-predict the long run behavior of the Moran process, polymorphisms will persist, and the behavior of the replicator dynamics and Moran process will realign. In §4, I will discuss real-world applications where one can expect my results will matter, and suggest a methodological moral for evolutionary modeling. In §6, I conclude.

2. The Dynamics

2.1. **The Replicator Dynamics.** The replicator dynamics is the "first and most important model of evolutionary game theory" (Cressman and Tao 2014, 1081). This

[6]These include the closely related Markov processes of Fermi and Wright-Fisher.

	A	B
A	a	b
B	c	d

TABLE 2. A 2×2 symmetric game

is due to the fact that it allows us to isolate the qualitative influence of selection on evolution, unperturbed by the complicating factors of mutation, drift, recombination, and so on. The leading idea behind the replicator dynamics is that types that are more fit than the population average fitness grow in relative proportion, and types that are less fit than average shrink in proportion. This can be described by a system of differential equations[7]

$$\dot{x}_i = x_i[u(i,x) - u(x,x)] \quad \text{for } i \in S$$

where S is the set of possible types, $\dot{x}_i$ denotes the rate of change of the population proportion of type i, x_i denotes the population proportion of type i, $u(i,x)$ denotes the expected fitness for type i from interacting with the population, and $u(x,x)$ denotes the population average fitness.

We derive predictions from the replicator dynamics by finding the *asymptotically stable states* of the dynamics for a given game, and equating these with the plausible outcomes of evolution for that game.[8] A population state is asymptotically stable just in case it is both *stable* and *attracting*. Intuitively, a state is stable if states near it remain near it, and attracting if states near it tend toward it. This gives us our prediction of the behavior of a process described by the replicator dynamics.

For the simple class of 2×2 symmetric games under the replicator dynamics, five qualitatively distinct outcomes are possible. These can bee seen in TABLE 3. Our puzzle concerns the class of anti-coordination games, shown in the third row of the table, and labeled the 'polymorphic case'. This is where we find polymorphisms that are asymptotically stable under the replicator dynamics. Anti-coordination games constitute and important class of interaction structures, and have been used in

[7]The Replicator dynamics can also be formulated for discrete time—the Maynard-Smith formulation (1982)—by a system of difference equations, which, under some conditions, yield subtly different results from their continuous time counterpart (Cressman 2003). However, for 2×2 games, the qualitative predictions of the two formulations coincide. Thus, here, without loss of generality, I will work with the continuous time formulation exclusively.

[8]Asymptotic stability does not exhaust the plausible outcomes of the replicator dynamics. In more complex games, disequilibrium behavior such as cycles and strange attractors, along with sets of collectively but not individually stable states, will not be asymptotically stable but may still constitute plausible outcomes of the dynamics. For a survey and analysis of this issue, see (Mohseni 2017). However, for the class of 2×2 symmetric games considered here, there is a one-to-one correspondence between evolutionarily significant outcomes and asymptotically stable states. So, we can proceed comfortably with asymptotic stability as our stability concept for the replicator dynamics.

Game Type	Payoffs	Phase Portrait	Asymptotically Stable States
A dominates B	$a > c$ $b > d$	A ●←○ B	All-A state
Bi-stable case	$a > c$ $b < d$	A ●←○→● B	All-A and All-B states, with basins of attraction divided at $\frac{d-b}{d+a-c-b}$ A-types
Polymorphic case	$a < c$ $b > d$	A ○→●←○ B	A mixed state, with $\frac{b-d}{b+c-a-d}$ A-types
B dominates A	$a < c$ $b < d$	A ○→● B	All-B state
Neutral case	$a = c$ $b = d$	A ····· B	None

TABLE 3. 2×2 symmetric games under the replicator dynamics. Opaque circles denote asymptotically stable states, empty circles denote unstable fixed points, dotted lines denote sets of unstable fixed points, and arrows indicate the direction of selection.

explanations of ritualized animal conflict (Maynard Smith 1974), sex ratios (Hamilton 1967), and bargaining norms (Skyrms 1996).

We derive the prediction of the replicator dynamics for anti-coordination games (TABLE 3, third row) as follows. We solve for the fixed points of the dynamics, where the rate of change in population proportions of each type is zero, i.e., $\dot{x}_1 = \dot{x}_2 = 0$. This yields three states: the two monomorphic states composed entirely of one type or the other, and a polymorphic state, i.e., $\{0, \frac{b-d}{b+c-a-d}, 1\}$. We assess the stability of these states by examination of the eigenvalues of Jacobian matrix for the dynamics, which reveals that only the mixed state is asymptotically stable. Given this, we know that a population starting at the polymorphism with proportion $\frac{b-d}{b+c-a-d}$ A-types will remain there, and that, from most all initial conditions,[9] the dynamics will converge to the polymorphism.

2.2. **The Frequency-Dependent Moran Process.** The Moran process is a birth-death process in which, for each time step, two individuals are chosen: one for reproduction and the other for elimination. The individual chosen for birth is determined, probabilistically, by the relative fitness of the types within the population, and the individual chosen for death is selected at random. So, if we consider a population of N individuals whose payoff from interaction are described by TABLE 2, then the fitnesses f_i, g_i of the types A, B can be described as functions of the number i of A-types,

$$f_i = 1 - w + w\frac{a(i-1) + b(N-i)}{N-1} \quad \text{and} \quad g_i = 1 - w + w\frac{ci + d(N-i-1)}{N-1},$$

[9]Initial conditions in which some proportion of each type is present in the population.

where w denotes the intensity of selection, or the game's contribution to the net fitness of the organism. Observe that $w = 1$ implies that an individual's fitness is entirely determined by her interactions in this game, and $w = 0$ implies that the game makes no contribution to her fitness.

Individuals reproduce at a rate proportional to their fitness. The rate of reproduction then for A -types is if_i and for B-types is $(N - i)g_i$. Each period, one offspring is chosen at random to enter the population. So, the probability of adding an A-type offspring is $\frac{if_i}{if_i+(N-i)g_i}$, and the probability of adding a B-type offspring is $\frac{(N-i)g_i}{if_i+(N-i)g_i}$. After reproduction, one individual is chosen at random to be removed from the population, so that with probability $\frac{i}{N}$ an A-type is removed, and with probability $\frac{N-i}{N}$ a B-type is removed. This makes it so that the population size remains constant.

Formally, we define the Moran process with population size N as a Markov process $\{X_t^N\}$ over the finite state space $\chi = \{1, \ldots, N\}$ of possible population states, with transition probabilities between states given by

$$P_{i,j} = \begin{cases} \dfrac{N-i}{N}\dfrac{if_i}{if_i + (N-i)g_i}, & \text{if } j = i+1 \\ \dfrac{i}{N}\dfrac{(N-i)g_i}{if_i + (N-i)g_i}, & \text{if } j = i-1 \\ 1 - P_{i,i+1} - P_{i,i-1}, & \text{if } j = i \\ 0, & \text{otherwise,} \end{cases}$$

which composes a tri-diagonal matrix. Note that $P_{0,0} = P_{N,N} = 1$, so that the process has two absorbing states, $i = 0$ and $i = N$, and that all other states are transient. An absorbing state is a state that, once visited by the process, is never escaped. A transient state then is one which will only be visited a finite number of times before the process arrives at some absorbing state. Note that, in the limit of time, with probability one, the process will reach one or the other absorbing state.

