Theoretical Economics 8 (2013), 829–874 1555-7561/20130829

Deterministic equations for stochastic spatial evolutionary games

Sung-Ha Hwang
School of Economics, Sogang University

Markos Katsoulakis
Department of Mathematics and Statistics, University of Massachusetts at Amherst,
and Department of Applied Mathematics, University of Crete

Luc Rey-Bellet
Department of Mathematics and Statistics, University of Massachusetts at Amherst

Spatial evolutionary games model individuals playing a game with their neighbors in a spatial domain and describe the time evolution of the strategy profile of individuals over space. We derive integro-differential equations as deterministic approximations of strategy revision stochastic processes. These equations generalize the existing ordinary differential equations such as replicator dynamics and provide powerful tools for investigating the problem of equilibrium selection. Deterministic equations allow the identification of many interesting features of the evolution of a population's strategy profiles, including traveling front solutions and pattern formation.

Keywords. Spatial evolutionary games, deterministic approximation, long-range interactions, equilibrium selection, traveling front solutions, pattern formation.

JEL classification. C70, C72, C73.

1. Introduction

Various economic phenomena are typically aggregate outcomes of interactions among a large number of seemingly unrelated agents. The occurrence of depression may depend on decisions of many agents to save or consume. Some institutional changes such as the transformation of conventional crop share systems involve a large number of actors with limited information and occur based on the collective action of these actors through decentralized processes. Interactions between agents are inherently local because they are

Sung-Ha Hwang: sunghah@sogang.ac.kr
Markos Katsoulakis: markos@math.umass.edu
Luc Rey-Bellet: luc@math.umass.edu
We thank William Sandholm, a co-editor, and three anonymous referees for their valuable suggestions and comments. The work of S.-H. H. was partly supported by National Science Foundation Grant NSF-DMS-0715125 and was partly supported by the Sogang University Research Grant of 201210055.01. The work of M. A. K. was partly funded by the European Union (ESF) and Greece (NSRF) through the Thalis Grant Opt-Man. The work of L. R.-B. was supported by National Science Foundation Grants NSF-DMS-06058 and NSF-DMS-1109316.

DOI: 10.3982/TE829

separated by spatial locations, languages, and cultures, and agents are more likely to interact with some agents than with others. Schelling (1971) examines spatial interactions in deciding residential areas and shows that the preference for racial integration may lead to spatial segregation in residential areas. Bisin and Verider (2011) study cultural transmission and socialization, and identify geographical diffusion as one of the major mechanisms underlying this transmission (see Bowles 2004 for more applications).

To examine such spatial interactions, we consider a class of spatial stochastic processes in which agents at various spatial locations interact with their neighbors by playing games. We develop and investigate deterministic approximations of spatial stochastic processes. A number of studies have examined spatial stochastic models in the context of evolutionary game theory (Blume 1993, 1995, Ellison 1993, 2000, Young 1998, Chapter 5, and Young and Burke 2001; see Szabo and Fath 2007 for a survey of spatial evolutionary games). In such models, agents typically interact with a finite set of neighbors and thus they are referred to as local interaction models. Important questions about equilibrium selection and the speed of convergence to equilibrium have been addressed by analyzing stochastic dynamics directly. For instance, Ellison (1993) shows that local interactions lead to a rapid transition to a long-run equilibrium in learning dynamics and Blume (1995) shows that spatial variations can replace persistent randomness, which is necessary for equilibrium selection in global interaction models.

However, existing methods and results are confined to special classes of games such as potential or coordination games or to a limited number of behavioral and learning rules such as perturbed best-response and logit choice rules. It is known that the implications of specific behavioral or learning rules for evolutionary dynamics are quite different. In the stochastic dynamics of uniform interaction models, as in Kandori et al. (1993) and Young (1993), the dynamics based on perturbed best-response rules select a risk-dominant equilibrium, while those based on imitation behaviors select a Pareto-superior Nash equilibrium (Josephson and Matros 2004, Robson and Vega-Redondo 1996). Grauwin and Jensen (2012) examine Schelling's models of residential segregation by adopting the logit choice rule, but the choice of residence can be based on other behavioral rules, notably imitation rules. After all, imitation is one of the most important decision-making rules in everyday life (Levine and Pesendorfer 2007, Apesteguia et al. 2007, Bergin and Bernhardt 2009, Apesteguia et al. 2010). The model of imitation necessarily involves questions such as who to choose as a role model, and spatial considerations such as neighbors and proximity are intrinsic aspects of sophisticated models of such behaviors.

What are the implications of various learning and behavioral rules for spatial stochastic processes? Specifically, how do imitation behaviors (in comparison to perturbed best-response behaviors) contribute to spatial decision-making processes such as the choice of residential areas? What are the implications of various behavioral rules for the speed of convergence to equilibrium and equilibrium selection? Addressing such questions by using existing methods for analyzing stochastic processes directly can easily become a mathematically challenging problem (see, e.g., Durrett 1999). In addition, existing methods for examining equilibrium selection problems (Kandori et al. 1993, Young 1993, Ellison 1993, 2000) are not applicable to the case in which perturbed stochastic processes violate some regularity conditions such as ergodicity. To tackle this difficulty, we adopt deterministic approximation methods for analyzing stochastic processes, derive a class of differential equations that approximate the original stochastic process, and rigorously justify such approximations. The deterministic approximation method is widely used to analyze uniform or global interaction models (see Benaim and Weibull 2003), but the extension of this method to spatial stochastic processes involves nontrivial and technically demanding tasks.

Our main results rely on the insight that deterministic approximation methods are still applicable to spatial stochastic processes under the key condition that the range of spatial interactions is sufficiently long. We thus focus on the spatial models of *long-range interactions* in which a player interacts with a substantial portion of the population, but spatial variations in the strength of interactions are nonetheless allowed.

In the current literature of evolutionary game and learning theory (e.g., Hofbauer and Sigmund 2003, Weibull 1995, Fudenberg and Levine 1998, Sandholm 2010b), the time evolution of the proportion of agents with strategy i at time t, $f_t(i)$, is described by an ordinary differential equation (ODE) as

$$\frac{d}{dt}f_t(i) = \sum_{k\in S}\mathbf{c}^M(k,i,f_t)f_t(k) - f_t(i)\sum_{k\in S}\mathbf{c}^M(i,k,f_t) \quad \text{for } i\in S. \tag{1}$$

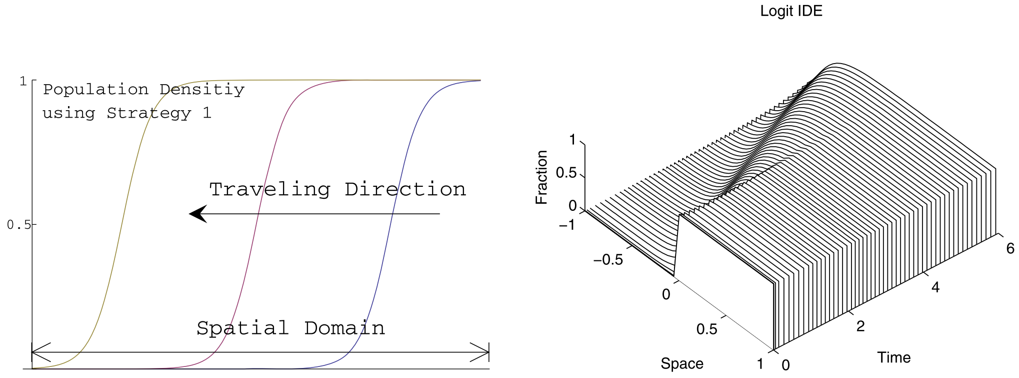
Examples of such equations include the well known replicator dynamics, logit dynamics, and Brown–von Neumann Nash dynamics. The first term in (1) describes the rate at which agents switch to strategy i from some other strategy, whereas the second term describes the rate at which agents switch to some other strategy from strategy i. For this reason, (1) is also called an *input–output* equation.

It is well known (Kurtz 1970, Benaim and Weibull 2003, Darling and Norris 2008) that a solution to (1), $f_t(i)$, approximates, on a finite time interval, a suitable mean stochastic process (or a stochastic process with uniform interactions) in the limit of an infinite population. That is, $f_t(i)$ is the average of the proportion of agents playing strategy i for the entire domain. Our spatial model with long-range interactions describes instead the state of the system by a local density function $f_t(u,i)$. Here u, a spatial location, belongs to the spatial domain $\mathbb{A}\subset\mathbb{R}^n$, where agents are continuously distributed and $f_t(u,i)$ indicates the proportion of agents playing strategy i at u. Our main result is that spatial stochastic processes with long-range interactions are approximated on finite time intervals and in the limit of an infinite population through the equation

$$\frac{\partial}{\partial t}f_t(u,i) = \sum_{k\in S}\mathbf{c}(u,k,i,f_t)f_t(u,k) - f_t(u,i)\sum_{k\in S}\mathbf{c}(u,i,k,f_t) \quad \text{for } i\in S. \tag{2}$$

Equation (2) provides a natural generalization of (1). For example, the term $\mathbf{c}(u,k,i,f)$ describes the rate at which agents at spatial location u switch from strategy k to i. This rate depends on strategies of agents at other spatial locations and, typically, $\mathbf{c}(u,k,i,f)$ takes the functional form

$$\mathbf{c}(u,k,i,f) = G(k,i,\mathcal{J}*f(u,i)), \quad \text{where } \mathcal{J}*f(u,i) := \int \mathcal{J}(u-v)f(v,i)\,dv.$$



(a) Traveling front solution (b) Logit equation

Figure 1. Traveling front solutions. This figure illustrates how a traveling front solution describes the propagation of a strategy over the whole spatial domain for a two-strategy game. The heights of the solutions in (a) and (b) are local densities for strategy 1. For (b), $N = 256$, $\Lambda = [-1, 1]$, $dt = 0.001/(0.05N^2)$, $a_{11} = 20/3$, $a_{22} = 10/3$, and $a_{12} = a_{21} = 0$; $\chi = 2$ for the Gaussian kernel. The initial condition is a unit step function at 0.

Here $\mathcal{J} * f$ is the convolution product of $\mathcal{J}$ and f, and $\mathcal{J}(u)$ is a nonnegative probability kernel that describes the strength of the interaction between players whose relative distance is u. If $\mathcal{J}$ is constant, then (2) reduces to (1). This equation is referred to as an *integro-differential equation* (IDE).

One of the major advantages of our framework and equations lies in their flexibility and generality. Following the heuristic derivation method in Section 3.4, one can easily derive specific equations from various behavioral assumptions for any normal form games, study the problem of equilibrium selection, and compare the roles of various behavioral rules in spatial interactions. Traveling front solutions to spatial equations (see Figure 1) show how the adoption of a strategy propagates throughout a spatial domain, selecting an equilibrium from among multiple Nash equilibria.

In the case of two-strategy coordination games, when agents' behavioral rules are perturbed best-response rules, we observe that the system converges everywhere to the risk-dominant equilibrium exponentially fast. By contrast, in the dynamics in which agents adopt imitation rules for strategy revisions, we observe a slow transition to the risk-dominant equilibrium. Similarly, we observe the persistent spatial separation of the choice of strategies in imitation dynamics, whereas no such patterns arise in the dynamics of perturbed best-response rules.

Existing approaches to pattern formation and the existence of traveling front solutions in evolutionary games traditionally employ reaction-diffusion partial differential equations. Such models are typically obtained by adding a diffusion term to the mean dynamic equation, which in turn models the fast but homogeneous spatial diffusion of agents (Hutson and Vickers 1992, Vickers et al. 1993, Hofbauer et al. 1997, Hofbauer, 1997, 1999, Durrett 1999). By contrast, in our scaling-limit approach, spatial effects are introduced at the microlevel, and thus diffusive effects are generally density dependent and differ markedly from equation to equation. This introduces new and interesting spatial phenomena, which are absent in reaction-diffusion equations.

The rest of this paper is organized as follows. Section 2 illustrates the main results in the example of a two-strategy symmetric game under perturbed best-response rules and demonstrates that our deterministic equations can provide a useful technique for analyzing the original stochastic processes. Section 3 introduces the stochastic process and scaling limits, and presents the main results (Section 3.3). Section 3.4 provides a heuristic derivation of the equations, and Section 3.5 elucidates the relationships between spatial models and uniform interaction models. Section 4 analyzes equilibrium selection. In the final section, we informally discuss pattern formation in two-strategy games through a combination of linear stability analysis and numerical simulations. The Appendix provides all proofs, thereby streamlining the presentation of the main results and analyses.

2. Two-strategy symmetric games: Equilibrium selection

In this section, we present the main results of the paper and their implications for the problem of equilibrium selection in a two-strategy symmetric game. Our general results for various games and a large class of updating rules are presented in Section 3. Consider a two-strategy game with the payoff matrix $\{a(i,j)\}_{i,j=1,2}$, where $a(i,j)=a_{ij}$ is the payoff for playing strategy i against j, and $a_{11}>a_{21}$ and $a_{12}<a_{22}$. Suppose that agents are located at various sites of a graph Λ. The strategy of the agent at site x is $\sigma(x)$ and the strategy profile σ is a collection of all agents' strategies in the population, i.e., $\sigma=(\sigma(x_1),\dots,\sigma(x_n))$, where n is the total number of agents in the population. We assign nonnegative weights $\mathcal{W}(x,y)$ to any two sites x and y to capture the importance or intensity of the interaction between neighbors. As in Young (1998, Chapter 6), we define the payoffs for an agent playing strategies 1 and 2 as

$$u(x,\sigma,1)=\sum_{y\in\Lambda}\mathcal{W}(x,y)a(1,\sigma(y)),\qquad u(x,\sigma,2)=\sum_{y\in\Lambda}\mathcal{W}(x,y)a(2,\sigma(y)).$$

Thus, the total payoff for the agent at site x is the weighted sum of payoffs from all games played with his neighbors. Suppose that each agent possesses a random alarm clock. An agent receives a strategy revision opportunity when his clock goes off and he revises his strategy according to some choice rule. This setting defines a Markov process $\{\sigma_t\}$. Suppose that the choice rule is a logit choice rule (Blume 1993) modeling the perturbed best-response behavior of agents. That is, the agent at x chooses strategy 1 (strategy 2, respectively) with probability

$$\frac{\exp(\beta u(x,\sigma,1))}{\exp(\beta u(x,\sigma,1))+\exp(\beta u(x,\sigma,2))}\quad\left(\text{or }\frac{\exp(\beta u(x,\sigma,2))}{\exp(\beta u(x,\sigma,1))+\exp(\beta u(x,\sigma,2))}\right),$$

where $\beta>0$ is the parameter measuring the (inverse) noise level or the degree of perturbation. Then the spatial Markov process admits a stationary distribution of the form

$$\mu(\{\sigma\})\propto\exp\left[\beta\frac{1}{2}\left(\sum_{z,y\in\Lambda}\mathcal{W}(y-z)a(\sigma(y),\sigma(z))+\sum_{y\in\Lambda}\mathcal{W}(0)a(\sigma(y),\sigma(y))\right)\right]. \tag{3}$$

(See Blume (1997) for the analogous expression in the case of local interaction models.)

Recall that strategy 1 is risk dominant if and only if $a_{11} - a_{21} > a_{22} - a_{12}$ (Harsanyi and Selten 1988). It is easy to show that under suitable conditions for $\mathcal{W}$ (as in (4) or Young 1998, p. 97), as the noise level becomes negligible ($\beta \to \infty$), the stationary distribution (3) converges to a point mass on the state in which every agent chooses the risk-dominant strategy. The selected state is called a stochastically stable state and it has playing a significant role in the questions of equilibrium selection.

Instead of analyzing the long-run behaviors of the spatial stochastic process directly, we focus on the time trajectories of the spatial stochastic process and derive a differential equation as a deterministic approximation. To do this, we make the following *scaling* assumption for the interaction:

$$\mathcal{W}(x-y) = n^{-1}\mathcal{J}(n^{-1}(x-y)). \tag{4}$$

Under this assumption, we show that as n approaches infinity, the time trajectory of the empirical distribution for the strategy profile (see (13)) converges to the time trajectory of a local density for the strategy profile f, which satisfies a system of differential equations (Theorems 1 and 2). The function $\mathcal{J}(x)$ captures the spatial variations in the model. Specifically, our deterministic approximation yields the integro-differential equation

$$\text{logit IDE} \quad \frac{\partial}{\partial t} f_t(w) = \frac{e^{\beta(a_{11}\mathcal{J}*f(w)+a_{12}(1-\mathcal{J}*f(w)))}}{e^{\beta(a_{11}\mathcal{J}*f(w)+a_{12}(1-\mathcal{J}*f(w)))} + e^{\beta(a_{21}\mathcal{J}*f(w)+a_{22}(1-\mathcal{J}*f(w)))}} - f_t(w), \tag{5}$$

where $\mathcal{J} * f(w) := \int \mathcal{J}(w-v)f(v)\,dv$ and $f(v)$ is the density of strategy 1 at location v. Equation (5) is a spatial generalization of mean dynamics of the logit choice rule (Sandholm 2010b). As in the analysis of mean dynamics, a handy tool for analyzing the dynamics given by (5) is the functional

$$V(f) := \frac{1}{2}(a_{11} - a_{21} + a_{22} - a_{12}) \int \mathcal{J} * f(w) f(w)\,dw - (a_{22} - a_{12}) \int f(w)\,dw - \frac{1}{\beta} \int \phi(f(w))\,dw, \tag{6}$$

where $\phi(x) := x \log x + (1-x)\log(1-x)$. In Proposition 1, we show that $V(f)$ is the Lyapunov functional of the logit IDE (5); the critical values of $V(f)$ coincide with the stationary solutions and the value of $V(f)$ along the solution to (5) increases over time.