For the simple class of 2×2 symmetric games under the Moran process, as with the replicator dynamics, we can examine five distinct cases (Table 4) when the population size is large.[10] For each case, the outcomes are described in terms of the relative probability of arrival of the process at each of the absorbing states. In particular, we compare the probability of a single mutant coming to replace the incumbent type, and take over the population. This yields the *replacement probabilities*[11] ρ_{AB} and ρ_{BA}, where ρ_{AB} denotes the probability of a single A-type individual leading to the takeover of an otherwise B-type population, and ρ_{BA} denotes the probabilities of the inverse process. The replacement probabilities of types are

[10]We take the large population limit for the Moran process to allow for meaningful comparison with the replicator dynamics. For analysis of the changes in the behavior of the Moran process as a function of population size see (Taylor et al. 2004).

[11]For an exposition of the details of this approach, see (Nowak et al. 2004).

Game Type	Payoffs	Replacement Probabilities	Description
A dominates B	$a > c$ $b > d$	$\rho_{BA} < \frac{1}{N} < \rho_{AB}$	Selection opposes B and favors A.
Bi-stable case	$a > c$ $b < d$	Not $(\frac{1}{N} < \rho_{BA}, \rho_{AB})$	Selection may favor A or B, but not both.
Polymorphic case	$a < c$ $b > d$	Not $(\rho_{BA}, \rho_{AB} < \frac{1}{N})$	Selection may oppose A or B, but not both.
B dominates A	$a < c$ $b < d$	$\rho_{AB} < \frac{1}{N} < \rho_{BA}$	Selection opposes A and favors B.
Neutral case	$a = c$ $b = d$	Not $(\frac{1}{N} < \rho_{BA}, \rho_{AB})$, or $\rho_{BA} = \rho_{AB} = \frac{1}{N}$	Selection favors A or B depending on the sign of $(a+b)-(c+d)$, or is neutral if $a+c=b+d$.

TABLE 4. 2×2 symmetric games under the Moran process with large populations.

compared to those of a neutral mutant (where $a = b = c = d$), which will come to fixation with probability $1/N$. We say that *selection favors* a type if its replacement probability is greater than that of a neutral mutant, and that *selection opposes* a type if its replacements probability is less than that of a neutral mutant.

We derive the predictions of the Moran process for anti-coordination games (TABLE 4, third row) by calculating th replacement probabilities for each type. This yields three possibilities: $\rho_{BA} < \frac{1}{N} < \rho_{AB}$, $\rho_{AB} < \frac{1}{N} < \rho_{BA}$, or $\frac{1}{N} < \rho_{AB}, \rho_{AB}$. That is, either selection favors one type replacing the other, or it favors both replacing one another.[12] What we see is that, for anti-coordination games, selection must favor at least one type in coming to dominate the population. In the absence of mutation, polymorphism is temporary, and evolution inevitably attains homogeneity.

2.3. **The Frequency-Dependent Moran Process with Mutation.** With the introduction of mutation the behavior of the Moran process changes qualitatively. Absorbing states disappear, and there is positive probability that the process will transit within finite time from any given state to any other. Thus, in the limit of time, the process visits each state infinitely often. Since absorption will not occur, replacement probabilities are no longer appropriate, and a different method of analyzing the behavior of the process is needed. This method is to find the long run distribution of time spent by the process over the possible population states.

Formally, we define the Moran process with population size N and mutation rate η as an ergodic process[13] $\{X_t^{N,\eta}\}$ over the finite state space $\chi = \{1, \ldots, N\}$, with

[12] In such cases, one can examine the sign of $\rho_{AB} - \rho_{BA}$ to determine which type is more or less favored by selection.

[13] An ergodic process is a Markov process that is both irreducible (every state is reachable from any other), and aperiodic (the greatest common divisor for the number of steps to return to each state is one). It is easily verified that the Moran process with mutation is indeed ergodic.

transition probabilities given by

$$\hat{P}_{i,j} = \begin{cases} (1-\eta)\dfrac{N-i}{N}\dfrac{if_i}{if_i+(N-i)g_i} + \eta\dfrac{N-i}{N}\dfrac{(N-i)g_i}{if_i+(N-i)g_i}, & \text{if } j = i+1 \\ (1-\eta)\dfrac{i}{N}\dfrac{(N-i)g_i}{if_i+(N-i)g_i} + \eta\dfrac{i}{N}\dfrac{if_i}{if_i+(N-i)g_i}, & \text{if } j = i-1 \\ 1-\hat{P}_{i,i+1}-\hat{P}_{i,i-1}, & \text{if } j = i \\ 0, & \text{otherwise} \end{cases}$$

where $\hat{P}_{0,0} = \hat{P}_{N,N} = 1-\eta$. Note the mutation terms. What they capture is that, most of the time $(1-\eta)$, selection behaves as normal, but in a minority of instances η, an offspring that was to be an A-type will become a B-type, and vice versa.[14]

Now, to understand the long term behavior of the Moran process we can compute its stationary distribution which captures proportion of time spent at each population state. Formally, a probability distribution $\mu \in \mathbb{R}^{\chi}$ is a *stationary distribution* of the ergodic process $\{X_t^{N,\eta}\}$ if

$$\sum_{i\in\chi} \mu_i P_{i,j} = \mu, \qquad \text{for all } j \in \chi.$$

That is, a stationary distribution is a probability vector such that taking its product with the matrix of transition probabilities simply returns itself.

We know that such a distribution exists, since every ergodic process has a unique stationary distribution, and that it is history independent. That is, from any initial distribution over states, the distribution of time spent by the process over population states converges to that given by stationary distribution.

Typically, however, we do not derive predictions from the Moran process by finding its stationary distribution. This is due to the fact that general analytic forms of the stationary distribution for the Moran process for complex games are not known,[15] and because the stationary distribution applies positive probability to every state, as opposed to yielding unique predictions (Harper and Fryer 2016).

Instead, the drive for analytic tractability and unique equilibrium prediction motivates the use an alternative: stochastically stability analysis. Stochastically stable states are just those that retain mass in the stationary distribution when we take the

[14]Here, we have assumed that mutation is symmetric, but it need not be so. Asymmetric mutation can be accounted for by formulating the rate of mutation for one type as a ratio of the other $r\eta$, where r is a positive constant. For an analysis of the affects of asymmetric mutation rates see (Traulsen and Hauert 2010).

[15]The exceptions to this are for 2×2 games under arbitrary revision protocols, and for potential games under exponential revision protocols (Sandholm 2009).

Game Type	Payoffs	Stochastically Stable States	Description
A dominates B	$a > c$ $b > d$	$\mu_A \to 1$	The stationary distribution is a point-mass at the all-A state.
Bi-stable case	$a > c$ $b < d$	$\mu_A \to 1$ xor $\mu_B \to 1$	The stationary distribution is a point-mass at either the all-A or all-B state.
Polymorphic case	$a < c$ $b > d$	$\mu_A \to 1$ xor $\mu_B \to 1$	The stationary distribution is a point-mass at either the all-A or all-B state.
B dominates A	$a < c$ $b < d$	$\mu_B \to 1$	The stationary distribution is a point-mass at the all-B state.
Neutral case	$a = c$ $b = d$	$\mu_A \to 1$ xor $\mu_B \to 1$ xor $\mu_A, \mu_B \to \frac{1}{2}$	The stationary distribution is a point-mass at all-A or all-B depending on the sign of $(a+b)-(c+d)$, or evenly split between the states if $a+c=b+d$.

TABLE 5. 2×2 symmetric games under the Moran process with mutation and large populations.

limit as mutation approaches zero.[16] Formally, a state $i \in \chi$ is *stochastically stable* if

$$\lim_{\eta \to 0} \mu_i^{N,\eta} > 0.$$

We saw that, for the Moran process, in the absence of mutation, all and only monomorphic states were absorbing states. Now, with vanishing mutation, the stationary distribution collapses (typically) to a point-mass on just one of these absorbing states. As mutation vanishes, the behavior of the ergodic process approaches that of the absorbing chain, and so spends most of its time near one or another monomorphic state.

la We derive the predictions of the Moran process with mutation for anti-coordination games (TABLE 5, third row) by finding which states retain positive mass in stationary distribution as mutation vanishes.[17] This yields two possibilities: $\mu_A \to 1$, or $\mu_B \to 1$.[18] That is, either the all-A or all-B state, but not both, can be stochastically stable. Once again, polymorphism cannot be selected.

3. THE LONG RUN PERSISTENCE OF DIVERSITY

Why can't stochastically stable states be polymorphisms? This is by construction: a stochastically stable state of an ergodic process will be an absorbing state of its corresponding absorbing chain. Stochastic stability is defined for an ergodic process, and is determined by identifying the states that retain mass in the stationary distribution of the process as mutation vanishes. As mutation vanishes, the behavior

[16]Stochastic stability is often solved for using particular well-chosen graphs that capture the difficulty of transitioning from each absorbing state (of the original absorbing chain) to any other. For a presentation of the relevant techniques, see Ch. 3.2 of Young (1998). I present a more general formulation better suited to my project.

[17]See (Fudenberg and Imhof 2004) for the relevant technique.

[18]In the knife-edge case where $a = d$ and $b = c$ we get $\mu_A, \mu_b \to 1/2$.

of the ergodic process approaches that of the absorbing chain. Polymorphisms cannot be absorbing states, and thus cannot be stochastically stable.

In most game types, the qualitative predictions of asymptotic stability for the replicator dynamics and the predictions of stochastic stability for the Moran process are in basic agreement (see Tables 3, 4, 5). In the case of coordination games and dominating strategy games (rows 1, 2, 4), for large populations, the asymptotically stable state with the largest basin of attraction has the greatest replacement probability and is uniquely stochastically stable. Predictions differ in the in the polymorphic case (compare row 3 of Tables 3, 4, 5), and the neutral case (compare row 5 of Tables 3, 4, 5). The latter is to be expected, as the replicator dynamics explicitly abstracts away from the effects of drift. Disagreement about the polymorphic case is more puzzling. For an anti-coordination game, polymorphism is uniquely asymptotically stable under the replicator dynamics, but only monomorphic states can be stochastically stable under the Moran process.

The standard explanation we have seen accounts for this divergence in predictions by positing that the replicator dynamics fails to capture the long run behavior of the Moran process. The Moran process will, due to stochasticity, eventually arrive at an absorbing state of the process, where it will spend most of its time, trapped by low mutation rates. The following excerpt from Sandholm (2009, 208) tells this story:

> "The stochastic process typically moves in the direction indicated by the mean dynamic. If the process begins in the basin of attraction of a rest point or other attractor of this dynamic, then the initial period of evolution generally results in convergence to and lingering near this locally stable set. ... However, [since the process is irreducible] this cannot be the end of the story. Indeed, the process eventually reaches all states, and in fact visits all states infinitely often. This means that the process must leave the basin of the stable set visited first; it then enters the basin of a new stable set, at which point it is extremely likely to head directly to the set itself. The evolution of the process continues in this fashion, with long periods near each attractor punctuated by sudden jumps between them."

Young (1998, 20) calls this the "punctuated equilibrium effect," echoing that, "When the stochastic shocks are small, the mode of this frequency distribution [the stationary distribution] will tend to be close to the stochastically stable [states] predicted by the theory."

But this characterization turns out to be insufficient for the polymorphic case. How small must the mutation rate be? And what role do population size and intensity of selection play? In anti-coordination games under the Moran process, the population may indeed spend the majority of its time at or near a polymorphic equilbrium, even

	A	B
A	1	3
B	2	1

TABLE 6. A 2×2 anti-coordination game.

in the long run. This occurs when there is small but non-vanishing mutation, and sufficiently large population size and intensity of selection.

To illustrate the potential divergence of the actual and predicted behaviors of the Moran process, consider the anti-coordination game given by TABLE 6. We fix the following parameter settings: population size $N = 100$, mutation rate $\eta = 0.01$, and intensity of selection $w = 0.2$. Now, consider the predictions of each of our stability concepts: The replacement probabilities are $\rho_{AB} \approx 0.1644$ and $\rho_{BA} \approx 0$, so selection favors A and opposes B; absorption into the all-A state is the most probable outcome of the process. Stochastic stability analysis yields that $\mu_A = 1$, so all-A is the unique stochastically stable state; in the long run, the process will spend almost all of its time near the all-A state. Finally, asymptotic stability analysis yields the unique asymptotically stable state $x^* = 2/3$; from all mixed initial conditions the population will converge to a polymorphism where $2/3$ of the population are A-types and $1/3$ are B-types.

To see whether these predictions hold up, I analytically derive the actual long run behavior of the Moran process by calculating its stationary distribution,[19] without vanishing mutation, but rather with fixed parameter values of mutation rate.

This stationary distribution is plotted in FIGURE 1. Notice that the mode of the stationary distribution is at the state where there are 67 A-types in the 100-individual population. The process spends the most time precisely at the polymorphic state predicted by asymptotic stability under the replicator dynamics, and not at the all-A state that is stochastically stable.

To get a feel for the medium run behavior of the Moran process, we can simulate several dozen individual population trajectories, starting from random initial conditions, and evolving over a thousand birth-death events. This is plotted in FIGURE 2. Again, it is clear that asymptotic stability under the replicator dynamics gives us a more accurate prediction of the behavior of the Moran process.

[19]In the simple case of 2×2 games, we can obtain an explicit formula for the stationary distribution: $\mu_k = \mu_0 \prod_{i=1}^{k} \frac{\hat{P}_{i-1,i}}{\hat{P}_{i,i-1}}$ for $k \in \{1, \ldots, N\}$, and $\mu_0 = \left(\sum_{k=1}^{N} \prod_{i=1}^{k} \frac{\hat{P}_{i-1,i}}{\hat{P}_{i,i-1}}\right)$, where the empty product equals one. This can also be verified, computationally, using the Chapman-Kolmogorov equation, $P^t = (P)^t$, which says that the nth-step transition matrix for a Markov process is equal to the first-step transition matrix raised to the nth power. For very large t, this can be used to approximate the stationary distribution of a given Markov process (Karlin and Taylor 2012).