Proposition 1 (Lyapunov functional for spatial logit dynamics). *For a solution f_t that satisfies (5), we have*

$$\frac{d}{dt}V(f_t) \geq 0 \quad \textit{and} \quad \frac{d}{dt}V(f_t) = 0 \quad \textit{if and only if} \quad \frac{\partial}{\partial t} f_t = 0.$$

The Lyapunov functional gives all stationary solutions (spatially homogeneous or inhomogeneous), and the stabilities of such solutions can be studied by analyzing the shape of V. The first two terms in (6) are called the energy part and the last term is called the entropy part. If the noise level β^{-1} is high, then the entropy part dominates in (6) and the dynamic has one spatially homogeneous stationary solution. By contrast, if the noise level β^{-1} is low, then the agents' behaviors are closer to the best-response rules and the system admits three spatially homogeneous stationary solutions. Two of these solutions correspond precisely to the local maxima of the functional V, with one being close to 0 and the other being close to 1 across all spatial locations. This immediately leads to the issue of selecting an equilibrium from two competing stable solutions at deterministic dynamics.

Such equilibrium selection problems in deterministic spatial equations are typically addressed using traveling front solutions. To study these solutions, we examine the time evolution of solutions from initial data in which two stationary states are connected through the spatial domain (see the initial condition in the left panel of Figure 1). The existence of traveling front solutions and the asymptotic stability of such solutions in various cases of logit dynamics follow from results for essentially equivalent equations ((36); see Theorem 5; De Masi et al. 1995, Orlandi and Triolo 1997, Chen 1997). In the right panel of Figure 1, a traveling front solution shows how the choice of strategy 1 propagates over the spatial domain over time.

How are equilibrium selection results for deterministic dynamics related to the corresponding results for the original stochastic processes obtained through the analysis of the invariant measure (3)? Although our approximation results hold for a fixed time window, we will illustrate that the propagation of a strategy can happen within the time window during which the deterministic approximation is still valid. Once the propagation is complete, the entire population persists in playing the strategy, whether the deterministic approximation still holds or not. Hence, the deterministic equation can provide the correct prediction about the long-run behaviors of the spatial stochastic process.

To illustrate this, we first refer to a more general result on deterministic approximations: in the case of logit equations, deterministic approximations are valid for a longer time horizon as long as it is shorter than $\log(n)$ (De Masi et al. 1994). To be more concrete, we consider a large number of agents in the system (e.g., $n = 10^5$). According to the result, deterministic equations are informative up to time $t \approx 11.6$ ($\approx \log(10^5)$). In various numerical simulations of the coordination game (Figures 1, 2, and 4) with interaction intensity $\mathcal{J}(x) \propto \exp(-|x|^2)$, we observe that it usually takes time $t \approx 7$ or 8 for the risk-dominant equilibrium in the coordination game to complete the propagation over the whole spatial domain. The time scale for the deterministic equation is the same as that for the spatial stochastic process in our choice of scaling (see Section 3), which implies that the stochastic process starting from the same initial condition as the deterministic equation remains close to the traveling front solution with high probability until the complete propagation of strategy 1. Moreover, the traveling wave solution to the logit equation is asymptotically stable, attracting various solutions (see, e.g., the condition (38)), which implies that various strategy profiles of the original stochastic process could be attracted to the traveling front solution.

This example illustrates that (i) the speed of convergence to equilibrium in our stochastic model of long-range interaction is fast, as in the case of local interaction models, and (ii) the traveling front solution to the deterministic equation provides a useful tool for analyzing equilibrium selection problems in the spatial stochastic process. In our models, the function $\mathcal{J}$ varies according to the spatial distance, typically putting more weight on interactions between close neighbors, which provides spatial clustering that expedites the propagation of the risk-dominant equilibrium. In this regard, the way that traveling front solutions move can be regarded as Ellison's evolution path in a large population limit toward the risk-dominant equilibrium (Ellison 1993, 2000).

A rigorous theoretical analysis relying on new techniques is required for establishing more precise relationships between the waiting time for the stochastic processes and the speed of the traveling front solution, but we leave such an analysis to future research. However, our deterministic approximation methods yield a promising first step, and these methods themselves often provide a powerful tool for analyzing the original stochastic processes. For example, Kreindler and Young (2013) in their recent work on learning in stochastic models first show that fast learning obtains in the deterministic approximation of the stochastic process and then that the property of fast learning is preserved under the deterministic approximation. Durrett (2002) proves the coexistence of species in the stochastic spatial model by proving persistent results for the corresponding deterministic reaction diffusion equation.

Ellison (1993) shows that in the uniform model, it takes an extremely long time to shift from one equilibrium to the other, while in the local interaction model, convergence to the risk-dominant equilibrium is fast. Our spatial model of interactions also confirms Ellison's result. In the right panel of Figure 2, the spatial model accounting for the locality of interactions shows a fast transition to the state in which all agents adopt the risk-dominant strategy equilibrium (strategy 1). By contrast, the left panel demonstrates that the uniform interaction model (here $\mathcal{J}(x)$ is identically constant over the domain) shows convergence to the state that is risk dominated (strategy 2).

In the uniform interaction case, agents are concerned only about the aggregate behavior of other agents, and the initial proportion of agents playing strategy 2 in the whole population is great enough to offset the payoff advantage of strategy 1. As a result, the system converges to a state in which all agents play strategy 2. However, in a spatial model that accounts for agents' local behaviors, neighbors around the island of agents with strategy 1 switch to strategy 1, and these behaviors propagate over the space. By reducing complicated spatial stochastic processes to spatial differential equations, we obtain more tractable mathematical settings in which important questions concerning the original stochastic processes can be addressed. The uniqueness, existence, and stability of traveling front solutions to IDE's similar to ours have been studied and established under fairly general conditions (see Theorem 5 in Section 4 of Chen 1997). This general result can provide a rigorous characterization for equations arising from a number of interesting strategy revision rules.

How is the very long-term behavior of differential equations (which may be longer than the time scale of $\log(n)$) related to that of the original stochastic processes? In this time scale, long-term predictions of differential equation may be misleading, neglecting

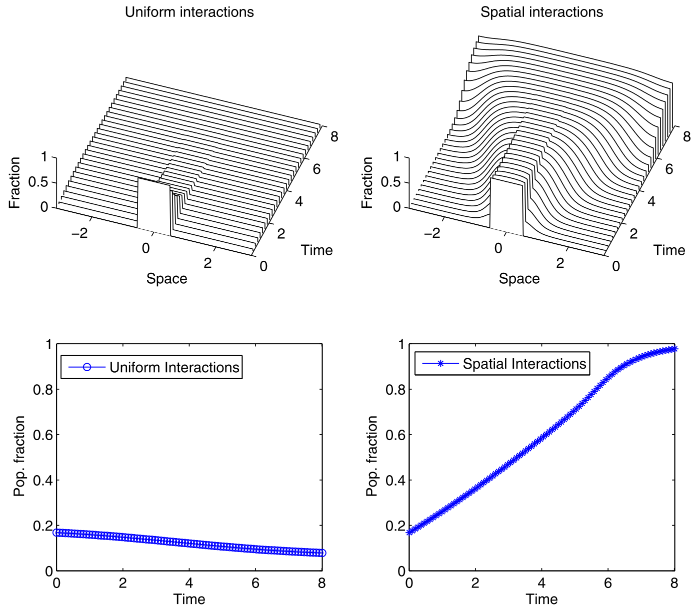


FIGURE 2. Comparison of equilibrium selection between uniform and spatial interactions. The upper panels show how front solutions evolve over time in the uniform interaction model (the left panel) and in the spatial interaction model (the right panel). We consider the same initial conditions under which agents using strategy 1 form an island in the population using strategy 2. The bottom panels show the population densities for strategy 1 when the local density is aggregated over the space. Here, the logit choice rule is used, $N = 512$, $\Lambda = [-\pi, \pi]$, $dt = 0.001/(0.25N^2)$, $a_{11} = 20/3$, $a_{22} = 10/3$, and $a_{12} = a_{21} = 0$; $\chi = 2$ for the Gaussian kernel in Section 4.2. The initial density in the upper panel is $1/6$ and the initial condition is $\mathbf{1}_{[-\pi/6, \pi/6]}$, where $\mathbf{1}_A$ hereafter means a function that takes the value of 1 if u belongs to A and 0 otherwise.

stochasticity and fluctuations. For a large system, the very long-run behavior of stochastic processes can be suitably analyzed using large deviation techniques in conjunction with deterministic approximations. It is well known that in the case of the logit rule and the two-strategy game, the empirical distribution of the strategy profile satisfies a large deviation principle with a rate functional that has exactly the same form as the Lyapunov functional (6) (see Eisele and Ellis 1983). In this case, the analysis of the Lyapunov functional associated with deterministic equations can provide a fruitful prediction of the long-run behavior of stochastic processes via large deviation analysis. In particular, one can study metastable states, and compute the mean exit time and the optimal exit path from the neighborhood of a metastable point. This is an infinite-dimensional version of Freidlin and Wentzell's (1998) theory, and De Masi et al. (1996a, 1996b) and Presutti (2009) have obtained in this way the best results to date on the long-time behavior of complex spatial stochastic processes.

3. Spatial evolutionary games

3.1 *Strategy revision processes*

In models of spatial evolutionary games, agents at various spatial locations play a normal form game with their neighbors. Specifically, we suppose that agents are located at the vertices of a graph Λ, which is a subset of the integer lattice $\mathbb{Z}^d$. We consider one population playing a normal form game, but the generalization to multiple population games is straightforward. A normal form game is specified by a finite set of strategies S and a payoff function $a(i, j)$ that gives the payoff for a player using strategy $i \in S$ against strategy $j \in S$.

The strategy of the agent at site $x \in \Lambda$ is $\sigma_\Lambda(x) \in S$ and we denote by $\sigma_\Lambda = \{\sigma_\Lambda(x) : x \in \Lambda\}$ the configuration of strategies for every agent in the population. With these notations, the state space, i.e., a set of all possible configurations, is S^Λ. The subscript of $\sigma_\Lambda = \sigma$ is suppressed whenever there is no confusion. As in Section 2, we assign nonnegative weights $\mathcal{W}(x - y)$ to any two sites x and y to capture the importance or intensity of the interaction between neighbors. Note that we assume that these weights depend only on the relative location $x - y$ between the players (i.e., we assume translation invariance; see the discussion at the end of Section 3.3). It is convenient to assume that the total weight that an agent at site x assigns to all his neighbors is normalized to 1, that is, $\sum_{y\in\Lambda} \mathcal{W}(x - y) \approx 1$. We say an agent at site y is a neighbor of an agent at site x when $\mathcal{W}(x - y) > 0$. Given a configuration σ, an agent at site x who plays strategy i receives the payoff

$$u(x, \sigma, i) := \sum_{y\in\Lambda} \mathcal{W}(x - y) a(i, \sigma(y)).$$

If we regard $\mathcal{W}$ as the probability with which an agent samples his neighbors, then $u(x, \sigma, i)$ is the expected payoff for an agent at x playing strategy i if the population configuration is σ. An alternative interpretation is that an agent receives an instantaneous payoff flow from his interactions with other neighbors (Blume 1993, Young 1998).

For the special case in which $\mathcal{W}(x - y)$ is constant, the interaction is uniform and there is no spatial structure. In this case, if there is a total of n^d agents in the population, then $\mathcal{W}(x - y) \approx 1/n^d$ because of the normalization condition for $\mathcal{W}$. Alternatively, if $\mathcal{W}(x - y) = 1/2d$ for $\|x - y\| = 1$ and $\mathcal{W}(x - y) = 0$ otherwise, then there are interactions only between nearest sites and the model is referred to as a nearest neighbor model (Blume 1995, Szabo and Fath 2007).

In this paper, we focus on long-range interactions in which each agent interacts with as many agents as in the uniform interaction case, but the interaction is spatial and thus local. We then analyze the limit of spatial stochastic processes under long-range interactions by using unscaled time and scaled space. This model is known as a *local mean field model* (Comets 1987) or a *Kac potential model* (Lebowitz and Penrose 1966, De Masi et al. 1994, Presutti 2009) and such limits are referred to as *mesoscopic scaling limits* in the physics literature. Limits similar to ours have been derived using several models in statistical mechanics (e.g., Comets 1987, De Masi et al. 1994, Katsoulakis et al. 2005). We generalize these results to spatial stochastic processes that arise from evolutionary game theory. Other scaling limits such as hydrodynamic limits in which the space and time are scaled simultaneously (e.g., Katsoulakis and Souganidis 1997) may be relevant to game theory but are not addressed in this paper.

More specifically, let $\mathcal{J}(x)$ be a nonnegative, compactly supported, and integrable function such that $\int \mathcal{J}(x)\,dx = 1$. We often take the support size for $\mathcal{J}$ to be less than or equal to the domain (see (33) and (34) in Section 4). We assume that $\mathcal{W}$ is of the form

$$\mathcal{W}^\gamma(x-y) = \gamma^d \mathcal{J}(\gamma(x-y)), \tag{7}$$

and we take the limit $\gamma \to 0$ and $\Lambda \nearrow \mathbb{Z}^d$ such that $\gamma^{-d} \approx |\Lambda| \approx n^d$. Here n^d is the size of the population, $||$ denotes cardinality, and the factor γ^d is chosen to ensure that $\sum \mathcal{W}^\gamma(x-y) \approx \int \mathcal{J}(x)\,dx = 1$. Note that in (7), the interaction vanishes when $\|x-y\| \geq R\gamma^{-1}$ if $\mathcal{J}$ is supported on the ball of radius R. Thus, as $\gamma \to 0$, an agent interacts very weakly but with a growing number of neighbors in the population.

The time evolution of the system is given by a continuous-time Markov process $\{\sigma_t\}$ with state space S^Λ. Each agent receives, independently of all other agents, a strategy revision opportunity in response to his own Poisson alarm clock according to rate 1 and then updates his strategy according to a rate $c(x,\sigma,k)$—the rate at which an agent at site x switches to strategy k when the configuration is σ. Then this process is specified by the generator (see Ethier and Kurtz 1986, Ligget 1985)

$$(Lg)(\sigma) = \sum_{x\in\Lambda}\sum_{k\in S} c(x,\sigma,k)(g(\sigma^{x,k}) - g(\sigma)),$$

where g is a bounded function on S^Λ, and $\sigma^{x,k}(y) = \sigma(y)$ if $y \neq x$ and $\sigma^{x,k}(y) = k$ if $y \neq x$. Thus $\sigma^{x,k}$ represents a configuration in which an agent at site x switches from his current strategy $\sigma(x)$ to some new strategy k.

If a stochastic process can introduce a new strategy that is not currently used in the population, then we refer to this process as an *innovative* process. If a strategy that is not present in the population does not reappear under the dynamics, we refer to the process as a *noninnovative* process. In addition, if, on switching, agents consider only the payoff of the new strategy, then we refer to the process as a *targeting* process. In contrast, if agents' decisions depend on the difference in the payoff between the current strategy and the new strategy, then we refer to the process as a *comparing* process (Szabo and Fath 2007, Sandholm 2010b).

Precise technical assumptions about the strategy revision rates are discussed later (conditions C1–C3 in Section 3.3). Here we provide only a few concrete examples that are commonly used for applications. Several more examples of rates are discussed in the Appendix, and the assumptions about rates are satisfied by virtually all dynamics commonly used in evolutionary game theory (for a more comprehensive discussion on rates and more examples, see Sandholm 2010b). In addition, one can easily define a new strategy revision rate that models some other interesting behavioral and learning rules, and still satisfies C1–C3. We first introduce a notation for the probability that an agent at site x finds a neighbor with strategy k,

$$w(x,\sigma,k) := \sum_{y\in\Lambda} \mathcal{W}(x-y)\delta(\sigma(y),k), \tag{8}$$

where $\delta(i, j) = 1$ if $i = j$ and 0 otherwise. Then, because $\sum_{k \in S} w(x, \sigma, k) = 1$, (8) indeed defines a probability distribution over S, according to which an agent at site x samples a neighbor with strategy k. Let F be a nondecreasing function.

Examples of rates

• *Targeting and innovative*: This case arises if $c(x, \sigma, k) = F(u(x, \sigma, k))$ and $F > 0$. If

$$c(x, \sigma, k) = \frac{\exp(\beta u(x, \sigma, k))}{\sum_l \exp(\beta u(x, \sigma, l))}, \tag{9}$$

then the rate is known as the *logit choice rule* (Blume 1993). It is also called Gibbs sampler in statistics and Glauber dynamics in physics. The inverse of β captures the noise level. Here $\beta = 0$ means the randomization of strategies, and the choice rule approaches the best-response rule as β approaches infinity. For this reason, (9) is often referred to as the perturbed best-response rule. When the underlying normal form game is a potential game, the spatial stochastic process is known to be time reversible, and the invariant measure is often called the Gibbs measure (Blume 1993).

• *Comparing*: This rate is of the form $c(x, \sigma, k) = F(u(x, \sigma, k) - u(x, \sigma, \sigma(x)))$, and is comparing and innovative provided that $F > 0$. The dynamics induced by this kind of rate are also known as pairwise comparison dynamics. Smith (1984) introduces them to study a dynamic model of traffic assignment and shows that the dynamic converges to Nash equilibria. Recently, Sandholm (2010a) shows that under various kinds of pairwise comparison dynamics, the set of stationary states of the dynamics is identical to the set of Nash equilibria. As an another example, we let

$$c(x, \sigma, k) = \min\{1, \exp(\beta(u(x, \sigma, k) - u(x, \sigma, \sigma(x))))\}.$$

For this behavioral rule, the maximum rate is truncated at 1, and if a strategy gives a higher or equal payoff than the current strategy, then an agent chooses that strategy with rate 1. If a strategy gives a lower payoff than the current one, then an agent chooses that strategy with rate $\exp(\beta(u(x, \sigma, k) - u(x, \sigma, \sigma(x))))$. As β approaches infinity, the rate approaches the choice rule in which the agent does not choose any strategy that gives lower payoffs than his current one, but still chooses the strategy that gives higher or equal payoffs than his current one with rate 1. As β approaches zero, an agent chooses any strategy with the same rate, namely 1. Similarly to the logit choice rule, when the normal form game is a potential game, the corresponding Markov chain is reversible and has the same invariant measure as the dynamics of the logit choice rule. This rate is known to give a Metropolis algorithm in Monte Carlo simulations.