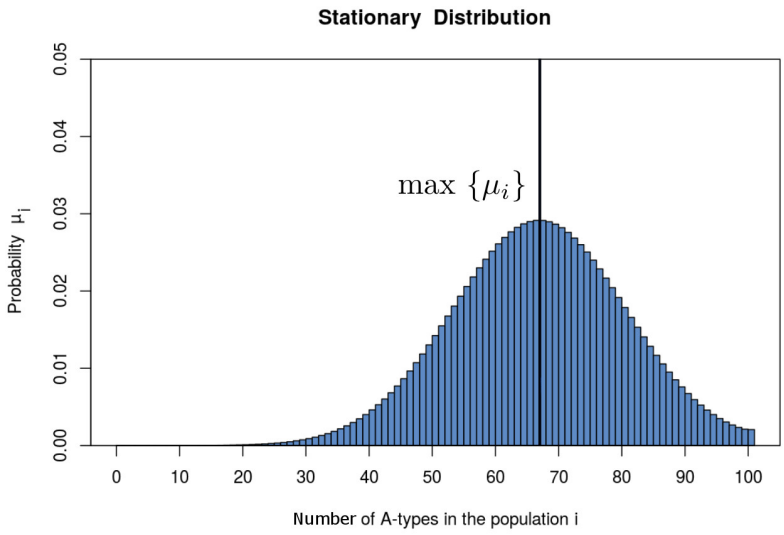

FIGURE 1. The stationary distribution for the Moran process with: population size $N = 100$, mutation rate $\eta = 0.01$, and intensity of selection $w = 0.2$. The vertical line marks the mode of the distribution.

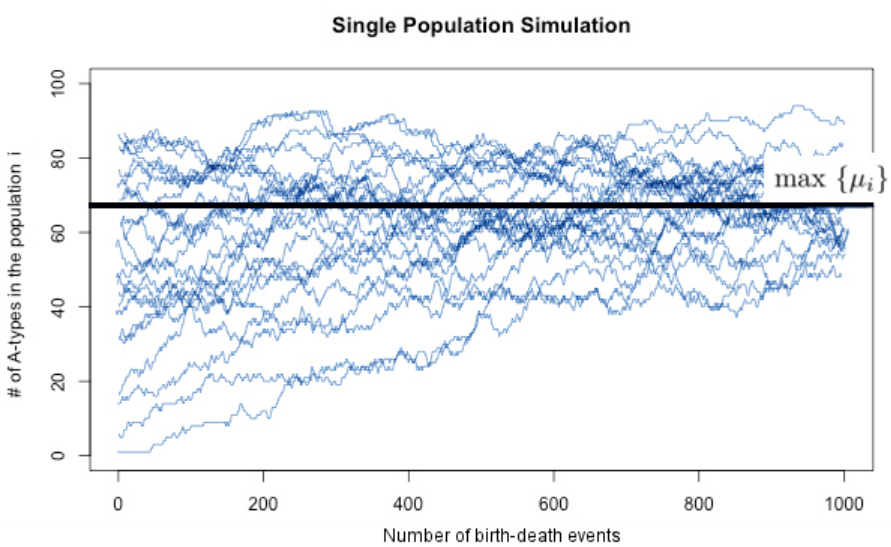

FIGURE 2. Plot of 35 population trajectories for the Moran process over 1000 birth-death events. The horizontal bar corresponds to the peak of the stationary distribution.

What we see is stochastic stability mis-predicting the actual behavior of the Moran process under particular conditions. But we want a more general characterization of when this will occur. To obtain this, we turn first to the case where there is no selection $w = 0$. Here, I use the detailed balance conditions of ergodic processes

(A) $\eta(N+2) > 1$ (B) $\eta(N+2) = 1$ (C) $\eta(N+2) < 1$

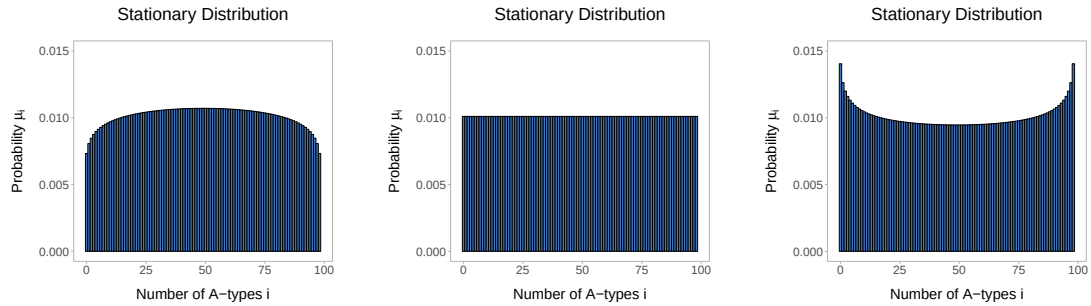

FIGURE 3. In the absence of selection pressures $w = 0$, satisfaction of the *strong mutation condition* $\eta(N+1) > 1$ determines the shape of the stationary distribution.

(Karlin and Taylor 2012) to deduce when the mass of the stationary distribution will be increasing toward the center of the state space. That is, the conditions under which the peak of the stationary distribution will be at a polymorphic state. This is captured by the following lemma.

Lemma 1. *For any 2×2 game under the Moran process, in the absence of selection $w = 0$, the strong mutation condition $\eta(N+2) > 1$ is necessary and sufficient for the mode of the stationary distribution to be a polymorphic state.*

What I am calling the 'strong mutation condition' corresponds to when the expected number of mutants entering a population in $N+2$ birth-death events is greater than 1. Intuitively, here strong mutation gives us when either the population is sufficiently large such that the process rarely arrives at monomorphic states, or the mutation rate is sufficiently high such that, when the population does arrive at monomorphic states, it does not spend too much time there.

Note that strong mutation is both necessary and sufficient for the stationary distribution to exhibit a polymorphic peak (Figure 3). That is, just when $\eta(N+2) > 1$, the stationary distribution is concave, and climbs gradually toward its peak near the middle point $\frac{N+1}{2}$ from either side of the state space (Figure 3a). When $\eta(N+2) = 1$, the stationary distribution is uniform (Figure 3b). When $\eta(N+2) < 1$, the stationary distribution is convex, and climbs outward from its nadir near the middle point toward its peaks at the monomorphic states 0 and N (Figure 3c).

We can use this insight as we turn to consider the case of nonzero intensity of selection $w > 0$. Consider the dynamics of an anti-coordination game characterized by the payoffs $a < c$ and $b > d$. It may be intuitive that, when the stationary distribution is already increasing in mass toward a polymorphic state, the addition of selection

pressure toward an interior equilibrium will continue to produce an interior mode. This is the essential insight from which I will derive my main result.

Before doing so, however, there are two stipulations that need to be made. First, I follow *Tayloret al.* (2004) in requiring that a coordination game, characterized by $a < c$ and $b > d$, further satisfy the condition that $N(b-d) > a-d$ for finite populations. This is because, for finite populations, the qualitative dynamics of a game can be affected by the anti-correlation produced by individuals not interacting with themselves. Anti-correlation can alter the qualitative dynamics of the game. Indeed, in sufficiently small populations, each of our four game types can, in principle, be transformed into a different game.

To see why this is so, consider the case with a population composed of two individuals $N = 2$. Here, the process has three possible population states: two A-types, two B-types, and one of each type. Since transition out of each monomorphic state occurs solely via mutation, only the state where there is one of each type involves selection. In this state, only the difference of the values in the off-diagonal of the payoff matrix, $b - c$, matters. If $b - c > 0$, then A dominates B. If $b - c < 0$, then B dominates A. The game is no longer, qualitatively, an anti-coordination game.

To correct for this, we require that payoffs further satisfy $N(b-d) > a-d$, ensuring that the game retains the qualitative dynamics of anti-coordination.[20] When N grows large, this condition is easily satisfied, and the qualitative dynamics are once again determined by the signs of the differences of the values of the column vectors, $a - c$ and $b - d$, just as with the replicator dynamics.