• *Comparing and noninnovative*: This rate is of the form

$$c(x, \sigma, k) = w(x, \sigma, k)\big(F\big(u(x, \sigma, k) - u(x, \sigma, \sigma(x))\big)\big). \tag{10}$$

This rate can model imitation behaviors as follows. The first factor $w(x, \sigma, k)$ is the probability that an agent at site x chooses an agent with strategy k ((8)) and the second factor $F(u(x, \sigma, k) - u(x, \sigma, \sigma(x)))$ gives the rate at which some new strategy k is adopted. Therefore, the first term specifies who to imitate and the second term specifies how to imitate, as in the standard literature (Weibull 1995, Benaim and Weibull 2003, Hofbauer and Sigmund 2003). An important example is

$$c(x,\sigma,k)=w(x,\sigma,k)\big[u(x,\sigma,k)-u(x,\sigma,\sigma(x))\big]_+, \tag{11}$$

where $[s]_+=\max\{s,0\}$. In the uniform interaction case, the rate (11) gives rise to famous replicator ODE's as deterministic approximations. More generally if F in (10) satisfies $F(s)-F(-s)=s$, then the corresponding mean ODE is the replicator dynamic. Note that the function $[s]_+$ satisfies this condition. In this paper, we frequently adopt

$$F_\kappa(s):=\frac{1}{\kappa}\log(\exp(\kappa s)+1), \tag{12}$$

and it is easily seen that the function (12) satisfies the condition ($F(s)-F(-s)=s$) too and converges uniformly to $[s]_+$ as $\kappa\to\infty$. Thus (12) can serve as the smooth regularization of (11).

3.2 *Mesoscopic scaling and long-range interactions*

We consider the limit $\gamma\to 0$ under the assumption of (7). As the order of the interaction range γ^{-1} approaches infinity, an agent at site x interacts with a growing number of agents. To obtain limiting equations, we rescale the space and take a continuum limit with the time unscaled. Specifically, let $\mathbb{A}\subset\mathbb{R}^d$ (called the mesoscopic domain) and $\mathbb{A}^\gamma:=\gamma^{-1}\mathbb{A}\cap\mathbb{Z}^d$ (the microscopic domain). If $\mathbb{A}$ is a smooth region in $\mathbb{R}^d$, then $\mathbb{A}^\gamma$ contains approximately $\gamma^{-d}|\mathbb{A}|$ lattice sites, and as $\gamma\to 0$, $\gamma\mathbb{A}^\gamma$ approximates $\mathbb{A}$. Here we recall that $|\mathbb{A}|$ denotes the cardinality of $\mathbb{A}$. Concerning the time scale, we use microscopic time. Thus, our scaling consists of (i) a long-range interaction scale ((7)), (ii) a continuum space (rescaled space), and (iii) microscopic time (unscaled time), and is referred to as mesoscopic scaling.

At the mesoscopic scale, the state of the system is described by the *strategy profile* $f_t(u,i)$—the density of agents with strategy i at site u. The bridge between the microscopic and mesoscopic scales is provided by an *empirical measure* π^γ_σ defined as follows. For $(v,j)\in\mathbb{A}\times S$, let $\delta_{(v,j)}$ denote the Dirac delta measure at (v,j).

Definition 1 (Empirical measure). The empirical measure $\pi^\gamma_\sigma : S^{\mathbb{A}^\gamma}\to\mathcal{P}(\mathbb{A}\times S)$ is a map given by

$$\sigma\mapsto\pi^\gamma_\sigma:=\frac{1}{|\mathbb{A}^\gamma|}\sum_{x\in\mathbb{A}^\gamma}\delta_{(\gamma x,\sigma(x))}, \tag{13}$$

where $\mathcal{P}(\mathbb{A}\times S)$denotes a set of all probability measures on $\mathbb{A}\times S$.

In addition to the empirical measure, for convenience, we define a measure m on $\mathbb{A}\times S$ to be a product measure of the Lebesgue measure on $\mathbb{A}$ and the counting measure on S, i.e., $m:=du\otimes di$, where du is the Lebesgue measure on $\mathbb{A}$ and di is the counting measure on S. To state the main theorem, we denote by fm a measure $v(E):=\int_E f\,dm$. Then the main result shows that, under suitable conditions,

$$\pi^\gamma_{\sigma_t}\to f_t m \quad \text{in probability} \tag{14}$$

and f_t satisfies an integro-differential equation. Because σ_t is the state of the microscopic system at time t, $\pi^{\gamma}_{\sigma_t}$ is a random measure, and f_t is a solution to a deterministic equation. Thus, (14) is in a sense a form of a time-dependent law of large numbers. For this result to hold, we need to assume that the initial distribution of σ_0 is sufficiently regular. For our purpose it is sufficient to assume that the distribution of σ_0 is given by a *product measure with a slowly varying parameter*, which is defined as follows.

Definition 2 (Product measure with a slowly varying parameter). For a given continuous profile f, we define a measure

$$\mu_\gamma := \bigotimes_{x\in\mathbb{A}^\gamma} \rho_x \text{ on } S^{\mathbb{A}^\gamma}, \quad \text{where } \rho_x(\{i\}) = f(\gamma x, i).$$

We call such a collection of measures $\{\mu^\gamma\}_\gamma$ a family of product measures with a slowly varying parameter associated to f.

More general initial distributions can be accommodated (see Kipnis and Landim 1999) if they can be associated with a mesoscopic strategy profile. In addition, we consider two types of boundary conditions as follows.
(a) *Periodic boundary conditions*. Let $\mathbb{A} = [0,1]^d$. We assume that $\mathbb{A}^\gamma = \gamma^{-1}\mathbb{A}\cap\mathbb{Z}^d = [0,\gamma^{-1}]^d\cap\mathbb{Z}^d$, and then extend the profile $f_t(u,i)$ and the configuration $\sigma_{\mathbb{A}^\gamma}$ periodically to $\mathbb{R}^d$ and $\mathbb{Z}^d$. Equivalently, we can identify $\mathbb{A}$ with the torus $\mathbf{T}^d$ and, similarly, identify $\mathbb{A}^\gamma$ with the discrete torus $\mathbf{T}^{d,\gamma}$.
(b) *Fixed boundary conditions*. In applications, it is useful to consider the case in which the domain is simply a subset of $\mathbb{R}^d$ or $\mathbb{Z}^d$. To accommodate this domain, we suppose that the strategies of agents outside the subset do not change over time. Specifically, let $\Lambda\subset\mathbb{R}^d$, where Λ is bounded. For a given domain Λ, we can define the boundary region as follows: Because we consider compactly supported $\mathcal{J}$, we can take, for suitable $r>0$, $\Gamma := \bigcup_{u\in\Lambda} B(u,r)$, where B denotes a ball centered at u with radius r that covers the support of $\mathcal{J}$. The region Γ includes all agents who are relevant in the evolution of the dynamics. Then we consider $\partial\Lambda := \Gamma\setminus\Lambda$ as the "boundary region," where agents do not revise their strategies. Based on these assumptions, we define the microscopic spaces $\Lambda^\gamma := \gamma^{-1}\Lambda\cap\mathbb{Z}^d$ and $\Gamma^\gamma := \gamma^{-1}\Gamma\cap\mathbb{Z}^d$.

3.3 *Main results*

We first consider the case with periodic boundary conditions. For the interaction weights $\mathcal{W}^\gamma(x-y)$, we make the following assumption.

(F) We have $\mathcal{W}^\gamma(x-y) = \gamma^d\mathcal{J}(\gamma(x-y))$, where $\mathcal{J}$ is nonnegative, continuous with compact support, and normalized, $\int\mathcal{J}(x)\,dx = 1$.

Let $\{\sigma^\gamma_t\}_{t\geq 0}$ be a stochastic process with the generator L^γ given by

$$(L^\gamma g)(\sigma) = \sum_{x\in\mathbf{T}^{d,\gamma}}\sum_{k\in S} c^\gamma(x,\sigma,k)(g(\sigma^{x,k}) - g(\sigma))$$

for $g \in L^\infty(S^{\mathbf{T}^{d,\gamma}})$. For the strategy revision rate $c^\gamma(x,\sigma,k)$, we assume that there is a real-valued function

$$\mathbf{c}_0(u,i,k,\pi), \quad u \in \mathbf{T}^d, i,k \in S, \pi \in \mathcal{P}(\mathbf{T}^d \times S)$$

such that the following conditions hold.

C1. The function $\mathbf{c}_0(u,i,k,\pi)$ satisfies

$$\lim_{\gamma\to 0} \sup_{x\in\mathbf{T}^{d,\gamma},\sigma\in S^{\mathbf{T}^{d,\gamma}},k\in S} |c^\gamma(x,\sigma,k) - \mathbf{c}_0(\gamma x,\sigma(x),k,\pi_\sigma^\gamma)| = 0.$$

C2. The function $\mathbf{c}_0(u,i,k,\pi)$ is uniformly bounded, i.e., there exists M such that

$$\sup_{u\in\mathbf{T}^d,i,k\in S,\pi\in\mathcal{P}(\mathbf{T}^d\times S)} |\mathbf{c}_0(u,i,k,\pi)| \le M.$$

C3. The function $\mathbf{c}_0(u,i,k,fm)$ satisfies the Lipschitz condition with respect to f, i.e., there exists L such that for all $f_1, f_2 \in \mathcal{M}(\mathbf{T}^d \times S)$,

$$\sup_{u\in\mathbf{T}^d,i,k\in S} |\mathbf{c}_0(u,i,k,f_1 m) - \mathbf{c}_0(u,i,k,f_2 m)| \le L\|f_1 - f_2\|_{L^1(\mathbf{T}^d\times S)}.$$

When a measure π is absolutely continuous with respect to m, so there exists a measurable function f such that $\pi = fm$, we write $\mathbf{c}(u,i,k,f) := \mathbf{c}_0(u,i,k,\pi)$. In the Appendix, we show that all classes of rates given in the examples in Section 3.1 and several others satisfy C1–C3. Note that if f_1 and f_2 are constant over $\mathbf{T}^d$ or there is no spatial dimension, then f_1 and f_2 can be regarded as points in the simplex Δ. In this case, C3 reduces to the Lipschitz continuity condition in (Benaim and Weibull 2003, p. 878) and, in this way, generalizes their condition. In Section 3.4, we explain how the function $\mathbf{c}(u,i,k,f)$ can be obtained from these rates. We now state our main result.

Theorem 1 (Periodic boundary condition). *Suppose that the revision rate satisfies C1–C3. Let $f \in \mathcal{M}(\mathbf{T}^d \times S)$ and assume that the initial distribution $\{\mu^\gamma\}_\gamma$ is a family of measures with a slowly varying parameter associated with the profile of f. Then for every $T > 0$,*

$$\lim_{\gamma\to 0} \pi_{\sigma_t}^\gamma = f_t m \quad \textit{in probability}$$

uniformly for $t \in [0,T]$ and f_t satisfies the differential equation, for $u \in \mathbf{T}^d$, $i \in S$,

$$\begin{aligned} \frac{\partial}{\partial t} f_t(u,i) &= \sum_{k\in S} \mathbf{c}(u,k,i,f) f_t(u,k) - f_t(u,i) \sum_{k\in S} \mathbf{c}(u,i,k,f) \\ f_0(u,i) &= f(u,i). \end{aligned} \tag{15}$$

As an example of $\mathbf{c}(u,i,k,f)$, when the $c^\gamma(x,\sigma,k)$ is of the form $c(x,\sigma,k) = F(u(x,\sigma,k) - u(x,\sigma,\sigma(x)))$ (comparing and innovative), we have

$$\mathbf{c}(u,i,k,f) = F\Bigl(\sum_{l\in S} a(k,l)\mathcal{J} * f(u,l) - a(i,l)\mathcal{J} * f(u,l)\Bigr) \tag{16}$$

(recall that $\mathcal{J} * f(u,l) := \int_{\mathbf{T}^d} \mathcal{J}(u-v) f(v,l)\,dv$ is the convolution of $\mathcal{J}$ with f). A slight modification of (16) yields corresponding expressions for each choice of $c^\gamma(x,\sigma,k)$ in Section 3.1 (see the Appendix for a complete list of these rates).

Next we consider fixed boundary conditions in Section 3.2. In this case, the stochastic process, $\{\sigma_t\}_{t\geq 0}$, is specified by the generator L^γ as

$$(L^\gamma g)(\sigma_{\Gamma^\gamma}) = \sum_{x\in\Lambda^\gamma}\sum_{k\in S} c^\gamma(x,\sigma_{\Gamma^\gamma},k)(g(\sigma_{\Gamma^\gamma}^{x,k}) - g(\sigma_{\Gamma^\gamma})) \tag{17}$$

for $g \in L^\infty(S^{\Gamma^\gamma})$. Note that the summation in terms of x in (17) is taken over Λ^γ, which shows that only those agents in Λ^γ revise their strategies, whereas the rate itself depends on the configuration throughout Γ^γ. For a given $f \in \mathcal{M}$, we define its restriction on Λ as $f_\Lambda(u,i): f_\Lambda(u,i) = f(u,i)$ if $u\in\Lambda$ and $f_\Lambda(u,i) = 0$ if $u \in \Lambda^C$.

Theorem 2 (Fixed boundary condition). *Suppose that the revision rate satisfies C1–C3. Let $f \in \mathcal{M}(\Gamma^d \times S)$ and assume that the initial distribution $\{\mu^\gamma\}_\gamma$ is a family of measures with a slowly varying parameter associated with the profile of f. Then, for every $T > 0$,*

$$\lim_{\gamma\to 0}\pi^\gamma_{\sigma_t} = \frac{1}{|\Gamma|} f_t m \quad \textit{in probability}$$

uniformly for $t\in[0,T]$ and $f_t = f_{\Lambda,t} + f_{\partial\Lambda,t}$ satisfies the differential equation, for $u\in\Gamma$, $i\in S$,

$$\begin{aligned}\frac{\partial}{\partial t} f_{\Lambda,t}(u,i) &= \sum_{k\in S}\mathbf{c}(u,k,i,f) f_{\Lambda,t}(u,k) - f_{\Lambda,t}(u,i)\sum_{k\in S}\mathbf{c}(u,i,k,f) \\ f_0(u,i) &= f(u,i).\end{aligned} \tag{18}$$

Note that $\mathbf{c}(u,k,i,f) = \mathbf{c}(u,k,i,f_\Lambda + f_{\partial\Lambda})$ is given by a formula similar to (16), with $\mathcal{J} * f(u) = \int_\Gamma \mathcal{J}(u-v) f(v)\,dv$ for $u\in\Lambda$ and, thus, the rate depends on $f_{\partial\Lambda}$ as well as f_Λ. The existence of solutions to IDE's (15) or (18) follows from the proof of the theorems. That is, the convergence of $\pi^\gamma_{\sigma_t}$ to a limit point shows this existence. We provide the proof of uniqueness of the solutions in the Appendix.

The assumption of an integer lattice is for simplicity. First, we can accommodate any other regular lattice (e.g., a hexagonal lattice). More generally, we can extend our theorem by using the same techniques as follows: If a stochastic spatial process is such that there are approximately n^d players (up to corrections of lower order in n) in every region of size 1, then for large n, we can still show the convergence of the empirical measure following the same line as the proof for Theorems 1 and 2. In this sense, the lattice assumption is not necessary and, thus, can be replaced by a local homogeneity condition. Formulating this in a mathematically precise sense would only obscure our results and, therefore, we choose not to do so. The assumption of translation invariance is also not necessary for Theorems 1 and 2. We can replace an interaction of the form $\mathcal{W}(x-y)$ with a general interaction $\mathcal{W}(x,y)$ if $1/n^d \mathcal{W}(x/n, y/n)$ converges to $\mathcal{J}(u,v)$. Then the deterministic equation is suitably modified.

3.4 *Heuristic derivation of the differential equations*

In this section, we heuristically justify the IDE's obtained in Theorems 1 and 2. For simplicity, we assume the periodic boundary condition, but the fixed boundary case is similar. The differential equations (15) and (18) are examples of input–output equations. In particular, by summing f_t over the strategy set, it is easy to see that $\sum_{i\in S} f_t(u,i)$ is independent of t and, thus, if $f_0 \in \mathcal{M}$, then $f_t \in \mathcal{M}$ for all t. In addition, the space $\mathcal{M}$ can be regarded as a product space of the standard strategy simplex Δ of game theory, that is, $\mathcal{M} = \prod_{u\in \mathbf{T}^d} \Delta$. As is shown in evolutionary game theory textbooks (Weibull 1995, Hofbauer and Sigmund 1998, Sandholm 2010b), one can derive heuristically the ODE's from corresponding uniform interaction stochastic processes. Here the main assumption is that the rate depends only on the population average of players with a given strategy. We provide, for the convenience of readers, a similar heuristic derivation of the spatial IDE (15) from stochastic processes with long-range interactions. We replace the global average with spatially localized averages in the limit of the empirical measure (13).