Second, I must also stipulate that mutation rates be reasonable: $\eta < 1/2$. It should be clear why this is so. If it is more probable that birth events are produced by mutation than by selection, then the fitnesses of the types will be reversed, and we will once again be playing a different game; a coordination game, in fact.

With these two stipulations in hand, we can turn to a sufficient condition for the persistence of diversity of types under the Moran process.

Theorem 1. *For any 2×2 symmetric anti-coordination game under the Moran process $a < c, b > d$, and $N(b-d) > a-d$, for any intensity of selection $w > 0$ and mutation $\eta < 1/2$, when the strong mutation condition $\eta(N+2) > 1$ is satisfied, the mode of the stationary distribution will be at a polymorphic state located between the critical point $i^* = \frac{N(b-d)+d-a}{b+c-a-d}$ and the midpoint of the state space $\frac{N+1}{2}$.*

What we have is that, when the strong mutation condition is satisfied, the peak of the stationary distribution is guaranteed to be between the midpoint of the state space and a critical point $i^* = \frac{N(b-d)+d-a}{b+c-a-d}$ that rapidly approaches the asymptotically

[20]See (Taylor et al. 2004) for a characterization of qualitative dynamics of finite games.

stable state of the same game under the replicator dynamics $x^* = \frac{b-d}{b+c-a-d}$ as N grows large.[21]

What remains is for us to confirm that, as intensity of selection increases, the peak of the stationary distribution will move toward the critical point. We can answer this in the affirmative.

Corollary 1. *For any 2×2 symmetric anti-coordination game under the Moran process $a < c, b > d$, and $N(b-d) > a-d$, for any intensity of selection $w > 0$ and mutation $\eta < 1/2$, when the strong mutation condition $\eta(N+2) > 1$ is satisfied, the mass of the states nearest the critical point $i^* = \frac{N(b-d)+d-a}{b+c-a-d}$ monotonically increase with intensity of selection.*

This is good. I note, however, that strong mutation provides *sufficient*, and *not* necessary, conditions for a polymorphic mode of the stationary distribution. An anti-coordination games can fail to satisfy the strong mutation condition, but have it so that a peak of its stationary distribution is at a polymorphic state. Strong mutation ensures that the stationary distribution increases monotonically toward some interior equilibrium, and thus ensures that there are no other peaks at the monomorphic states. For anti-coordination games where $\eta(N+2)$ is slightly less than one, but where selection pressure is great $a \ll c$ or $b \gg d$, the highest peak of the stationary distribution may still be a polymorphic state, with other smaller peaks at the monomorphic states.

In sum, when strong mutation obtains for an anti-coordination game, we know—with certainty—that stochastic stability analysis will mis-predict a monomorphic outcome when polymorphism is to be expected. But, when strong mutation does not obtain, there is still the possibility of mis-prediction.

To get an idea of the conditions under which games will exhibit polymorphic modes near the replicator dynamics prediction, we can examine the peak of the stationary distribution of a representative anti-coordination game (Table 6) for different values of N, η, and w.

In Figure 4, the darkness of each point in a plot encodes the distance, in terms of population proportions, between the peak of the stationary distribution, and the replicator dynamics prediction. In the plots, population sizes $N \in \{2, 3, \ldots, 100\}$ vary along the x-axes, mutation rates $\eta \in \{0, 0.01, \ldots, 0.5\}$ vary along the y-axes, and intensities of selection $w \in \{10^{-2}, 10^{-1}, 1\}$ vary between plots.

This accords with what we have learned so far, and illustrates our results. Where strong mutation obtains (in the space above the black curves), the peak of the stationary distribution is near the replicator dynamics prediction x^*. When the intensity of selection is low $w = 10^{-2}$, the demarcation is quite precise. As intensity of

[21]This is clear when we state the critical point in terms of a population proportion $\frac{i^*}{N} = \frac{b-d+\frac{(d-a)}{N}}{b+c-a-d}$.

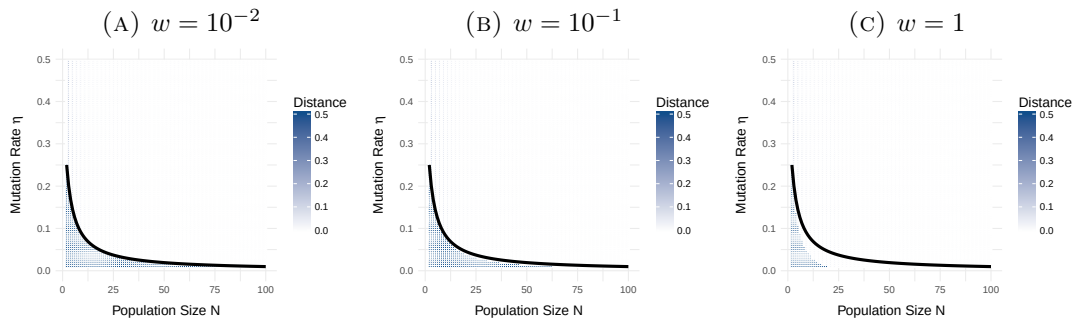

FIGURE 4. Distances of the mode of the stationary distribution from the replicator dynamics prediction for different values of N, η, w. The curve in black corresponds to $\eta(N+2) = 1$, above which the strong mutation condition is satisfied.

selection increases $w = 10^{-1}$, a growing range of population sizes and mutation rates (just beneath the black curves) will be compatible with an interior mode near x^*. When intensity of selection is at its maximum $w = 1$, strong mutation will continue to provide a sufficient, but not necessary, condition for polymorphism.

4. Discussion

I have characterized conditions under which we can anticipate the behavior of the Moran process will be mis-characterized by stochastic stability, and where long term diversity will persist. But should we expect these conditions to obtain in nature? If, indeed, the strong mutation condition were never to be satisfied, then we might comfortably rely on stochastic stability analysis without fear of it leading us astray. To see if this is so, we can survey representative population sizes, mutation rates, and intensities of selection from relevant real-world evolving populations.

Considering the canonical case of *E. coli* bacteria, we have that the per-site mutation rates are typically of the order of 10^{-4} mutations per allele per replication (Tenaillon et al. 2016). That is, an expected 1 out of $5,000$ bacteria produced carry at least one mutation at a locus of interest. Typical bacterial populations, however, are of the order of $10^6 - 10^8$. This yields a mutation strength of $\eta(N+2) \approx 20,000$. That is, there will be an average of twenty thousand mutations per population per generation. This is deep into the territory of strong mutation. Moreover, the population sizes and mutations rates of many bacteria are comparable (Drake et al. 1998). Bacterial populations, it seems, will exhibit population sizes and mutation rates that suggest their long run evolutionary behavior will typically be at odds with the predictions of stochastic stability.

Humans, on other hand, exhibit per-allele mutation rates ranging from 10^{-5} to 10^{-10} (Drake et al. 1998). Throughout much of our evolutionary history, *H. sapiens*

subsisted in hunter-gatherer groups averaging 50 to 150 individuals (Bowles and Gintis 2011). Hence, human biological evolution will typically not satisfy strong mutation. The same will be true of many complex organisms, such as mammals (Kumar and Subramanian 2002).

However, in the case of cultural evolution, we expect mutation rates—or noise in the transmission of behavior via social learning and imitation—to be potentially much higher (Boyd and Richerson 1985). Taking the example of humans, for a group of 100 individuals, an innovation or error rate in behavior transmission of just over 10% would satisfy strong mutation. Given that the Moran process is often used to model processes of cultural evolution, it will be important to know when an evolutionary process satisfies the strong mutation condition.

We should note that the relationship between strong mutation and persistent polymorphism is not guaranteed to hold in other game structures. We can expect the analysis will vary for extensive-form games, with more players and strategies, and with the introduction of social or spatial structure, an so on. But it is reasonable to imagine that qualitatively similar conditions may hold for other classes of games. The question as to the limits of the agreement of the dynamics for the case of strong mutation provides an interesting topic for further study.

5. Conclusion

The puzzle of the divergence between the predictions of the replicator dynamics and the Moran process finds its resolution in identifying a shortcoming of stochastic stability analysis. The cause of mis-prediction by stochastic stability is the assumption of vanishing mutation. Polymorphism, which cannot be stochastically stable, can be the most probable long run outcome of the Moran process.

I have shown that, under a range of values of population size, mutation rate, and intensity of selection, the Moran process leads to polymorphisms which dominate the long run behavior of the process. My results show that anti-coordination games, and games containing anti-coordination subgames, can exhibit this behavior for a broad range of conditions. For the 2×2 anti-coordination games considered here, 'strong mutation' provides a sufficient condition for mis-prediction by stochastic stability analysis of the long run behavior of the Moran process. Moreover, in the presence of strong mutation, the Moran process will typically spend most of its time near the specific polymorphic state that is asymptotically stable under replicator dynamics.

We have also seen that strong mutation will be satisfied by a range of real-world evolutionary processes. This is particularly true when population sizes are large, such as with bacterial colonies, and when mutation or noise rates are high, as is typical in the transmission of behavior in models of cultural evolution. In such cases, we can

anticipate that the behavior of the Moran process will be mis-characterized by stochastic stability, and will realign with the predictions of the replicator dynamics.

The upshots of our analysis are that we can characterize the conditions under which an evolutionary process described by the Moran process (1) will sustain long run diversity, (2) realign with the predictions of the replicator dynamics, and (3) should not be analyzed using stochastic stability. Our moral is that, when we anticipate attracting polymorphic equilibria—that is, when a population interaction structure is characterized by anti-coordination—stochastic stability may be an unreliable predictor of even the long term behavior of evolution. In such cases, analysis should proceed by computing the stationary distribution explicitly using representative values of population size, mutation rate, and intensity of selection. When such an approach is not feasible, simulation methods may suffice. In normal science, we can keep our best analytic tools, shortcomings and all, but we must also learn when not to use them.

Mathematical Appendix

For the following proofs, we consider a game under the Moran process with population size $N \in \mathbb{N}$, mutation rate $\eta \in (0, 1/2)$, and intensity of selection $w \in [0, 1]$, characterized by any 2×2 payoff matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ where $a, b, c, d > 0$,[22] payoff functions f_i and g_i, and transition matrix $P_{i,j}$ over the finite state space $\chi = \{0, 1, \ldots, N\}$. This yields the ergodic process $\{X_t^{N,\eta,w}\}$.

Let the fitness of each type at a particular intensity of selection be denoted $f^w \equiv f|_w$, $g^w \equiv g|_w$. Similarly, for transition probabilities, $P^w \equiv P|_w$, and stationary distributions $\mu^w \equiv \mu|_w$. We will omit the state subscript i, when there is no risk of confusion.

Proof of Lemma 1. We want to show that, in the absence of selection, the peak of the stationary distribution is a polymorphic state if, and only if, strong mutation $\eta(N + 2) > 1$ holds.

Since $\{X_t^{N,\eta,w}\}$ is an ergodic process when $\eta > 0$, we are guaranteed that it has a unique stationary distribution $\mu = \langle \mu_0, \ldots, \mu_N \rangle$. Further, we know that μ satisfies the detailed balance condition $\mu_i P_{i,i-1} = \mu_{i-1} P_{i-1,i}$. From this, it follows that $\mu_i > \mu_{i-1}$ just in case $P_{i,i-1} < P_{i-1,i}$. That is, a state i has greater mass in the stationary distribution than its preceding state $i - 1$ just in case the transition probability from i to $i - 1$ is less than the transition probability from $i - 1$ to i.

Set intensity of selection to zero $w = 0$, giving $f^0 = g^0 = 1$. From these fitnesses we determine the relevant transition probabilities.

$$P_{i,i-1} = \frac{i\,(N - i + \eta(2i - N))}{N^2}$$

[22]Note that the stipulation of positive payoffs is required as positive fitness values are needed for the Moran process to be well-defined.

$$P_{i,i+1} = \frac{(N-i)\,(i+\eta(N-2i))}{N^2}$$
$$P_{i-1,i} = \frac{(N-i-1)\,(i-1+\eta(N-2i-2))}{N^2}$$

Now, we find the conditions under which $P_{i,i-1} < P_{i-1,i}$ and $P_{i,i-1} > P_{i-1,i}$ in terms of i, N, and η. This will tell us when the mass of states in the stationary distribution is increasing, and when it is decreasing. Unpacking the inequality $P_{i,i-1} < P_{i-1,i}$, we get

$$\frac{i\,(N-i+\eta(2i-N))}{N^2} < \frac{(N-i-1)\,(i-1+\eta(N-2i-2))}{N^2}$$

which, with some algebra, yields

$$(1-\eta(N+2))\,(N-2i+1) < 0. \qquad (*)$$

We make the necessary restrictions, $2 \leq N$ and $1 \leq i \leq N$, and denote the term on the left hand side of the inequality $(*)$ by h. We see that, when $\eta(N+2) > 1$, $i < \frac{N+1}{2}$ implies $h < 0$ and $i > \frac{N+1}{2}$ implies $h > 0$. Whereas, when $\eta(N+2) < 1$, $i < \frac{N+1}{2}$ implies $h > 0$ and $i > \frac{N+1}{2}$ implies $h < 0$.

That is, when strong mutation obtains, the mass μ_i of a state i in the stationary distribution is greater than that of its preceding state $i-1$ over the first half of the state space $i < \frac{N+1}{2}$, and less than that of its preceding state over the second half of the state space $i > \frac{N+1}{2}$. Thus, the stationary distribution μ must exhibit a unique mode exactly at (or, when N is even, at the states directly adjacent to) the center of the state space $i = \frac{N+1}{2}$. And when strong mutation does not obtain, the relation between the mass of adjacent states are precisely reversed, and the stationary distribution must exhibit two modes, one at each of the monomorphic states $i = 0$ and $i = N$.

Thus, in the absence of selection, strong mutation is necessary and sufficient for the mode of the stationary distribution to be a polymorphic state. □

To tackle our theorem, first we prove some helpful lemmas.