The key idea behind this heuristic derivation is to replace the expected payoff at the microscopic level, $\sum_{y\in\mathbf{T}^{d,\gamma}} \gamma^d \mathcal{J}(\gamma x - \gamma y) a(k,\sigma(y))$, with the expected payoff at the mesoscopic level, $\sum_{j\in S} a(k,j) \mathcal{J} * f(u,j)$. For microscopic sites x and y, we denote by $u=\gamma x$ and $v=\gamma y$ the corresponding spatial positions at the mesoscopic level. For the sake of exposition, suppose that $c^\gamma(x,\sigma,k)$ is innovative and comparing, that is, $c^\gamma(x,\sigma,k) = F(u(x,\sigma,k) - u(x,\sigma,\sigma(x)))$, where F is positive and increasing. For a function g on $\mathbf{T}^d \times S$, by the definition of the empirical measure (13), we can write the sum of values of g over the microscopic domain as the integration of g with respect to the empirical measure π^γ_σ:

$$\frac{1}{|\mathbf{T}^{d,\gamma}|} \sum_{y\in\mathbf{T}^{d,\gamma}} g(\gamma y, \sigma(y)) = \int_{\mathbf{T}^d\times S} g(v,j)\, d\pi^\gamma_\sigma(v,j). \tag{19}$$

Because the number of all sites in the microscopic domain is approximately γ^{-d} ($|\mathbf{T}^{d,\gamma}| \approx \gamma^{-d}$) and the empirical measure π^γ_σ is shown to converge to a smooth density f, we obtain

$$\lim_{\gamma\to 0} \sum_{y\in\mathbf{T}^{d,\gamma}} \gamma^d g(\gamma y, \sigma(y)) = \int_{\mathbf{T}^d\times S} g(v,j) f(v,j)\, dm(v,j)$$

for any g. Thus, if we choose $g(v, j) = \mathcal{J}(u-v)a(k, j)$, then, by using (19), we find that

$$\lim_{\gamma\to 0} \sum_{y\in \mathbf{T}^{d,\gamma}} \gamma^d \mathcal{J}(\gamma(x-y))a(k, \sigma(y)) = \int_{\mathbf{T}^d\times S} a(k, j)\mathcal{J}(u-v)f(v, j)\, dm(v, j)$$
$$= \sum_{j\in S} a(k, j)\mathcal{J} * f(u, j).$$

This shows that as the size of the system increases (as γ goes to 0), the expected payoff for strategy k in the microscopic spatial model becomes the expected payoff evaluated by using the "spatially weighted" fraction of each strategy, namely $\mathcal{J} * f(u, j)$. Thus, if $\sigma(x) = i$, then we obtain

$$\begin{aligned} c^\gamma(x, \sigma, k) &= F\Bigg(\sum_{y\in \mathbf{T}^{d,\gamma}} \gamma^d \mathcal{J}(\gamma x - \gamma y)a(k, \sigma(y)) \\ &\qquad - \sum_{y\in \mathbf{T}^{d,\gamma}} \gamma^d \mathcal{J}(\gamma x - \gamma y)a(\sigma(x), \sigma(y))\Bigg) \qquad (20) \\ &\underset{\gamma\to 0}{\longrightarrow} F\bigg(\sum_{j\in S} a(k, j)\mathcal{J} * f(u, j) - \sum_{j\in S} a(i, j)\mathcal{J} * f(u, j)\bigg) = \mathbf{c}(u, i, k, f), \end{aligned}$$

and this gives (16).

Having identified the rate, we now explain how the IDE (15) is derived. We write

$$\langle \pi_\sigma^\gamma, g\rangle := \int_{\mathbf{T}^d\times S} g(u, i)\, d\pi_\sigma^\gamma, \qquad \langle f, g\rangle := \int_{\mathbf{T}^d\times S} g(u, i)f(u, i)\, dm(u, i),$$

where we view $\langle \pi_\sigma^\gamma, g\rangle$ as a function of the configuration σ. Then, by using (20), we can compute the action of the generator on this function as

$$L_\gamma\langle \pi_\sigma^\gamma, g\rangle = \sum_{k\in S} \int_{\mathbf{T}^d\times S} \mathbf{c}(u, i, k, \pi_\sigma^\gamma)(g(u, k) - g(u, i))\, d\pi_\sigma^\gamma(u, i).$$

From the martingale representation theorem for Markov processes (see, e.g., Ethier and Kurtz 1986), there exists a martingale $M_t^{g,\gamma}$ such that

$$\begin{aligned} \langle \pi_{\sigma_t}^\gamma, g\rangle &= \langle \pi_{\sigma_0}^\gamma, g\rangle \\ &\quad + \int_0^t ds \sum_{k\in S} \int_{\mathbf{T}^d\times S} \mathbf{c}(u, i, k, \pi_{\sigma_s}^\gamma)(g(u, k) - g(u, i))\, d\pi_{\sigma_s}^\gamma(u, i) + M_t^{g,\gamma}. \end{aligned} \qquad (21)$$

The representation (21) shows that a change in the value of the Markov chain, $\langle \pi_{\sigma_t}^\gamma, g\rangle - \langle \pi_{\sigma_0}^\gamma, g\rangle$, can be decomposed into two parts: a change from the generator, which is sometime referred to as a reaction part (the second expression on the right hand side of (21)) and a change from an unbiased random walk (the martingale term, the third expression on the right hand side of (21)). As $\gamma \to 0$, one can usually prove that the reaction part

$$
\begin{tikzcd}
\text{Macroscopic Proc. } \eta_t \arrow[r, "\mathit{approximate}", "\mathit{Corollary}\ 4"'] & \text{Mean ODE } \rho_t(i) \\
\text{Microscopic Proc. } \sigma_t \arrow[u, "\mathit{aggregate}", "\mathit{Theorem}\ 3"'] \arrow[r, "\mathit{approximate}", "\mathit{Theorems}\ 1,\ 2"'] & \text{Mesoscopic Equation } f_t(u,i) \arrow[u, "\mathit{aggregate}"']
\end{tikzcd}
$$

FIGURE 3. Relationships between the stochastic process and the deterministic approximation.

gives a law of motion of a deterministic dynamic and that the martingale term vanishes. Thus, if $\pi^{\gamma}_{\sigma_t} \to f_t m$ as $\gamma \to 0$, then (21) becomes

$$
\langle f_t, g\rangle = \langle f_0, g\rangle + \int_0^t ds \sum_{k\in S} \int_{\mathbf{T}^d \times S} \mathbf{c}(u, i, k, f_s)(g(u, k) - g(u, i)) f_s(u, i)\, dm(u, i),
$$

and upon differentiating with respect to time, we find

$$
\left\langle \frac{\partial f_t}{\partial t}, g\right\rangle = \sum_{k\in S} \int_{\mathbf{T}^d \times S} \mathbf{c}(u, i, k, f_t)(g(u, k) - g(u, i)) f_t(u, i)\, dm(u, i). \tag{22}
$$

Equation (22) is equivalent to the IDE (15), which is referred to as the weak formulation, and can be obtained by integrating (18) over u and i.

The proof of Theorems 1 and 2 (see the Appendix) is a variation on the proof given in Comets (1987), Kipnis and Landim (1999), and Katsoulakis et al. (2005). Unlike these studies, we do not assume that the spatial stochastic processes are reversible. Rather, we identify the general conditions under which scaling limits hold. Specifically, by using the martingale representation (21), we show that $\{\mathbf{Q}^{\gamma}\}_{\gamma}$, a sequence of probability laws of $\{\pi^{\gamma}_{\sigma_t}\}_{\gamma}$, is relatively compact. We then show that all limit points are concentrated on weak solutions to (18) and on measures that are absolutely continuous with respect to the Lebesgue measure. Finally, we demonstrate that weak solutions to (18) are unique and, thus, we conclude the convergence of $\mathbf{Q}^{\gamma}$ to the Dirac measure concentrated on solutions to (18).

3.5 *Spatially uniform interactions: Mean dynamics*

In this section, we show that under the assumption of uniform interactions, a spatially aggregated process is still a Markov chain (such a process is called *lumpable*). Furthermore, as expected, our IDE's then reduce to usual ODE's in evolutionary game theory. Figure 3 shows the relationships between various processes and differential equations. We take periodic boundary conditions and uniform interactions, that is, $\mathcal{J} := 1$ on $\mathbf{T}^d$, where 1 denotes a constant function with the value of 1 on $\mathbf{T}^d$. We define an aggregation variable

$$
\eta^{\gamma}(i) := \frac{1}{|\mathbf{T}^{d,\gamma}|} \sum_{x\in \mathbf{T}^{d,\gamma}} \delta(\sigma(x), i)
$$

that gives the empirical population proportion of agents with strategy i in the entire domain $\mathbf{T}^{d,\gamma}$. Suppose that there are n^d agents in the population. Thus, $|T^{d,\gamma}| = n^d = \gamma^{-d}$. Note that this can be obtained, equivalently, by integrating the empirical measure π_σ^γ over the spatial domain $\mathbf{T}^d$. Furthermore, because $\mathcal{J} = 1$, the payoff $u(x, \sigma, k)$ depends on σ only through the aggregated variable $\eta^n(i)$. Indeed, we have

$$u(x, \sigma, k) := \frac{1}{n^d} \sum_{y \in \mathbf{T}^{d,n}} \sum_{l \in S} \delta(\sigma(y), l) a(k, l) = \sum_{i \in S} a(k, i) \eta^n(i).$$

Thus, for the strategy revision rate, if $\sigma(x) = j$, then we define $c^M(j, k, \eta^n) := c^\gamma(x, \sigma, k)$, because the right hand side is independent of x and depends only on σ through the corresponding aggregate variable η^n. So $\{\eta_t^n\}_t$ itself is a Markov process, as we show below in Theorem 3, and the state space for η_t^n is the discrete simplex $\Delta^n = \{\{\eta(i)\}_{i \in S}; \sum_{i \in S} \eta(i) = 1, n^d \eta(i) \in \mathbb{N}_+\}$. To capture the transition induced by an agent's strategy switching, we write $\eta^{n,j,k}(i) = \eta^n(i)$ if $i \neq k, j$, $\eta^{n,j,k}(i) = \eta^n(i) - 1/n^d$ if $i = j$, and $\eta^{n,j,k}(i) = \eta^n(i) + 1/n^d$ if $i = k$. Thus, $\eta^{n,j,k}$ is the state obtained from η^n if one agent switches his strategy from j to k.

Theorem 3. *Suppose that the interaction is uniform. Then η_t^n is a Markov chain over the state space Δ^n and its generator is*

$$L^{M,n} g(\eta) = \sum_{k \in S} \sum_{j \in S} n^d \eta^n(j) c(j, k, \eta)(g(\eta^{n,j,k}) - g(\eta^n)). \tag{23}$$

The factor n^d in (23) derives from the fact that in a time interval of size 1, on average, n^d agents switch their strategies, and among those, $n^d \eta^n(j)$ agents switch from strategy j. Theorem 3 shows that the stochastic process with uniform interactions coincides with the multitype birth and death process in population dynamics (Blume 1997, Benaim and Weibull 2003). Furthermore, at the mesoscopic level, IDE's directly reduce to ODE's as follows (see Figure 3). Note that when $\mathcal{J} = 1$, we can define

$$\rho(i) := \int f(u, i)\, du = \mathcal{J} * f(i)$$

such that $\mathbf{c}(u, k, i, f)$ is independent of u and this again allows for the definition $\mathbf{c}^M(k, i, \rho) := \mathbf{c}(u, k, i, f)$. Thus, from the IDE (15), we immediately obtain

$$\frac{d\rho_t(i)}{dt} = \sum_{k \in S} \mathbf{c}^M(k, i, \rho) \rho_t(k) - \rho_t(i) \sum_{k \in S} \mathbf{c}^M(i, k, \rho).$$

Well known mean ODE's such as replicator dynamics, logit dynamics, and Smith dynamics can be derived by choosing F appropriately. Finally, as a consequence of Theorem 1, we have the following corollary, which can be compared to the continuous-time model of Benaim and Weibull (2003) (see Section 6 and Appendix I in Benaim and Weibull 2003). To state the result, we write $\|\eta^n\|_u := \sup_{i \in S} |\eta^n(i)|$.

Corollary 4 (Uniform interaction (Benaim and Weibull 2003)). *Suppose that the interaction is uniform and that the strategy revision rate satisfies C1–C3. Further suppose that there exists $\rho \in \Delta$ such that the initial condition η_0^n satisfies $\lim_{n\to\infty} \eta_0^n = \rho$ in probability. Then for every $T > 0$,*

$$\lim_{n\to\infty} \eta_t^n(i) \longrightarrow \rho_t(i) \quad \textit{in probability}$$

uniformly for $t \in [0, T]$ and $\rho_t(i)$ satisfies the differential equation

$$\begin{aligned} \frac{d\rho_t(i)}{dt} &= \sum_{k\in S} \mathbf{c}^M(k, i, \rho)\rho_t(k) - \rho_t(i) \sum_{k\in S} \mathbf{c}^M(i, k, \rho) \\ \rho_0(i) &= \rho(i) \end{aligned} \tag{24}$$

for $i \in S$, where $\mathbf{c}^M(k, i, \rho) := \mathbf{c}(u, k, i, f)$. Moreover, there exist C and ϵ_0 such that for all $\epsilon \le \epsilon_0$, there exists n_0 such that for all $n \ge n_0$,

$$\Pr\Big\{\sup_{t\le T} \|\eta_t^n - \rho_t\|_u \ge \epsilon\Big\} \le 2|S|e^{-n^d\epsilon^2/(TC)}. \tag{25}$$

Estimates such as (25) describe the validity regimes of the approximation by uniform interaction models (24) in terms of both the agent number n and the time window $[0, T]$. Observe that the bound in (25) is essentially the same as that in Lemma 1 in Benaim and Weibull (2003). That is, the bound increases linearly as the number of strategies increases, and it decreases exponentially as the size of the system (n^d) and the deviation (ϵ^2) increase.

4. Equilibrium selection

4.1 *Linear stability analysis*

In this section, we present linear stability analysis of IDE's around stationary solutions. The linearization of IDE's around stationary solutions is a widely used technique for analyzing nonlinear IDE's, as in the case of ODE's (see Murray 1989, Fisher and Marsden 1975, Collet and Eckmann 1990, Fife 1979). For example, if all eigenvalues for a linearized system have a negative real part, then one can show that the stationary solution to the nonlinear equation is stable. Furthermore, if the linearized system around the stationary solution is hyperbolic (i.e., no eigenvalue has a zero real part), then one can analyze the local behavior of the nonlinear equation around the stationary solution by constructing stable and unstable manifolds. We do not provide a precise statement for the Hartman–Gorbman type theorem that relates the linearized equations to original nonlinear equations. Such a theorem can be proved by using standard methods on a case-by-case basis.

At a deeper level the linear stability analysis is the first step toward understanding the generation and propagation of spatial structures. As in the case of the example in Section 2, traveling front solutions are constructed by joining two stable and spatially homogeneous stationary solutions to the linearized system, and the exponential convergence to such a front solution can be proved by analyzing the spectrum of the linear operator obtained from the original nonlinear system (Bates et al. 1997; Theorem 1.1 in Chapter 5 in Volpert et al. 1994).

This linearization also allows for an analysis of bifurcations in the system when the nature of eigenvalues for the linearized system changes. The appearance of eigenvalues with a positive real part indicates the instability of the system, and this instability often leads to the formation of complex spatial structures such as patterns. For example, Vickers et al. (1993) show the existence of patterns—spatially inhomogeneous stationary solutions—by analyzing bifurcations from eigenvalues for a linearized system (Theorems 3.1 and 3.2 in Vickers et al. 1993). Several such examples are demonstrated at length in Murray (1989).

To simplify the linear stability analysis, in the remaining part of this section, we consider a general type of integro-differential equation that incorporates logit and replicator equations as special cases. We then determine explicit solutions for a linearized system of general equations and apply the result to specific equations in Section 4.2. Consider the type of integro-differential equations

$$
\begin{aligned}
\frac{\partial f}{\partial t} &= \Phi(\mathcal{J} * f, f) \quad \text{in } \Lambda \times (0, T] \\
f(0, x) &= f^0(x) \quad \text{on } \Lambda \times \{0\},
\end{aligned}
\tag{26}
$$

where $\Lambda \subseteq \mathbb{R}^d$ or $\Lambda = \mathbf{T}^d$, $f \in \mathcal{M}(\Lambda \times S)$, $\mathcal{J} * f := (\mathcal{J} * f_1, \mathcal{J} * f_2, \ldots, \mathcal{J} * f_{|S|})^T$, and $\Phi : \mathbb{R}^{|S|} \times \mathbb{R}^{|S|} \to \mathbb{R}^{|S|}$, $\Phi(r, s)$ is smooth in both arguments, where r and s are variables for $\mathcal{J} * p$ and p, respectively, and T denotes the transpose operation. Here Φ is a vector-valued function taking two vectors as arguments. The first vector argument $\mathcal{J} * f$ is a collection of spatially weighted densities of each strategy, $\mathcal{J} * f_1, \mathcal{J} * f_2, \ldots, \mathcal{J} * f_{|S|}$, and the second vector argument f is a collection of densities of each strategy, $f_1, f_2, \ldots, f_{|S|}$. Note that replicator and logit equations can be written in the form of (26) by choosing Φ appropriately.

Observe that if f is spatially homogeneous, that is, $f(u, t) = f(t)$, then $\mathcal{J} * f = f(\mathcal{J} * 1) = f$, where 1 again denotes a constant function with the value of 1 on Λ. Thus, the IDE (26) reduces to the ODE $\partial f / \partial t = \Phi(f, f)$. In turn, this ODE is identical to the one obtained when the interaction is uniform ($\mathcal{J} = 1$). This shows that spatially homogeneous solutions to (26) are precisely the stationary solutions to the corresponding mean ODE. In particular, every spatially homogeneous stationary solution f_0 satisfies $\Phi(f_0, f_0) = 0$. We record this observation in Lemma 1.