Lemma 2. *All else being equal, increasing intensity of selection exaggerates selection in favor of the fitter type. That is, if $w < w'$, then $f^w > g^w$ just in case*

$$\frac{if^w}{if^w + (N-i)g^w} < \frac{if^{w'}}{if^{w'} + (N-i)g^{w'}} \quad \textit{and} \quad \frac{(N-i)g^w}{if^w + (N-i)g^w} > \frac{(N-i)g^{w'}}{if^{w'} + (N-i)g^{w'}}.$$

Proof. Consider two versions of the same process, $\{X_t^{N,\eta,w}\}$ and $\{X_t^{N,\eta,w'}\}$, differing only in that the latter has greater intensity of selection, $w < w'$. Suppose $f_i^w > g_i^w$ for some $i \in \chi$. Then $f^w - g^w = wk$ and $f^{w'} - g^{w'} = w'k$ where
$k = (a(i-1) + (b(N-i))) - (ci + d(N-i-1))/(N-1)$. Hence
$\frac{f^w - g^w}{f^{w'} - g^{w'}} = \frac{wk}{w'k} = \frac{w}{w'} < 1$. We now have that $0 < f^w - g^w < f^{w'} - g^{w'}$, and so $\frac{f^w}{g^w} < \frac{f^{w'}}{g^{w'}}$.
We turn to the selection terms of our transition probabilities, and observe that the

following inequalities are equivalent.

$$\frac{f^w}{g^w} < \frac{f^{w'}}{g^{w'}}$$
$$\frac{i}{N-i}\frac{f^w}{g^w} < \frac{i}{N-i}\frac{f^{w'}}{g^{w'}}$$
$$if^w(N-i)g^{w'} < if^{w'}(N-i)g^w$$
$$if^w(N-i)g^{w'} + (if^w \cdot if^{w'}) < if^{w'}(N-i)g^w + (if^w \cdot if^{w'})$$
$$if^w((N-i)g^{w'} + if^{w'}) < if^{w'}((N-i)g^w + if^w)$$
$$\frac{if^w}{if^w + (N-i)g^w} < \frac{if^{w'}}{if^{w'} + (N-i)g^{w'}}.$$

By similar reasoning, the following are equivalent

$$\frac{f^w}{g^w} < \frac{f^{w'}}{g^{w'}}$$
$$\frac{(N-i)g^{w'}}{if^{w'} + (N-i)g^{w'}} < \frac{(N-i)g^w}{if^w + (N-i)g^w},$$

as required. □

Lemma 3. *All else being equal, increasing intensity of selection exaggerates transition probabilities in favor of the fitter type. That is, if $w < w'$, then $f^w > g^w$ just in case*

$$P^w_{i,i+1} < P^{w'}_{i,i+1} \quad \textit{and} \quad P^w_{i,i-1} > P^{w'}_{i,i-1}.$$

Proof. Denote $A \equiv \frac{if^w}{if^w+(N-i)g^w}$, $A' \equiv \frac{if^{w'}}{if^{w'}+(N-i)g^{w'}}$, $B \equiv \frac{(N-i)g^w}{if^w+(N-i)g^w}$, and $B' \equiv \frac{(N-i)g^{w'}}{if^{w'}+(N-i)g^{w'}}$. Suppose $w < w'$, and $f^w_i > g^w_i$ for some $i \in \chi$. From Lemma 2, we have that $f^w > g^w$ just in case $A < A'$ and $B > B'$. We will make use of the fact that $B = 1 - A$ and $B' = 1 - A'$. Let $\eta < 1/2$. Then the following inequalities are equivalent.

$$A < A'$$
$$A(1-2\eta) + \eta < A'(1-2\eta) + \eta$$
$$(1-\eta)A + \eta(1-A) < (1-\eta)A' + \eta(1-A')$$
$$(1-\eta)A + \eta B < (1-\eta)A' + \eta B'$$
$$(1-\eta)\frac{N-i}{N}A + \eta\frac{N-i}{N}B < (1-\eta)\frac{N-i}{N}A' + \eta\frac{N-i}{N}B'$$
$$P^w_{i,i+1} < P^{w'}_{i,i+1}.$$

By similar reasoning, the following inequalities are equivalent.

$$B > B'$$

$$(1-\eta)\frac{i}{N}B + \eta\frac{i}{N}A > (1-\eta)\frac{i}{N}B' + \eta\frac{i}{N}A'$$
$$P^w_{i,i-1} > P^{w'}_{i,i-1},$$

as required. □

Proof of Theorem 1. Let all else be as before, except our 2×2 symmetric game is now characterized by anti-coordination payoffs $a < c$, $b > d$, with the extra condition required for finite games that $N(b-d) > a-d$.

From Lemma 1, we have that, in the absence of selection $w = 0$, strong mutation $\eta(N+2) > 1$ is necessary and sufficient for $\mu_{i-1} < \mu_i$ for $i \leq \lfloor \frac{N+1}{2} \rfloor$ and $\mu_{i-1} > \mu_i$ for $i > \lceil \frac{N+1}{2} \rceil$.

For nonzero intensity of selection $w > 0$, we will show that $f^w > f^0$ and $g^w < g^0$ for some range of states before a polymorphic critical point i^*. As we will show, it follows that $P^w_{i-1,i} > P^0_{i-1,i}$ and $P^w_{i,i-1} < P^0_{i,i-1}$ which in turn implies that $\frac{P^w_{i-1,i}}{P^w_{i,i-1}} > \frac{P^0_{i-1,i}}{P^0_{i,i-1}} > 1$. From the detailed balance conditions, this yields $\mu_{i-1} < \mu_i$ when $i \leq \lfloor i^* \rfloor$. Similarly, we will find that $\frac{P^w_{i-1,i}}{P^w_{i,i-1}} < \frac{P^0_{i-1,i}}{P^0_{i,i-1}} < 1$, and hence $\mu_{i-1} < \mu_i$ after the critical point, when $i > \lceil i^* \rceil$. This will conclude the proof.

Let $w > 0$. First, we find our critical point i^*. Recall the fitness functions for each type

$$f^w_i = 1 - w + w\frac{a(i-1) + b(N-i)}{N-1} \quad \text{and} \quad g^w_i = 1 - w + w\frac{ci + d(N-i-1)}{N-1}.$$

To find the critical point, we solve for when each type is fitter than the other.

$$f^w - g^w > 0$$
$$a(i-1) + b(N-i) - ci - d(N-i-1) > 0$$
$$i(a-b-c+d) + N(b-d) + (d-a) > 0$$
$$i < \frac{N(b-d) + (d-a)}{b+c-a-d}$$

Hence, $f^w > g^w$ just in case $i < i^* = \frac{N(b-d)+(d-a)}{b+c-a-d}$.

From Lemma 1, we know that, whenever the strong mutation condition is satisfied, the mass of stationary distribution of the process in the absence of selection μ^0_i is increasing over the first half of the state space $i < \frac{N+1}{2}$, and decreasing over the second half $i > \frac{N+1}{2}$.

From Lemma 3, it follows from $w > 0$ that $f^w > g^w$ implies $P^0_{i,i+1} < P^w_{i,i+1}$ and $P^0_{i,i-1} < P^w_{i,i-1}$, which obtains for $i < i^*$, and $f^w < g^w$ implies $P^0_{i,i+1} > P^w_{i,i+1}$ and $P^0_{i,i-1} > P^w_{i,i-1}$, which obtains for $i > i^*$. Hence, when $f^w > g^w$ and $\mu^0_i > \mu^0_{i-1}$, we have that $\frac{P^w_{i-1,i}}{P^w_{i,i-1}} > \frac{P^0_{i-1,i}}{P^0_{i,i-1}} > 1$. And, when $f^w < g^w$ and $\mu^0_i < \mu^0_{i-1}$, we have that $\frac{P^w_{i-1,i}}{P^w_{i,i-1}} < \frac{P^0_{i-1,i}}{P^0_{i,i-1}} < 1$.

From this, and the detailed balance conditions, we know that μ^w must be increasing for $i \leq \min\{\lfloor i^* \rfloor, \lfloor \frac{N+1}{2} \rfloor\}$, and decreasing for $i > \max\{\lceil i^* \rceil, \lceil \frac{N+1}{2} \rceil\}$. Thus, we have that the stationary distribution μ^w must find it maximum value at a polymorphic state somewhere in a state between $i^* = \frac{b-d+(\frac{d-a}{N})}{b+c-a-d}$ and $\frac{N+1}{2}$. □

Proof of Corollary 1. Consider an anti-coordination game under the Moran process, as before. Suppose the strong mutation condition $\eta(N+2) > 1$ is satisfied, and consider two intensities of selection w, w' where $0 < w < w' \leq 1$. Then, for every population state prior to the critical point $i^* = \frac{d-b+(\frac{a-d}{N})}{d-c-b+a}$ we know that $f^w > g^w$ and $f^{w'} > g^{w'}$. By lemma 3, $w < w'$ implies that $P^w_{i,i-1} > P^{w'}_{i,i-1}$ and $P^w_{i-1,i} < P^{w'}_{i-1,i}$. So $\frac{P_{i,i-1}}{P^w_{i-1,i}} > \frac{P^{w'}_{i,i-1}}{P^{w'}_{i-1,i}}$ which gives us, from the detailed balance conditions, that $\frac{\mu^w_i}{\mu^w_{i-1}} < \frac{\mu^{w'}_i}{\mu^{w'}_{i-1}}$.

This means that the increase in mass $(\mu_i - \mu_{i-1})$ in every state states prior to the critical point i^* is greater (though, of course, it may be still be negative for some states between i^* and $\frac{N+1}{2}$) for $\mu^{w'}$ than for μ^w. It is easy to see that the inverse inequalities obtain for states after the critical point i^*, and so the rate of decrease in mass is greater for $\mu^{w'}$ than for μ^w for $i > i^*$.

By the conservation of mass of the stationary distribution, $\sum_i \mu_i = 1$, if the rate of increase (of mass) for every state of a distribution $\mu^{w'}$ is greater than another μ^w to the left of a critical point $i < i^*$ and the rate of decrease for every state of $\mu^{w'}$ is greater than for μ^w to the right of that critical point $i^* > i$, then $\mu^{w'}$ must place greater mass than μ^w on the state(s) nearest the critical point. Hence, the mass of the state(s) nearest the critical point $i^* = \frac{N(b-d)+(d-a)}{b+c-a-d}$ is increasing in intensity of selection. □

References

Benaïm, M. and J. Weibull (2003). Deterministic Approximation of Stochastic Evolution in Games. *Econometrica 71*(3), 873–903.

Benaïm, M. and J. Weibull (2009). Mean-field approximation of stochastic population processes in games. Technical Report 1979.

Binmore, K. and L. Samuelson (1997). Muddling Through: Noisy Equilibrium Selection. *Journal of Economic Theory 74*(2), 235–265.

Binmore, K., L. Samuelson, and R. Vaughan (1995). Musical chairs: modeling noisy evolution. *Games and economic behavior 11*(1), 1–35.

Bowles, S. and H. Gintis (2011). *A Cooperative Species: Human Reciprocity and its Evolution.*

Boyd, R. and P. J. Richerson (1985). *Culture and the Evolutionary Process*, Volume 175.

Cressman, R. (2003). *Evolutionary Dynamics and Extensive Form Games.* MIT Press.

Cressman, R. and Y. Tao (2014). The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences 111*(Supplement_3), 10810–10817.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow (1998). Rates of spontaneous mutation. *Genetics 148*(4), 1667–1686.

Fudenberg, D. and L. Imhof (2004). Stochastic Evolution as a Generalized Moran Process. *Unpublished Manuscript*, 1–26.

Fudenberg, D. and L. A. Imhof (2006). Imitation processes with small mutations. *Journal of Economic Theory 131*(1), 251–262.

Fudenberg, D., M. A. Nowak, C. Taylor, and L. A. Imhof (2006). Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theoretical Population Biology 70*(3), 352–363.

García, J. and A. Traulsen (2012). The structure of mutations and the evolution of cooperation. *PLoS ONE 7*(4).

Hamilton, W. D. (1967). Extraordinary Sex Ratios. *Science 156*(3774), 477–488.

Harper, M. and D. Fryer (2016). Stationary stability for evolutionary dynamics in finite populations. *Entropy 18*(9).

Imhof, L. A. and M. A. Nowak (2006). Evolutionary game dynamics in a Wright-Fisher process. *Journal of Mathematical Biology 52*(5), 667–681.

Karlin, S. and H. E. Taylor (2012). *A First Course in Stochastic Processes: Second Edition.*

Kumar, S. and S. Subramanian (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America 99*(2), 803–808.

Maynard Smith, J. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology 47*(1), 209–221.

Mohseni, A. (2017). The Limitations of Equilibrium Concepts in Evolutionary Games. *Unpublished Manuscript*.

Moran, P. A. (1962). *The Statistical Processes of Evolutionary Theory.* Clarendon Press, Oxford.

Novak, M. (2007). *Evolutionary Dynamics: Exploring the Equations of Life*, Volume 82.

Nowak, M. A., A. Sasaki, C. Taylor, and D. Fudenberg (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature 428*(6983), 646–650.

Ohtsuki, H., P. Bordalo, and M. A. Nowak (2007). The one-third law of evolutionary dynamics. *Journal of Theoretical Biology 249*(2), 289–295.

Sandholm, W. H. (2007). Simple formulas for stationary distributions and stochastically stable states. *Games and Economic Behavior 59*(1), 154–162.

Sandholm, W. H. (2009). Population Games and Evolutionary Dynamics. *Population (English Edition)*, xvi, 556.

Sandholm, W. H. (2010). Orders of limits for stationary distributions, stochastic dominance, and stochastic stability. *Theoretical Economics 5*(1), 1–26.

Sandholm, W. H. (2012). Stochastic imitative game dynamics with committed agents. *Journal of Economic Theory 147*(5), 2056–2071.

Skyrms, B. (1996). *Evolution of the social contract.* Cambridge University Press.

Taylor, C., D. Fudenberg, A. Sasaki, and M. A. Nowak (2004). Evolutionary game dynamics in finite populations. *Bulletin of Mathematical Biology 66*(6), 1621–1644.

Taylor, P. D. and L. B. Jonker (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences 40*(1-2), 145–156.

Tenaillon, O., J. E. Barrick, N. Ribeck, D. E. Deatherage, J. L. Blanchard, A. Dasgupta, G. C. Wu, S. Wielgoss, S. Cruveiller, C. Médigue, D. Schneider, and R. E. Lenski (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature 536*(7615), 165–170.

Traulsen, A. and C. Hauert (2010). Stochastic Evolutionary Game Dynamics. In *Reviews of Nonlinear Dynamics and Complexity*, Volume 2, pp. 25–61.

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

Young, H. P. (1998). *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*, Volume 110.

Young, H. P. (2005). The Diffusion of Innovations in Social Networks. *The Economy as an Evolving Complex System III* (1966), 267–282.

Young, H. P. (2015). The Evolution of Social Norms. *Annual Review of Economics* *7*(1), 359–387.

UC Irvine, Irvine, CA 92697, USA
E-mail address: amohseni@uci.edu