Lemma 1 (Spatially homogeneous stationary solutions). *The constant function f_0 is a spatially homogeneous stationary solution to (26) if and only if $\Phi(f_0, f_0) = 0$.*

We now examine perturbations by linearizing around a spatially homogeneous stationary solution, f_0. Let $f = f_0 + \epsilon Z$, where $Z = Z(u, i)$. Substituting this into (26), we obtain

$$
\epsilon \frac{\partial Z}{\partial t} = \Phi(f_0 + \epsilon \mathcal{J} * Z, f_0 + \epsilon Z). \tag{27}
$$

We expand the right hand side of (27) and obtain

$$\Phi(f_0+\epsilon\mathcal{J}*Z, f_0+\epsilon Z)=\Phi(f_0,f_0)+\epsilon[D_r\Phi(f_0,f_0)\mathcal{J}*Z+D_s\Phi(f_0,f_0)Z]+O(\epsilon^2), \quad (28)$$

where D_r and D_s denote derivatives with respect to r and s, respectively, that is, $(D_r\Phi(f_0,f_0))_{i,j}=\partial\Phi_i/\partial r_j$, $(D_s\Phi(f_0,f_0))_{i,j}=\partial\Phi_i/\partial s_j$. Note that $\Phi(f_0,f_0)=0$. We substitute (28) into the right hand side of (27), divide each side by ϵ, and take $\epsilon\to 0$. Then we obtain

$$\frac{\partial Z}{\partial t}=D_r\Phi(f_0,f_0)\mathcal{J}*Z+D_s\Phi(f_0,f_0)Z. \quad (29)$$

This equation is a linear equation of the form $\partial Z/\partial t=AZ$, where A is a linear operator. One of the important properties of Fourier transformation is that it converts the complicated convolution operation into simple multiplication, which enables us to find the explicit solutions to (29) (for Fourier transformation, see, e.g., Stein and Shakarchi 2003). By applying the Fourier transformation to (29), we obtain an $|S|\times|S|$ matrix

$$D_r\Phi(f_0,f_0)\hat{\mathcal{J}}(k)+D_s\Phi(f_0,f_0) \quad (30)$$

for each $k\in\mathbb{Z}^d$, which is often called a frequency variable. Here $\hat{\mathcal{J}}(k):=\int_{\mathbf{T}^d}\mathcal{J}(u)e^{2\pi\mathbf{i}k\cdot u}\,du$ is called the Fourier coefficients of $\mathcal{J}$. We denote by $\lambda_j(k)$ the eigenvalues of the matrix (30) for $j=1,\dots,|S|$, $k\in\mathbb{Z}^d$, and denote by $Z_j(k)=e^{\mathbf{i}k\cdot x}Z_j$ the corresponding eigenfunction, where Z_j is some vector in $\mathbb{R}^{|S|}$ for each $j=1,\dots,|S|$. Then the general solution to (29) is the linear superposition of $e^{\lambda_j(k)t}e^{\mathbf{i}k\cdot x}Z_j$ over j and k (see (39) and Appendix A.5).

4.2 *Two-strategy symmetric games*

We consider the two-strategy symmetric games in Section 2. We call a game a coordination game if $a_{11}>a_{21}$ and $a_{22}>a_{12}$, and call it a Hawk–Dove type game if $a_{11}<a_{21}$ and $a_{22}<a_{12}$. If $p(u):=f(u,1)$, then from $f(u,1)+f(u,2)=1$, we can write a single equation for $p(u)$ and obtain equations of the form (26):

$$\text{replicator IDE}\quad \frac{dp}{dt}=(1-p)\mathcal{J}*pF_\kappa(\pi_1(\mathcal{J}*p)-\pi_2(\mathcal{J}*p)) \\ -p(1-\mathcal{J}*p)F_\kappa(\pi_2(\mathcal{J}*p)-\pi_1(\mathcal{J}*p)) \quad (31)$$

$$\text{logit IDE}\quad \frac{dp}{dt}=l_\beta(\pi_1(\mathcal{J}*p)-\pi_2(\mathcal{J}*p))-p, \quad (32)$$

where $l_\beta(t):=1/(1+\exp(-\beta t))$, $F_\kappa(t):=1/\kappa\log(\exp(\kappa t)+1)$ (recall (12)), $\pi_1(\mathcal{J}*p):=a_{11}\mathcal{J}*p+a_{12}(1-\mathcal{J}*p)$, and $\pi_2(\mathcal{J}*p):=a_{21}\mathcal{J}*p+a_{22}(1-\mathcal{J}*p)$. We write the replicator equation as $dp/dt=\Phi_R(\mathcal{J}*p,p)$ and the logit equation as $dp/dt=\Phi_L(\mathcal{J}*p,p)$.

We consider $[-\pi,\pi]^d$ for $d=1,2$ as a domain with the periodic boundary condition and $[-1,1]^d$ for $d=1,2$ as a domain with the fixed boundary condition. In addition to the conditions for $\mathcal{J}$ stated in Section 3.3, we assume that $\mathcal{J}$ is symmetric: $\mathcal{J}(x)=$

$\mathcal{J}(-x)$ for $x \in \Lambda$. Frequently, in examples and simulations, we consider the localized Gaussian-like kernel $\mathcal{J}(x) \propto \exp(-\chi\|x\|^2)$ for some $\chi > 0$. More specifically, we use

$$\text{periodic boundary:} \quad \mathcal{J}(x) = \frac{\exp(-\chi\|x\|^2)}{\int_{[-\pi,\pi]^d} \exp(-\chi\|z\|^2)\,dz} \quad \text{for } x \in [-\pi, \pi]^d \tag{33}$$

$$\text{fixed boundary:} \quad \mathcal{J}(x) = \frac{\exp(-\chi\|x\|^2)}{\int_{[-1,1]^d} \exp(-\chi\|z\|^2)\,dz} \quad \text{for } x \in [-1, 1]^d. \tag{34}$$

In the fixed boundary domain, we have $\Lambda = [-1, 1]^d$ and $\Gamma = [-2, 2]^d$. For the case of $d = 1$, an agent located at the center of the domain Λ (at 0) interacts with all other agents in Λ, whereas an agent at the endpoints of Λ (e.g., at -1) interacts with half of all agents in Λ (e.g., $[-1, 0]$) as well as half of all agents in $\partial\Lambda$ (e.g., $[-2, -1]$).

4.2.1 *Stationary solutions and their linear stability* To determine spatially homogeneous stationary solutions, we need to set $\Phi_R(p, p) = 0$ and $\Phi_L(p, p) = 0$. We let $\alpha := a_{11} - a_{21} + a_{22} - a_{12}$ and $\zeta := (a_{22} - a_{12})/\alpha$. Note that we have $\pi_1(p) - \pi_2(p) = \alpha(p - \zeta)$. Then, for a coordination or Hawk–Dove type game, we have $0 < \zeta < 1$. Consider the replicator equation. In this case, it is easy to see that $p = 0, 1$, and ζ are three spatially homogeneous stationary solutions. In the case of the logit equation, we recall that $l_\beta(\pi_1(p) - \pi_2(p))$ has the shape of a smoothed step function at $p = \zeta$. As β approaches zero, $l_\beta(\pi_1(p) - \pi_2(p))$ approaches a constant function with the value 1/2. As β approaches infinity, $l_\beta(\pi_1(p) - \pi_2(p))$ approaches a unit step function at $p = \zeta$. For the coordination game ($\alpha > 0$), $l_\beta(\pi_1(p) - \pi_2(p))$ is increasing in p. Thus, if β is small, then there is a unique p such that $l_\beta(\pi_1(p) - \pi_2(p)) = p$. If β is large, then there are three p's that satisfy the relationship $l_\beta(\pi_1(p) - \pi_2(p)) = p$. For the Hawk–Dove type game ($\alpha < 0$), $l_\beta(\pi_1(p) - \pi_2(p))$ is decreasing in p. As a result, there is a unique p that satisfies $l_\beta(\pi_1(p) - \pi_2(p)) = p$ for all $\beta > 0$. We summarize these observations in the following proposition.

Proposition 2 (Stationary solutions). *(i) Consider replicator dynamics. Then $p(u) = 0$, $p(u) = 1$, and $p(u) = \zeta$ for all u are stationary solutions.*

(ii) Consider logit dynamics. Assume a coordination game. Then there exists β_C such that for $\beta < \beta_C$, there is one spatially homogeneous stationary solution p_1, and for $\beta > \beta_C$, there are three spatially homogeneous stationary solutions p_1, p_2, and p_3.

(iii) Again consider the logit dynamics. Assume a Hawk–Dove type game. Then there is a unique spatially homogeneous stationary solution.

We now examine the linear stability of these stationary solutions. By differentiating Φ_R and Φ_L, and using (30), we find the expression for eigenvalues (the so-called *dispersion relations*) for the replicator IDE (Table 1). Note that by the assumptions for $\mathcal{J}$ (i.e., the symmetry of $\mathcal{J}$ and $\int \mathcal{J}(u)\,du = 1$), $\hat{\mathcal{J}}(k)$ is real-valued and $|\hat{\mathcal{J}}(k)| < 1$ for all k. We also note that the Gaussian kernel satisfies the hypothesis $0 < \hat{\mathcal{J}}(k)$ for all k. Using these facts, we obtain the first part of Proposition 3. Because $\alpha < 0$ in the Hawk–Dove type game, if $\hat{\mathcal{J}}(k) \geq 0$, then $\lambda_R(k)$ is negative for sufficiently large κ.

$p=0$	$\lambda_R(k)=F_\kappa(-\alpha\zeta)\hat{\mathcal{J}}(k)-F_\kappa(\alpha\zeta)$
$p=1$	$\lambda_R(k)=F_\kappa(\alpha(\zeta-1))\hat{\mathcal{J}}(k)-F_\kappa(\alpha(1-\zeta))$
$p=\zeta$	$\lambda_R(k)=(\log(2)/\kappa+\alpha\zeta(1-\zeta))\hat{\mathcal{J}}(k)-\log(2)/\kappa$

TABLE 1. Eigenvalues for the replicator IDE.

PROPOSITION 3 (Linear stability for the replicator IDE). *We have the following results.*

(i) The equality $p=0,1$ is linearly stable for replicator dynamics for coordination games.

(ii) The equality $p=\zeta$ is linearly stable for the replicator dynamics for Hawk–Dove type games.

In the case of logit dynamics, we note that $l'_\beta(t)=\beta l_\beta(t)(1-l_\beta(t))$ and, thus, obtain eigenvalues for any stationary solution p:

$$\lambda_L(k)=\beta\alpha(1-p)p\hat{\mathcal{J}}(k)-1, \quad k\in\mathbb{Z}^d. \tag{35}$$

PROPOSITION 4 (Linear stability for the logit IDE). *We have the following results.*

(i) Suppose a coordination game. If $\beta<\beta_C$, then the unique stationary solution p_0 is linearly stable. If $\beta>\beta_C$, then two stationary solutions p_1 and p_3 are linearly stable, where $p_3<p_2<p_1$.

(ii) For a Hawk–Dove type game, the unique stationary solution p_0 is linearly stable.

The linear stability results for spatially homogeneous stationary solutions are consistent with the corresponding results for stationary solutions to ODE equations. In addition, if we consider the linearization of the replicator equation ($k=\infty$) around ζ, then from Table 1, we can find that $\lambda_R(k)=\alpha\zeta(1-\zeta)\hat{\mathcal{J}}(k)$, and by using (35), we find that $\lambda_L(k)=\beta\lambda_R(k)-1$, which shows that the eigenvalues for the linearized replicator and logit equations at their interior equilibria differ only by a positive affine transformation (see Hopkins 1999 and Sandholm 2010b, pp. 298–299). This shows that the important and interesting relationship between the linearizations of replicator and logit equations holds at the level of the mesoscopic equations as well.

The logit equation has a close relationship with the Glauber equation, which is a well known equation in statistical mechanics (De Masi et al. 1994, Katsoulakis and Souganidis 1997, Presutti 2009). Using $l_\beta(z)=1/2+1/2\tanh(\beta z/2)$ and changing the variable by $p\mapsto 2p-1:=w$, we derive the following equation from the logit equation:

$$\frac{\partial w}{\partial t}=-w+\tanh\Big(\beta\tfrac{1}{4}\alpha(\mathcal{J}*w+1-2\zeta)\Big), \tag{36}$$

which is a Glauber mesoscopic equation with β interpreted as the *inverse temperature*. All known results for (36), such as the existence of traveling front solutions in one-dimensional space and the geometric evolution of the interface between homogeneous stationary states in higher dimensions, are applicable to logit dynamics. In addition, we show directly the existence and stability of traveling front solutions to the logit equation in the next section. Brock and Durlauf (1999, 2001) also use a discrete time version of (36) to analyze social interactions and they examine the logit choice rules for individuals by incorporating various terms that reflect social factors (e.g., conformism with the behavior of others).

4.2.2 *Imitation versus perturbed best responses: Equilibrium selection through traveling front solutions* To examine traveling front solutions, we suppose that the domain is a subset of $\mathbb{R}$ with the fixed boundary condition or the whole real line $\mathbb{R}$. A solution is called a traveling front (or wave) solution if it moves at a constant speed; thus, a traveling front solution $p(x,t)$ can be written as $P(x-ct)$ for some constant c and some function P that satisfies $\lim_{\xi\to\infty} P(\xi)=1$, $\lim_{\xi\to\infty} P(\xi)=0$ (see Figure 1). Here, if the speed of the front solution, c, is negative, then the front solution travels toward the left, indicating that the choice of strategy 1 propagates over the whole domain. If the speed c is zero, then the front solution is called a standing front solution or an instanton.

A rigorous analysis of such traveling front solutions is an important and popular topic in mathematical physics and evolutionary biology. For example, De Masi et al. (1995), and Orlandi and Triolo (1997) show the existence of traveling front solutions to Glauber equations, which implies the existence of such solutions for the logit equations. However, no rigorous results exist for the replicator IDE (31). Hutson and Vickers (1992), Hofbauer et al. (1997), Hofbauer (1997), and Hofbauer (1999) examine traveling front solutions and the problem of equilibrium selection for (bistable) reaction diffusion equations, which are similar to replicator IDE's. Chen (1997) provides fairly general conditions for the existence and stability of traveling front solutions to IDE's similar to (31) and (32), and using his results, we show that replicator IDE's as well as logit IDE's admit traveling front solutions.

To state Chen's result, we again consider the general IDE $dp/dt=\Phi(\mathcal{J}*p,p)$, and denote by r and s the first and second arguments for Φ. In addition, we assume that $h(p):=\Phi(p,p)$ satisfies

$$
\begin{aligned}
&h(p)>0, \quad p\in(a,1); \qquad h(p)<0, \quad p\in(0,a)\\
&h'(0)<0, \qquad h'(1)<0, \qquad h'(a)>0.
\end{aligned}
\tag{37}
$$

The conditions (37) require that the ODE, specified by $dp/dt=\Phi(p,p)$, has two stable stationary states at 0 and 1, and an unstable interior state at a. For a coordination game, these conditions are satisfied.

Theorem 5 (Chen 1997). *Suppose that Φ is smooth, $\partial\Phi/\partial r>0$, and $\partial\Phi/\partial s<0$. Then $dp/dt=\Phi(\mathcal{J}*p,p)$ admits a unique traveling front solution. Furthermore, the traveling front solution is asymptotically stable in the sense that any initial data p_0 that satisfy*

$$
\limsup_{x\to-\infty} p_0(x) < a < \liminf_{x\to\infty} p_0(x) \tag{38}
$$

converge exponentially to the traveling front solution.

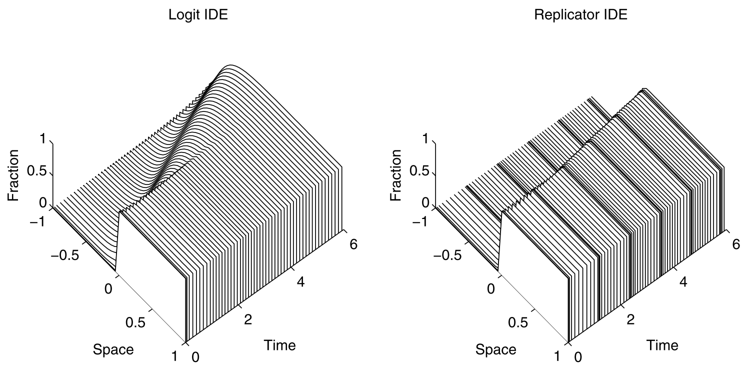


FIGURE 4. Comparison of traveling fronts between replicator and logit IDE's. We reproduce the right panel of Figure 1 in the left panel for comparison. The left panel shows the time path of the density of strategy 1 in logit dynamics, whereas the right panel shows the case of replicator dynamics: $N = 256$, $\Lambda = [-1, 1]$, $dt = 0.001/(0.05N^2)$, $a_{11} = 20/3$, $a_{22} = 10/3$, and $a_{12} = a_{21} = 0$; $\chi = 2$ for the Gaussian kernel. The initial condition is a unit step function at 0.

Then it is easy to see that the replicator IDE specified by Φ_R and the logit IDE specified by Φ_L at sufficiently high β satisfy the hypotheses in Theorem 5. This shows that the replicator and logit IDE's admit the unique and asymptotically stable traveling front solution described in Section 2.

The speed of the traveling front solutions is important too. The sign of the speed directly determines equilibrium selection, as discussed earlier. Orlandi and Triolo (1997) show that the speed of the traveling front solution is negative if $1 - 2\zeta > 0$ in the Glauber equation (36). This indicates that if $a_{11} - a_{21} > a_{22} - a_{12}$, then the equilibrium of strategy 1 is selected, whereas strategy 2 is driven out. This implies that, like other equilibrium selection models, logit IDE's select the risk-dominant equilibrium. A similar direct characterization of the sign of the speed of the replicator equation is not readily available. However, our numerical simulations (e.g., Figure 4) suggest that a similar characterization would hold for replicator equations. Which behavioral rules (imitating vs. best-response behaviors) select the risk-dominant equilibrium in the coordination game faster? To compare the speed of traveling front solutions, we conduct informal, but illuminating, numerical comparisons. In the literature studying the traveling front solutions to equations similar to (31) and (32), it is known that the speed of a traveling front solution is proportional to the mean curvature of the shape of the front solution (see Katsoulakis and Souganidis 1997, Carr and Pego 1989). This means that the sharper is the shape of the traveling front solution, the slower is the propagation of the interface. Therefore, a sharper solution shape implies a less diffusive system. To compare such shapes, we present the shapes of standing front solutions to the replicator and logit equations in Figure 5.

As shown in Figure 5, the shape of the standing wave in replicator dynamics is considerably sharper than that for logit dynamics. This implies that the speed of the traveling front solution for replicator dynamics is considerably slower than that for logit

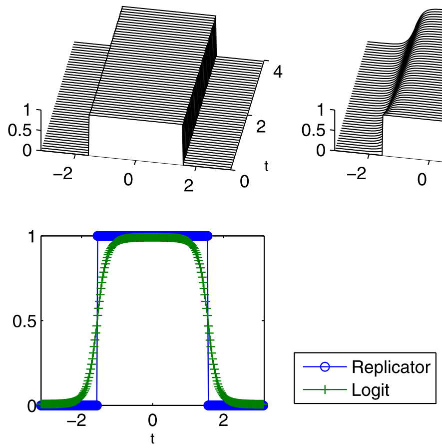


FIGURE 5. Comparison of standing waves between replicator and logit dynamics ($a_{11} = a_{22}$; the periodic boundary condition). The top left panel shows the time evolution of the population density of strategy 1 in replicator dynamics, whereas the top right panel describes the case of logit dynamics. The bottom panel shows the shapes of standing waves for both cases at time 4. We consider the replicator with $\kappa = \infty$, $N = 256$, $\Lambda = [-\pi, \pi]$, $dt = 0.001/(0.25N^2)$, $a_{11} = 5$, $a_{22} = 5$, and $a_{12} = a_{21} = 0$; $\chi = 2$ for the Gaussian kernel. The initial condition is $\mathbf{1}_{[-\pi/2, \pi/2]}$.

equations; Figure 4 illustrates these observations, demonstrating that when agents imitate other's behaviors instead of playing the best response, the propagation of a risk-dominant equilibrium is slow. We present this comparison from different perspectives in the next section.

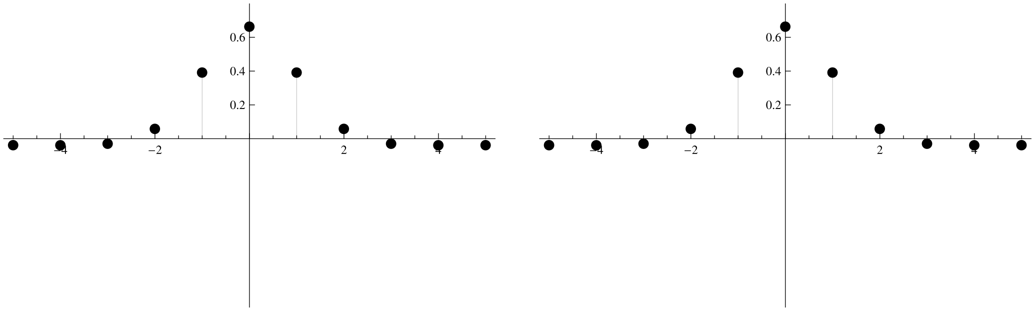
5. Pattern formation

In this final section, we explain the formation of patterns for replicator and logit dynamics in the case of coordination games. Here, the analysis is heuristic and informal; we leave the rigorous treatment of these issues for future research. Figure 6 shows an example of the dispersion relations (eigenvalues) for the replicator and logit equations at $p = \zeta$. Observe that in the case of two-strategy games and one- dimensional space, the solution to the linearized system can be expressed as

$$\sum_{k=-\infty}^{\infty} c_k e^{\lambda(k)t} \cos(2\pi k x) \quad \text{or} \quad \sum_{k=-\infty}^{\infty} d_k e^{\lambda(k)t} \sin(2\pi k x), \tag{39}$$

where c_k and d_k are some constants. Here the frequency variable k determines the period of the sine and cosine functions, and is often called the *mode*. We first consider the replicator equation. Notice that in the left panel of Figure 6, $\lambda_R(k) > 0$ if $k = 0, \pm 1$, and ± 2, and $\lambda_R(k) < 0$ otherwise. So, as t increases, the following three terms in each sum of (39),

$$\begin{array}{lll} c_0 e^{\lambda(0)t}, & c_1 e^{\lambda(1)t} \cos(2\pi x), & c_2 e^{\lambda(2)t} \cos(4\pi x) \\ d_0 e^{\lambda(0)t}, & d_1 e^{\lambda(1)t} \sin(2\pi x), & d_2 e^{\lambda(2)t} \sin(4\pi x), \end{array}$$



(a) Replicator (b) Logit

FIGURE 6. Eigenvalues for the mixed strategy equilibrium. The figure shows the dispersion relations $\lambda_R(k)$ and $\lambda_L(k)$ at $p = \zeta$, $\chi = 20$ for the Gaussian kernel, $\kappa = 20$, $\alpha = 3$, $\zeta = 1/3$, and $\beta = 1$.

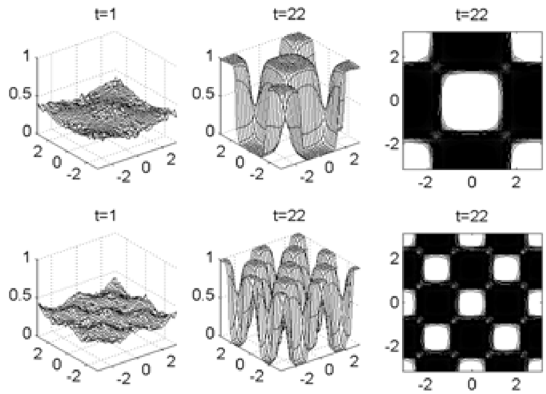


FIGURE 7. Pattern formation in replicator dynamics. The left and middle panels show the time evolution of population densities for strategy 1 in the spatial domain $T^d = [-\pi, \pi]^2$. The number of nodes is 64 and the time step is 0.0175. The initial conditions are $1/3 + \text{rand} \cos(2\pi x)\cos(2\pi y)$ (top panel) and $1/3 + \text{rand} \cos(4\pi x)\cos(4\pi y)$ (bottom panel), where rand denotes the realization of the uniform random variable [0, 1] at each node: $a_{11} = 2/3$, $a_{22} = 1/3$, and $a_{12} = a_{21} = 0$; $\chi = 15$ for the Gaussian kernel. The right panels show the contour maps of densities at $t = 22$.

increase exponentially and dominate other terms with negative $\lambda_R(k)$. If linear solutions approximate well the solution to the original nonlinear equation, then there may be nonlinear solutions that reflect the shape of periodic functions. In addition, such nonlinear solutions may be spatially inhomogeneous stationary solutions (for a detailed explanation, see Murray 1989). This is how we generate the patterns for replicator equations (Figures 7 and 8). In these simulations, we use the initial data obtained by perturbing the unstable mixed strategy periodically. That is, in the top panel of the coarse pattern, we use $p_0(x, y) = \zeta + \epsilon(x, y)\cos(2\pi x)\cos(2\pi y)$, where $\epsilon(x, y)$ is the realized value of a uniform random variable from [0, 0.1] at each spatial location (x, y). Alternatively, to generate a fine pattern in the bottom panel, we use $p_0(x, y) = \zeta + \epsilon(x, y)\cos(4\pi x)\cos(4\pi y)$.

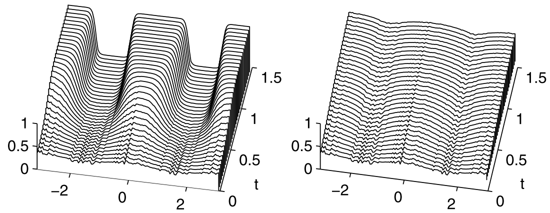


FIGURE 8. Replicator versus logit dynamics (periodic boundary conditions). The left panel shows the population density of strategy 1 for replicator IDE's with $\kappa = \infty$ and the right panel shows the population density for logit dynamics: $N = 512$, $\Lambda = [-\pi, \pi]$ with the periodic boundary condition, $dt = 0.001/(0.05N^2)$, $a_{11} = 20/3$, $a_{22} = 10/3$, and $a_{12} = a_{21} = 0$; $\chi = 10$ for the Gaussian kernel.The initial condition is $1/2 + 1/10 \,\mathrm{rand} \cos(2\pi x)$, where rand denotes the realization of a uniform random variable at each node.

As shown in the right panel of Figure 6, the dispersion relation (eigenvalues) for logit dynamics under exactly the same parameter values as replicator dynamics shows that $\lambda_L(k) < 0$ for all k, indicating that the logit dynamics may not develop a pattern under the same condition as the replicator dynamics. This conjecture is illustrated in Figure 8, which compares the replicator and logit equations. For a coordination game, both equations admit three stationary states: two stable states near the boundaries and an unstable interior state. Therefore, if we examine only the population aggregate, ignoring the spatial interactions, we may conclude that the two dynamics are indistinguishable from each other. Figure 8 contrasts the imitation behaviors with perturbed best-response behaviors in the spatial domain, demonstrating that such a conclusion may be misleading. As shown in Figure 8, imitation behaviors develop a spatial pattern, whereas perturbed best-response behaviors lead to the rapid convergence to strategy 1. This may be because the mixed strategy state in imitation dynamics is more destabilizing than that in perturbed best-response dynamics (see Figure 6). Note that the time evolution of the replicator IDE in the left panel of Figure 8 corresponds to a one-dimensional snapshot of the pattern in the two-dimensional replicator systems in Figure 7.

APPENDIX

A.1 *Proof of Proposition 1*

We set

$$B(f,u) := \frac{e^{\beta(a_{11}\mathcal{J}*f(u)+a_{12}(1-\mathcal{J}*f(u)))}}{e^{\beta(a_{11}\mathcal{J}*f(u)+a_{12}(1-\mathcal{J}*f(u)))} + e^{\beta(a_{21}\mathcal{J}*f(u)+a_{22}(1-\mathcal{J}*f(u))))}}$$

and note that $\phi'(p) = \log(p/(1-p))$. Thus, we have

$$\phi'(B(f,u)) = \beta(a_{11} - a_{21} + a_{22} - a_{12})\mathcal{J} * f(u) - \beta(a_{22} - a_{12}).$$

From the symmetry of $\mathcal{J}$, we have $\partial\langle \mathcal{J} * f_t, f_t\rangle/\partial t = 2\langle \mathcal{J} * f_t, \partial f_t/\partial t\rangle$, where $\langle f, g\rangle := \int f(u)g(u)\, du$ and, thus,

$$\begin{aligned}\frac{dE(f)}{dt} &= \frac{1}{\beta}\int_T \big[\beta(a_{11} - a_{21} + a_{22} - a_{12})\mathcal{J} * f_t(u) - \beta(a_{22} - a_{12}) - \phi'(f(u))\big]\frac{\partial f_t}{\partial t}(u)\, du \\ &= \frac{1}{\beta}\int_T [\phi'(B(f, u) - \phi'(f_t(u))][B(f, u) - f_t(u)]\, du \geq 0.\end{aligned}$$

We use the fact that $\phi''(p) \geq m > 0$ in the last line, and the equality holds if and only if the solution is stationary, that is, $\partial f_t/\partial t = 0$.

A.2 *Various strategy revision rates and the Proof of Theorem 2*

Strategy revision rates We show that C1–C3 are satisfied for the strategy revision rates

- $c^\gamma(x, \sigma, k) = F(u(x, \sigma, k))$

  $$\mathbf{c}(u, i, k, f) = F\Big(\sum_l a(i, l)\mathcal{J} * f(u, l)\Big)$$

- $c^\gamma(x, \sigma, k) = F\big(u(x, \sigma, k) - u(x, \sigma, \sigma(x))\big)$

  $$\mathbf{c}(u, i, k, f) = F\Big(\sum_l [a(k, l) - a(i, l)]\mathcal{J} * f(u, l)\Big)$$

- $c^\gamma(x, \sigma, k) = \sum_y w(x, y, \sigma, k)F(u(x, \sigma, k))$

  $$\mathbf{c}(u, i, k, f) = \mathcal{J} * f(u, k)F\Big(\sum_l a(k, l)\mathcal{J} * f(u, l)\Big)$$

- $c^\gamma(x, \sigma, k) = \sum_y w(x, y, \sigma, k)F\big(u(x, \sigma, k) - u(x, \sigma, \sigma(x))\big)$

  $$\mathbf{c}(u, i, k, f) = \mathcal{J} * f(u, k)F\Big(\sum_l [a(k, l) - a(i, l)]\mathcal{J} * f(u, l)\Big)$$

- $$c^\gamma(x, \sigma, k) = \frac{\exp(u(x, \sigma, k))}{\sum_l \exp(u(x, \sigma, l))}$$

  $$\mathbf{c}(u, i, k, f) = \frac{\exp(\mathcal{J} * f(u, k))}{\sum_l \exp(\mathcal{J} * f(u, l))}.$$

If F satisfies the global Lipschitz condition, i.e., for all $x, y \in \mathrm{Dom}(F)$ there exists $L > 0$ such that $|F(x) - F(y)| \leq L|x - y|$. Here we abuse the notation $\mathbf{c}(u, i, k, \pi) = \mathbf{c}(u, i, k, f)$. Note that the above list is far from being exhaustive, because one can easily invent a number of other rates that satisfy C1–C3. Because the verification of conditions is similar, we check the conditions for the rate in the periodic boundary domain:

$$c^\gamma(x,\sigma,k) = F\bigg(\sum_{y\in\Lambda^\gamma}\gamma^d a(k,\sigma(y))\mathcal{J}(\gamma(y-x)) - \sum_{y\in\Lambda^\gamma}\gamma^d a(\sigma(x),\sigma(y))\mathcal{J}(\gamma(x-y))\bigg). \quad (40)$$

Lemma 2. *The rate given by (40) satisfies C1–C3.*

Proof. Let

$$\bar{c}^\gamma(u,i,k,\sigma) := F\bigg(\sum_{y\in\Lambda^\gamma}\gamma^d a(k,\sigma(y))\mathcal{J}(u-\gamma y) - \sum_{y\in\Lambda^\gamma}\gamma^d a(i,\sigma(y))\mathcal{J}(u-\gamma y)\bigg)$$

$$\mathbf{c}(u,i,k,\pi) := F\bigg(|\Gamma|\int_{\Lambda^\gamma\times S} a(k,l)\mathcal{J}(u-v)\,d\pi(v,l) - |\Gamma|\int_{\Lambda^\gamma\times S} a(i,l)\mathcal{J}(u-v)\,d\pi(v,l)\bigg),$$

where we associate u with γx and v with γy. Here we note that

$$\begin{aligned}
&\bigg|\sum_{y\in\Lambda^\gamma}\gamma^d a(k,\sigma(y))\mathcal{J}(\gamma x-\gamma y) - |\Gamma|\int_{\Lambda\times S} a(k,l)\mathcal{J}(\gamma x - v)\,d\pi^\gamma_\sigma(v,l)\bigg| \\
&\qquad \le \bigg|\gamma^d - \frac{|\Lambda|}{|\Lambda^\gamma|}\bigg|\sum_{y\in\Lambda^\gamma} a(k,\sigma(y))\mathcal{J}(\gamma x-\gamma y) \le |\gamma^d|\Lambda^\gamma| - |\Lambda||M \\
&\qquad \to 0 \quad \text{uniformly in } x, \sigma, k,
\end{aligned}$$

where $M := \sup_{i,k,u,v} a(i,j)\mathcal{J}(u,v)$. Therefore, by using the Lipschitz condition for F, we have

$$\begin{aligned}
&|c^\gamma(x,\sigma,k) - \mathbf{c}(\gamma x,\sigma(x),k,\pi^\gamma_\sigma)| \\
&\quad \le |\bar{c}^\gamma(\gamma x,\sigma(x),k,\sigma) - \mathbf{c}(\gamma x, v,\sigma(x),k,\pi^\gamma_\sigma)| \\
&\quad \le L \sup_{\substack{x\in\Lambda^\gamma\\ \sigma\in S^{\Lambda^\gamma}\\ k\in S}}\bigg|\sum_{y\in\Gamma^\gamma}\gamma^d a(k,\sigma(y))\mathcal{J}(\gamma x-\gamma y) - |\Gamma|\int_{\Gamma\times S} a(k,l)\mathcal{J}(\gamma x-v)\pi_\sigma(v,l)\bigg| \\
&\qquad + L \sup_{\substack{x\in\Lambda^\gamma\\ \sigma\in S^{\Lambda^\gamma}\\ k\in S}}\bigg|\sum_{y\in\Gamma^\gamma}\gamma^d a(\sigma(x),\sigma(y))\mathcal{J}(\gamma x-\gamma y) \\
&\qquad - |\Gamma|\int_{\Gamma\times S} a(\sigma(x),l)\mathcal{J}(\gamma x-v)\pi_\sigma(v,l)\bigg| \\
&\quad \to 0 \quad \text{uniformly in } x,\sigma,k.
\end{aligned}$$

Hence C1 is satisfied. Because $\mathbf{c}(u,i,k,\pi)$ is uniformly bounded, C2 is satisfied. Again, C3 follows from the fact that $\mathbf{c}(u,i,k,\pi)$ is uniformly bounded and F satisfies the Lipschitz condition. □

Notation We use the following notation for the proof of Theorems 1 and 2.

- The process $\{\Sigma_t^\gamma\}$ is a stochastic process that takes the value σ_t with the generator L^γ given in (17) and the sample space $D([0,T], S^{\Gamma^\gamma})$.
- The process $\{\Pi_t^\gamma\}$ is a stochastic process for an empirical measure that takes the value π_t with the sample space $D([0,T], \mathcal{P}(\Lambda \times S))$, and we denote by $\mathbf{Q}^\gamma$ the law of the process $\{\Pi_t^\gamma\}$ and denote by $\mathbf{P}$ the probability measure in the underlying probability space. The proof of Theorem 1 is very similar to that of Theorem 2; therefore, we only prove Theorem 2 and leave the modification needed for proving Theorem 1 to the reader.

Martingale estimates For $g \in C(\Gamma \times S)$, we set

$$h(\sigma) := \langle \pi_\sigma^\gamma, g\rangle = \frac{1}{|\Gamma^\gamma|} \sum_{y \in \Gamma^\gamma} g(\gamma y, \sigma(y)). \tag{41}$$

We define $M_t^{g,\gamma}$ and $\langle M_t^{g,\gamma}\rangle$ as follows. For $g \in C(\Gamma \times S)$,

$$\begin{aligned} M_t^{g,\gamma} &= \langle \Pi_t^\gamma, g\rangle - \langle \Pi_0^\gamma, g\rangle - \int_0^t L^\gamma \langle \Pi_s^\gamma, g\rangle ds, \\ \langle M_t^{g,\gamma}\rangle &= \int_0^t [L^\gamma \langle \Pi_s^\gamma, g\rangle^2 - 2\langle \Pi_s^\gamma, g\rangle L^\gamma \langle \Pi_s^\gamma, g\rangle]\, ds. \end{aligned} \tag{42}$$

Because h is measurable, $M_t^{g,\gamma}$ and $\langle M_t^{g,\gamma}\rangle$ are $\mathcal{F}_t$-martingale with respect to $\mathbf{P}$, where $\mathcal{F}_t$ is the filtration generated by $\{\Sigma_t\}$ (Ethier and Kurtz 1986, Darling and Norris 2008).

Lemma 3. *For $g \in C(\Gamma \times S)$, there exists C such that*

$$|L^\gamma \langle \pi^\gamma, g\rangle| \le C, \qquad |L^\gamma \langle \pi^\gamma, g\rangle^2 - 2\langle \pi^\gamma, g\rangle L^\gamma \langle \pi^\gamma, g\rangle| \le \gamma^d C.$$

Proof. For h in (41), we have

$$h(\sigma^{x,k}) - h(\sigma) = \frac{1}{|\Gamma^\gamma|}\big(g(\gamma x, k) - g(\gamma x, \sigma(x))\big)$$

and, thus, we have (43). Now let $q(\sigma) := \langle \pi_\sigma^\gamma, g\rangle^2$. Then

$$\begin{aligned} &q(\sigma^{x,k}) - q(\sigma) \\ &= \frac{1}{|\Gamma^\gamma|^2}\Big(\sum_{y\in\Lambda^\gamma} g(\gamma y, \sigma^{x,k}(y))\Big)^2 - \frac{1}{|\Gamma^\gamma|^2}\Big(\sum_{y\in\Lambda^\gamma} g(\gamma y, \sigma(y))\Big)^2 \\ &= \frac{1}{|\Gamma^\gamma|^2}\big(g(\gamma x, k) - g(\gamma x, \sigma(x))\big)^2 + \frac{2}{|\Gamma^\gamma|^2}\big(g(\gamma x, k) - g(\gamma x, \sigma(x))\big) \sum_{y\in\Lambda^\gamma} g(\gamma y, \sigma(y)). \end{aligned}$$

Thus, we have

$$L^\gamma\langle\pi^\gamma, g\rangle = \frac{1}{|\Gamma^\gamma|}\sum_{k\in S}\sum_{x\in\Lambda^\gamma} c^\gamma(x,\sigma(x),k)\big(g(\gamma x,k)-g(\gamma x,\sigma(x))\big) \tag{43}$$

$$\begin{aligned}&L^\gamma\langle\pi^\gamma, g\rangle^2 - 2\langle\pi^\gamma, g\rangle L^\gamma\langle\pi^\gamma, g\rangle\\ &\qquad = \frac{1}{|\Gamma^\gamma|^2}\sum_{k\in S}\sum_{x\in\Lambda^\gamma} c^\gamma(x,\sigma(x),k)\big(g(\gamma x,k)-g(\gamma x,\sigma(x))\big)^2.\end{aligned}$$

Thus, from C1 and C2, we have $|\Gamma^\gamma|\approx|\Gamma|\gamma^{-d}$ and $|\Lambda^\gamma|\approx|\Lambda|\gamma^{-d}$, and the results follow. □

Proposition 5. *Let $g\in C(\Gamma\times S)$ and take τ^γ and δ^γ as follows.*

(a) The variable τ^γ is the stopping time on the process $\{\Pi^\gamma_t : 0\le t\le T\}$ with respect to the filtration $\mathcal{F}_t$.

(b) The variable δ^γ is a constant for which $0\le\delta^\gamma\le T$ and $\delta^\gamma\to 0$ as $\gamma\to 0$. Then for $\epsilon>0$, there exists C such that

$$\text{(i)}\quad \mathbf{P}\Big\{\omega : \sup_{t\in[0,T]}|M^{g,\gamma}_t|\ge\epsilon\Big\}\le\frac{\gamma^d CT}{\epsilon^2}\quad\text{and}$$

$$\text{(ii)}\quad \mathbf{P}\{\omega : |M^{g,\gamma}_{\tau^\gamma+\delta^\gamma}-M^{g,\gamma}_{\tau^\gamma}|\ge\epsilon\}\le\frac{\gamma^d C\delta^\gamma}{\epsilon^2},$$

and there exists γ_0 such that for $\gamma<\gamma_0$,

$$\text{(iii)}\quad \mathbf{P}\Big\{\omega : \Big|\int_{\tau^\gamma}^{\tau^\gamma+\delta^\gamma} L^\gamma\langle\Pi^\gamma_s, g\rangle\,ds\Big|\ge\epsilon\Big\}=0.$$

Proof. We first show (iii). Let C be as in Lemma 3. Because $\delta^\gamma\to 0$, there exists γ_0 such that $\delta^\gamma<\epsilon/2C$ for $\gamma\le\gamma_0$. Then by Lemma 3

$$\Big|\int_{\tau^\gamma}^{\tau^\gamma+\delta^\gamma} L^\gamma\langle\Pi^\gamma_s, g\rangle\,ds\Big|\le\delta^\gamma C<\frac{\epsilon}{2}\quad\text{for }\gamma\le\gamma_0.$$

For (i), let γ be fixed first. Because $(M^{g,\gamma}_0)^2-\langle M^{g,\gamma}_0\rangle=0$, $\mathbf{P}$ a.e. and $(M^{g,\gamma}_t)^2-\langle M^{g,\gamma}_t\rangle$ is $\mathcal{F}_t$-martingale, by the martingale inequality and Lemma 3, we have

$$\mathbf{P}\Big\{\omega : \sup_{t\in[0,T]}|M^{g,\gamma}_t|>\epsilon\Big\}\le\frac{1}{\epsilon^2}\mathbf{E}[(M^{g,\gamma}_T)^2]=\frac{1}{\epsilon^2}\mathbf{E}[\langle M^{g,\gamma}_T\rangle]\le\frac{\gamma^d CT}{\epsilon^2}.$$

For (ii), by Lemma 3, the Chebyshev inequality, and Doob's optional stopping theorem, we have

$$\begin{aligned}\mathbf{P}\{\omega : |M^{g,\gamma}_{\tau^\gamma+\delta^\gamma}-M^{g,\gamma}_{\tau^\gamma}|>\epsilon\}&\le\frac{1}{\epsilon^2}\mathbf{E}[(M^{g,\gamma}_{\tau^\gamma+\delta^\gamma}-M^{g,\gamma}_{\tau^\gamma})^2]\\ &=\frac{1}{\epsilon^2}\mathbf{E}[\langle M^{g,\gamma}_{\tau^\gamma+\delta^\gamma}\rangle-\langle M^{g,\gamma}_{\tau^\gamma}\rangle]\le\frac{\gamma^d C\delta^\gamma}{\epsilon^2}.\end{aligned}$$

□

We now prove an exponential estimate. Let $r_\theta(x) = e^{\theta|x|} - 1 - \theta|x|$ and $s_\theta(x) = e^{\theta x} - 1 - \theta x$ for $x, \theta \in \mathbb{R}$. We define

$$\phi(\sigma, \theta) := \sum_{k\in S}\sum_{x\in\Lambda^\gamma} c^\gamma(x, \sigma, k) r_\theta(h(\sigma^{x,k}) - h(\sigma))$$
$$\psi(\sigma, \theta) := \sum_{k\in S}\sum_{x\in\Lambda^\gamma} c^\gamma(x, \sigma, k) s_\theta(h(\sigma^{x,k}) - h(\sigma)).$$

Then, from Proposition 8.8 in Darling and Norris (2008), we have for $M_T^{g,\gamma}$ in (42),

$$Z_t^{g,\gamma} := \exp\left\{\theta M_t^{g,\gamma} - \int_0^t \psi(\Sigma_s^\gamma, \theta)\, ds\right\}$$

as a supermartingale for $\theta \in \mathbb{R}$. Now let $C_g := 2\sup|g(u, i)|$ and $C_c := \sup|c^\gamma(x, \sigma, k)|$.

Lemma 4 (Exponential estimate). *There exists C that depends on C_g, C_c, S, and ϵ_0 such that for all $\epsilon \le \epsilon_0$, we have*

$$\mathbf{P}\left\{\sup_{t\le T}|M_t^{g,\gamma}| \ge \epsilon\right\} \le 2e^{-|\Lambda^\gamma|\epsilon^2/(TC)}.$$

Proof. We choose $\epsilon_0 \le |S|C_g C_c T/2$ and let $A = |S|C_g^2 C_c e/|\Lambda^\gamma|$, $\theta = \epsilon/(AT)$. Then, because r_θ is increasing in $\mathbb{R}_+$,

$$r_\theta(h(\sigma^{x,k}) - h(\sigma)) \le r_\theta\left(\frac{1}{|\Lambda^\gamma|}C_g\right) \le \frac{1}{2}\left(\frac{1}{|\Lambda^\gamma|}C_g\theta\right)^2 e^{1/|\Lambda^\gamma|\theta C_g} \quad \text{for all } \sigma \in S^{\Lambda^\gamma},$$

where we use $e^x - 1 - x \le \frac{1}{2}x^2 e^x$ for all $x > 0$ in the last line. In addition, for $\epsilon \le \epsilon_0$,

$$\frac{1}{|\Lambda^\gamma|}\theta C_g = \frac{1}{|\Lambda^\gamma|}\frac{\epsilon}{AT}C_g \le \frac{1}{|\Lambda^\gamma|}\frac{1}{2}\frac{|S|C_g^2 C_c}{A} \le \frac{1}{2e} < 1.$$

Thus,

$$\begin{aligned}\int_0^T \phi(\Sigma_t^\gamma, \theta)\, dt &\le |S||\Lambda^\gamma|1/|\Lambda^\gamma|^2\frac{1}{2}C_g^2\theta^2 e^{\frac{1}{|\Lambda^\gamma|}\theta C_g}C_c T \le \frac{1}{2}\frac{1}{|\Lambda^\gamma|}|S|C_g^2 C_c e\theta^2 T\\ &= \frac{1}{2}A\theta^2 T \quad \text{for all } \omega \in \Omega.\end{aligned}$$

Thus, because $\psi(\sigma, \theta) \le \phi(\sigma, \theta)$,

$$\begin{aligned}\mathbf{P}\left\{\sup_{t\le T} M_t^{g,\gamma} > \epsilon\right\} &= \mathbf{P}\left\{\sup_{t\le T} Z_t^{g,\gamma} > \exp\left[\theta\epsilon - \int_0^T \psi(\Sigma_t^\gamma, \theta)\, dt\right]\right\}\\ &\le \mathbf{P}\left\{\sup_{t\le T} Z_t^{g,\gamma} > \exp\left[\theta\epsilon - \frac{1}{2}A\theta^2 T\right]\right\}\\ &\le e^{(1/2)A\theta^2 T - \theta\epsilon} = e^{-|\Lambda^\gamma|\epsilon^2/(TC)},\end{aligned}$$

where we choose $C := 2|S|C_g^2 C_c e$. Because the same inequality holds for $-M_t^{g,\gamma}$, we obtain the desired result. $\square$

Convergence

Lemma 5 (Relative compactness). *The sequence* $\{\mathbf{Q}^\gamma\}$ *in* $\mathcal{P}(D([0,T];\mathcal{P}(\Lambda \times S)))$ *is relatively compact.*

Proof. By Proposition 1.7 in Kipnis and Landim (1999, p. 54), we show that $\{\mathbf{Q}^\gamma g^{-1}\}$ is relatively compact in $\mathcal{P}(D([0,T];\mathbb{R}))$ for each $g \in C(\Lambda \times S)$, where the definition of $\mathbf{Q}^\gamma g^{-1}$ is as follows. For any Borel set A in $D([0,T];\mathbb{R})$,

$$\mathbf{Q}^\gamma g^{-1}(A) := \mathbf{Q}^\gamma\big\{\pi. \in D([0,T];\mathcal{P}(\Lambda \times S)) : \langle \pi., g\rangle \in A\big\}.$$

Thus, from Theorem 1 in Aldous (1978) and the Prohorov theorem in Billingsley (1968, p. 125), it is enough to show that the following statements hold.

(i) For $\eta > 0$, there exists a such that

$$\mathbf{Q}^\gamma g^{-1}\Big\{x \in D([0,T];\mathbb{R}) : \sup_t |x(t)| > a\Big\} \le \eta \quad \text{for } \gamma \le 1.$$

(ii) For all $\epsilon > 0$,

$$\mathbf{P}\big\{\omega : |\langle \Pi^\gamma_{\tau^\gamma+\delta^\gamma}, g\rangle - \langle \Pi^\gamma_{\tau^\gamma}, g\rangle| > \epsilon\big\} \to 0$$

for $(\tau^\gamma, \delta^\gamma)$ that satisfies conditions (a) and (b) in Proposition 5. For (i), because g is bounded, it is enough to choose $a = 2\sup|g(u,i)|$, that is, $\mathbf{Q}^\gamma g^{-1}\{x \in D([0,T];\mathbb{R}) : \sup_t |x(t)| > a\} = \mathbf{Q}^\gamma\{\pi. : \sup_t |\langle \pi_t, g\rangle| > a\} = 0$ because $|\langle \pi., g\rangle| < a$ for all π. For (ii),

$$\begin{aligned}
&\mathbf{P}\big\{\omega : |\langle \Pi^\gamma_{\tau^\gamma+\delta^\gamma}, g\rangle - \langle \Pi^\gamma_{\tau^\gamma}, g\rangle| > \epsilon\big\} \\
&\qquad \le \mathbf{P}\Big\{\omega : |M^{g,\gamma}_{\tau^\gamma+\delta^\gamma} - M^{g,\gamma}_{\tau^\gamma}| > \frac{\epsilon}{2}\Big\} + \mathbf{P}\Big\{\omega : \sup_{t\in[0,T]} |M^{g,\gamma}_t| > \frac{\epsilon}{2}\Big\} \\
&\qquad \le \frac{\gamma^d C \delta^\gamma}{\epsilon^2} \quad \text{for } \gamma \le \gamma_0 \text{ chosen in Proposition 5.}
\end{aligned}$$

$\square$

Let $\mathbf{Q}^*$ be a limit point of $\{\mathbf{Q}^\gamma\}$ and choose a subsequence $\{\mathbf{Q}^{\gamma_k}\}$ that converges weakly to $\mathbf{Q}^*$. Hereafter we denote the stochastic process defined on Λ^γ by $\{\Sigma^{\Lambda^\gamma}\}$ and denote its restriction on Γ^γ by $\{\Sigma^{\Gamma^\gamma}\}$. With this notations, (42) becomes

$$\begin{aligned}
\langle \Pi^{\Gamma^\gamma}_t, g\rangle &= \langle \Pi^{\Gamma^\gamma}_0, g\rangle \\
&\quad + \frac{|\Lambda^\gamma|}{|\Gamma^\gamma|}\int_0^t ds \sum_{k\in S}\int_{\Lambda\times S} c(u,i,k,\Pi^{\Gamma^\gamma}_s)(g(u,k) - g(u,i))\, d\Pi^{\Lambda^\gamma}_s(u,i) + M^{g,\gamma}_t.
\end{aligned} \tag{44}$$

Let $\pi \in \mathcal{P}(\Gamma \times S)$ and define $d\pi_\Lambda := \mathbf{1}_{\Lambda\times S}\, d\pi$.

Lemma 6 (Characterization of limit points). *For all $\epsilon > 0$,*

$$\mathbf{Q}^*\Big\{\pi.: \sup_{t\in[0,T]} \langle \pi_t, g\rangle - \langle \pi_0, g\rangle - \int_0^t ds \sum_{k\in S}\Big[\int_{\Lambda\times S} c(u,i,k,\pi_s)(g(u,k)-g(u,i))\,d\pi_{\Lambda,s}\Big] > \epsilon\Big\} = 0.$$

That is, the limiting process is concentrated on weak solutions to the IDE (18).

Proof. We first define $\Phi : D([0,T], \mathcal{P}(\Lambda\times S)) \to \mathbb{R}$,

$$\pi. \mapsto \Big|\sup_{t\in[0,T]} \langle \pi_t, g\rangle - \langle \pi_0, g\rangle - \int_0^t ds \sum_{k\in S}\Big[\int_{\Lambda\times S} c(u,i,k,\pi_s)(g(u,k)-g(u,i))\,d\pi_{\Lambda,s}\Big]\Big|.$$

Then Φ is continuous and, thus, $\Phi^{-1}((\epsilon,\infty))$ is open. From the weak convergence of $\{\mathbf{Q}^{\gamma_k}\}$ to $\mathbf{Q}^*$,

$$\mathbf{Q}^*\{\pi. : \Phi(\pi.) > \epsilon\} \le \liminf_{l\to\infty} \mathbf{Q}^{\gamma_l}\{\pi. : \Phi(\pi.) > \epsilon\}.$$

In addition,

$$\mathbf{Q}^{\gamma}\{\pi. : \Phi(\pi.) > \epsilon\} = \mathbf{P}\Big\{\omega : \sup_{t\in[0,T]} |M_t^{g,\gamma}| > \epsilon\Big\} \le \frac{\gamma^d CT}{\epsilon^2} \quad \text{(by Proposition 5) for } \gamma < \gamma_0.$$

The first equality follows from (44) and the equality

$$\Pi_{\Lambda,s} = \frac{1}{|\Gamma^\gamma|}\sum_{x\in\Gamma^\gamma\cap\Lambda} \delta_{(\gamma x, \Sigma_s^{\Gamma^\gamma}(x))} = \frac{1}{|\Gamma^\gamma|}\sum_{x\in\Lambda^\gamma} \delta_{(\gamma x, \Sigma_s^{\Lambda^\gamma}(x))} = \frac{|\Lambda^\gamma|}{|\Gamma^\gamma|}\Pi_s^{\Lambda^\gamma}.$$

□

Lemma 7 (Absolutely continuity). *We have*

$$\mathbf{Q}^*\{\pi. : \pi_t \textit{ is absolutely continuous with respect to } m \textit{ for all } t\in[0,T]\} = 1.$$

Proof. We define $\Phi : D([0,T]; \mathcal{P}(\Gamma\times S)) \to \mathbb{R}$, $\pi. \mapsto \sup_{t\in[0,T]} |\langle \pi_t, g\rangle|$. Then Φ is continuous. In addition,

$$|\langle \pi^\gamma, g\rangle| \le \frac{1}{|\Gamma^\gamma|}\sum_{x\in\Gamma^\gamma} |g(\gamma x, \sigma(x))| \le \sum_{l\in S}\frac{1}{|\Gamma^\gamma|}\sum_{x\in\Gamma^\gamma} |g(\gamma x, l)|.$$

Thus

$$\sup_{t\in[0,T]} |\langle \pi_t^\gamma, g\rangle| \le \sum_{l\in S}\frac{1}{|\Gamma^\gamma|}\sum_{x\in\Gamma^\gamma} |g(\gamma x, l)|.$$

We write $\pi_\cdot^*$ to be a trajectory on which all $\mathbf{Q}^*$'s are concentrated. Then $\Pi_\cdot^\gamma \xrightarrow{\mathcal{D}} \pi_\cdot^*$ (convergence in distribution) and, thus, $\mathbf{E}(\Phi(\Pi_\cdot^\gamma)) \to \mathbf{E}(\Phi(\pi_\cdot^*))$. In addition,

$(1/|\Gamma^\gamma|)\sum_{x\in\Gamma^\gamma}|g(\gamma x, l)| \to \int_\Lambda |g(u,l)|\,du$ for all l by the Riemann sum approximations. Therefore,

$$\sup_{t\in[0,T]} |\langle \pi_t^*, g\rangle| = \Phi(\pi_\cdot^*) = \lim_{\gamma\to 0} \mathbf{E}(\Phi(\Pi_\cdot^\gamma)) \le \lim_{\gamma\to 0}\sum_{l\in S}\frac{1}{|\Gamma^\gamma|}\sum_{x\in\Gamma^\gamma}|g(\gamma x, l)|$$
$$= \int_{\Gamma\times S} |g(u,l)|\,dm(u,i).$$

Therefore, for all $t\in[0,T]$ and $g\in C(\Gamma\times S)$,

$$\left|\int_{\Gamma\times S} g(u,l)\,d\pi_t^*\right| \le \int_{\Gamma\times S}|g(u,l)|\,dm(u,i),$$

and, thus, for all $t\in[0,T]$, π_t^* is absolutely continuous with respect to $dm(u,i)$. $\square$

We also see that all limit points of the sequence $\{\mathbf{Q}^\gamma\}$ are concentrated on the trajectories that are equal to $f^0 m$ at time 0 because

$$\mathbf{Q}^*\left\{\pi_\cdot : \left|\int g(u,i)\,d\pi_0 - \frac{1}{|\Gamma|}\int g(u,i)f^0(u,i)\,dm(u,i)\right| > \epsilon\right\}$$
$$\le \liminf_{k\to\infty}\mathbf{Q}^{\gamma_k}\left\{\pi_\cdot : \left|\int g(u,i)\,d\pi_0 - \frac{1}{|\Gamma|}\int g(u,i)f^0(u,i)\,dm(u,i)\right| > \epsilon\right\} = 0,$$

where the definition of a sequence of product measures with a slowly varying parameter implies the last equality by Proposition 0.4 in Kipnis and Landim (1999, p. 44).

We have thus far shown that $\mathbf{Q}^*$'s are concentrated on trajectories that are weak solutions to the integro-differential equations. We now show the uniqueness of weak solutions defined as follows. Let $\mathcal{A}(f)(u,i) := \sum_{k\in S}\mathbf{c}(u,k,i,f)f_\Lambda(t,u,k) - f_\Lambda(t,u,i)\sum_{k\in S}\mathbf{c}(u,i,k,f)$. For an initial profile $f^0\in\mathcal{M}$, $f\in\mathcal{M}$ is a weak solution to the Cauchy problem

$$\frac{\partial f_t}{\partial t} = \mathcal{A}(f_t), \quad f_0 = f^0 \tag{45}$$

if, for every function $g\in C(\Gamma\times S)$ and for all $t<T$, $\langle f_t, g\rangle = \int_0^t \langle \mathcal{A}(f_s), g\rangle ds$. Observe that from C3 that $\mathcal{A}$ satisfies the Lipschitz condition that there exists C such that for all f, $\tilde f \in L^\infty([0,T]; L^\infty(\Gamma\times S))$, $\|\mathcal{A}(f)-\mathcal{A}(\tilde f)\|_{L^2(\Gamma\times S)} \le C\|f-\tilde f\|_{L^2(\Gamma\times S)}$.

Lemma 8 (Uniqueness of weak solutions). *Weak solutions to the Cauchy problem (45) that belong to $L^\infty([0,T]; L^2(\Gamma\times S))$ are unique.*

Proof. Let f_t and $\tilde f_t$ be two weak solutions, and let $\bar f_t := f_t - \tilde f_t$. Then we have

$$\langle \bar f_t, g\rangle = \int_0^t \langle \mathcal{A}(f_s) - \mathcal{A}(\tilde f_s), g\rangle\,ds \quad \text{for all } g\in C(\Gamma\times S).$$

We show that $t\mapsto \|\bar f_t\|^2_{L^2(\Gamma\times S)}$ is differentiable. Define a mollifier $\eta(x) := C\exp(1/(|x|-1))$ if $|x|<1$, $:= 0$ if $|x|\ge 1$, where $C>0$ is a constant such that

$\int_{\mathbb{R}^d} \eta(x)\,dx = 1$. For $\epsilon > 0$, set $\eta_\epsilon(x) := \epsilon^{-d}\eta(\epsilon^{-1}x)$. For each $u \in \Gamma$, $i \in S$, define $h^\epsilon_{u,i}(v,k) = \eta_\epsilon(u-v)\mathbf{1}_{\{i=k\}}$ and

$$\bar{f}^\epsilon_t(u,i) := \int_{\Gamma\times S} (f_t(v,k) - \tilde{f}_t(v,k))h^\epsilon_{u,i}(v,k)\,dm(v,k).$$

Then

$$\begin{aligned} \left|\langle \mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s), h^\epsilon_{u,i}\rangle\right| &\le \|\mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s)\|_{L^2}\|h^\epsilon_{u,i}\|_{L^2} \le C\|f_s - \tilde{f}_s\|_{L^2}\|h^\epsilon_{u,i}\|_{L^2} \\ &\le C \sup_{s\in[0,T]} \|f_s - \tilde{f}_s\|_{L^2}\|h^\epsilon_{u,i}\|_{L^2}. \end{aligned}$$

Because $f_s - \tilde{f}_s \in L^\infty([0,T]; L^2(\Gamma\times S))$ and $h^\epsilon_{u,i} \in C(\Gamma \times S)$ for each u, i, $t \mapsto \bar{f}^\epsilon_t(u,i)$ is differentiable and its derivative is $\bar{f}^{\epsilon\prime}_t(u,i) = \langle \mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s), h^\epsilon_{u,i}\rangle$. In addition, it follows that $\|\bar{f}^\epsilon_t\|^2_{L^2}$ is differentiable with respect to t and that

$$\frac{d}{dt}\|\bar{f}^\epsilon_t\|^2_{L^2} = \int_{\Gamma\times S} 2\langle \mathcal{A}(f_t) - \mathcal{A}(\tilde{f}_t), h^\epsilon_{u,i}\rangle \bar{f}^\epsilon_t(u,i)\,dm(u,i),$$

so

$$\|\bar{f}^\epsilon_t\|^2_{L^2} = \int_0^t \left[\int_{\Gamma\times S} 2\langle \mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s), h^\epsilon_{u,i}\rangle \bar{f}^\epsilon_t(u,i)\,dm(u,i)\right] ds.$$

Then, because $f^\epsilon_t \to f_t$ in $\|\cdot\|_{L^2}$ and $\bar{f}_t \in L^\infty([0,T]; L^2(\Gamma\times S))$ for a given t, we have $|\|\bar{f}^\epsilon_t\|^2_{L^2} - \|\bar{f}_t\|^2_{L^2}| \to 0$. In addition, because $\langle \mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s), h^\epsilon_{u,i}\rangle \to \mathcal{A}(f_t)(u,i) - \mathcal{A}(\tilde{f}_t)(u,i)$ for a.e. u and all i, t, by the dominant convergence theorem, we have

$$\|\bar{f}_t\|^2_{L^2} = \int_0^t 2\langle \mathcal{A}(f_s) - \mathcal{A}(\tilde{f}_s), \bar{f}_s\rangle\,ds.$$

Thus, $\|\bar{f}_t\|^2_{L^2}$ is differentiable and

$$\frac{d}{dt}\|\bar{f}_t\|^2_{L^2} = \langle \mathcal{A}(f_t) - \mathcal{A}(\tilde{f}_t), \bar{f}_t\rangle \le 2\|\mathcal{A}(f_t) - \mathcal{A}(\tilde{f}_t)\|_{L^2}\|\bar{f}_t\|_{L^2} \le C\|\bar{f}_t\|^2_{L^2}.$$

Hence, from the Gronwall lemma, the uniqueness of the solutions follows. □

Lemma 9 (Convergence in probability). *We have*

$$\Pi^\gamma_t \longrightarrow \frac{1}{|\Gamma|} f_t m \quad \textit{in probability.}$$

Proof. We have thus far established $\mathbf{Q}^\gamma \Rightarrow \mathbf{Q}^*$ (converge weakly) and, equivalently, $\Pi^\gamma_\cdot \to \pi^*_\cdot$ in the Skorohod topology (topology on $D([0,T], \mathcal{P}(\mathbf{T}^d \times S))$). If we show that $\Pi^\gamma_t \to \pi^*_t$ weakly in $\mathcal{P}(\Gamma^d \times S)$ or, equivalently, $\Pi^\gamma_t \to^{\mathcal{D}} \pi^*_t$ in distribution for fixed time $t < T$, then we have

$$\Pi^\gamma_t \xrightarrow{\mathbf{P}} \pi^*_t \quad \text{in probability.} \tag{46}$$

Because $\Pi^\gamma_\cdot \to \pi^*_\cdot$ in the Skorohod topology implies $\Pi^\gamma_t \to \pi^*_t$ weakly for continuity points of $\pi^*_\cdot$ Billingsley (1968, p. 112), it is enough to show that $\pi^*_\cdot : t \mapsto \pi^*_t$ is continuous for all $t \in [0, T]$ to obtain (46). Let $t_0 < T$ and let $\{g_k\}$ be a dense family in $C(\Gamma \times S)$. Because

$$\left|\int_{t_0}^{t} \langle A(\pi^*_s), g_k\rangle\, ds\right| \le (t - t_0) \sup_{s\in[0,T]} \langle \mathcal{A}(\pi^*_s), g_k\rangle,$$

we choose $\delta \le \min\{1, \epsilon\}$. Then for $|t - t_0| \le \delta$,

$$\frac{|\int_{t_0}^{t}\langle A(\pi^*_s), g_k\rangle\, ds|}{1 + |\int_{t_0}^{t}\langle A(\pi^*_s), g_k\rangle\, ds|} \le \frac{\delta \sup_{s\in[0,T]}\langle \mathcal{A}(\pi^*_s), g_k\rangle}{1 + \delta \sup_{s\in[0,T]}\langle \mathcal{A}(\pi^*_s), g_k\rangle} \le \delta \frac{\sup_{s\in[0,T]}\langle \mathcal{A}(\pi^*_s), g_k\rangle}{1 + \sup_{s\in[0,T]}\langle \mathcal{A}(\pi^*_s), g_k\rangle} \le \delta$$

and, thus, $\|\pi_t - \pi_{t_0}\|_{\mathcal{P}(\Gamma\times S)} \le \epsilon$, and $\pi^*_\cdot : t \mapsto \pi^*_t$ is continuous for all $t \in [0, T]$. Thus, all $t \in [0, T]$ are continuity points of $\pi^*_\cdot$ □

Proof of Theorem 2 From Lemma 9, we have

$$\Pi^{\Lambda\gamma}_t \xrightarrow{\mathbf{P}} \frac{1}{|\Lambda|} f_{\Lambda,t} m$$

for $t < T$. Thus, from (21) we obtain

$$\langle f_t, g\rangle = \langle f_0, g\rangle + \int_0^t ds \sum_{k\in S} \int_{\Lambda\times S} c\Big(u, i, k, \frac{1}{|\Gamma|} f_s m(u, i)(g(u, k) - g(u, i))\Big) f_{\Lambda,s}\, dm(u, i).$$

Because $|\Gamma^\gamma|\Pi^{\Gamma\gamma}_t = |\Lambda^\gamma|\Pi^{\Lambda\gamma}_t + |\Lambda^{\gamma c}|\Pi^{\Lambda^{\gamma c}}_0$, $|\Gamma|\Pi^{\Gamma\gamma}_t \xrightarrow{\mathbf{P}} f_t m$, $|\Lambda|\Pi^{\Lambda\gamma}_t \xrightarrow{\mathbf{P}} f_{\Lambda,t} m$, and $|\Lambda^c|\Pi^{\Lambda^{\gamma c}}_0 \xrightarrow{\mathbf{P}} f_{\Lambda^c} m$, we have $f_t = f_{\Lambda,t} + f_{\Lambda^c}$ for all t.

A.3 *Proof of Theorem 3*

For this, we first define the reduction mapping $\phi : S^{\Lambda^n} \to \Delta^n$ as

$$\sigma \mapsto \phi(\sigma), \quad \phi(\sigma)(i) := \frac{1}{|\Lambda^n|} \sum_{y\in\Lambda^n} \delta_{\sigma(y)}(\{i\}).$$

For $g \in L^\infty(\Delta^n; \mathbb{R})$, we let $f := g \circ \phi \in L^\infty(S^{\Lambda^n}; \mathbb{R})$, where$f(\sigma) = g(\eta)$. Then for $\eta = \phi(\sigma)$, we have $f(\sigma^{x,k}) - f(\sigma) = g(\eta^{\sigma(x),k}) - g(\eta)$ because

$$\phi(\sigma^{x,k})(i) = \frac{1}{n^d} \sum_{y\in\Lambda_n} \delta_{\sigma(y)}(\{i\}) + \frac{1}{n^d}\delta_k(\{i\}) - \frac{1}{n^d}\delta_{\sigma(x)}(\{i\}) = \eta^{\sigma(x),k}(i).$$

Proof of Theorem 3. We check the case of noninnovative and comparing rates. Other cases can be treated as special cases. By writing $m^n(k) := \sum_l a(k, l)\eta^n(l)$, we find that

$$L_n f(\sigma) = \sum_{k\in S} \sum_{x\in\Lambda_n} \eta(k) F\big(m^n(k) - m^n(\sigma(x))\big)(g(\eta^{\sigma(x),k}) - g(\eta))$$

$$
\begin{aligned}
&= \sum_{k\in S}\sum_{j\in S}\Bigg[\sum_{x\in\Lambda_n} \delta_{\sigma(x)}(\{j\})\Bigg]\eta(k)F(m^n(k)-m^n(j))(g(\eta^{j,k})-g(\eta)) \\
&= \sum_{k\in S}\sum_{j\in S} n^d\eta(j)\eta(k)F(m^n(k)-m^n(j))(g(\eta^{j,k})-g(\eta)) \\
&:= \sum_{k\in S}\sum_{j\in S} n^d c^M(\eta,j,k)(g(\eta^{j,k})-g(\eta)).
\end{aligned}
$$

Thus we obtain

$$
L^n g(\eta) = \sum_{k\in S}\sum_{j\in S} n^d c^M(\eta,j,k)(g(\eta^{j,k})-g(\eta)).
$$

This makes $\{\eta_t\}$ a Markov chain and the rate is given by $c^M(\eta,j,k)$. □

A.4 *Proof of Corollary 4*

Proof. It is enough to prove the exponential estimate. From (21), we recall that

$$
\langle \Pi_t^\gamma, g\rangle = \langle \Pi_0^\gamma, g\rangle + \int_0^t \sum_{k\in S}\int_{\mathbf{T}^d\times S} c(u,i,k,\Pi_s^\gamma)(g(u,k)-g(u,i))\,d\Pi_s^\gamma(u,i)\,ds + M_t^{g,\gamma}
$$

for $g\in C(\mathbf{T}^d\times S)$. By taking $g(u,i)=1$ if $i=l$ and $g(u,i)=0$ otherwise, we find

$$
\eta_{t,l}^n = \eta_{0,l}^n + n^d\int_0^t\Bigg[\sum_{i\in S} c^M(i,l,\eta_s^n)\eta_{s,l}^n - \sum_{k\in S} c^M(l,k,\eta_s^n)\eta_{s,l}^n\Bigg]ds + M_t^{l,n}.
$$

We define $\beta_l(x) := \sum_{i\in S}\mathbf{c}^M(i,l,x)x_l - \sum_{k\in S}\mathbf{c}^M(l,k,x)x_l$. Thus, we have

$$
\eta_{t,l}^n = \eta_{0,l}^n + n^d\int_0^t \beta_l(\eta_s^n)\,ds + M_t^{l,n}, \qquad \rho_{t,l} = \rho_{0,l} + \int_0^t \beta_l(\rho_s)\,ds.
$$

From Lemma 4, we have $\mathbf{P}\{\sup_{t\le T}|M_t^{l,n}|\ge\delta\}\le 2e^{-n^d\delta^2/(TC_0)}$ for each l and for $\delta\le\delta_0$, where we note that the choices of C_0 and δ_0 do not depend on g because $|g(u,i)|\le 1$ for all u,i. Thus, $\mathbf{P}\{\sup_{t\le T}\|M_t^n\|_u\ge\delta\}\le 2|S|e^{-n^d\delta^2/(TC_0)}$. Thus, using the Lipschitz condition for β, we obtain

$$
\sup_{\tau\le t}\|\eta_\tau^n-\rho_\tau\|_u \le \|\eta_0^n-\rho_0\|_u + L\int_0^t \sup_{\tau\le s}\|\eta_\tau^n-\rho_\tau\|_u\,ds + \sup_{t\le T}\|M_t^n\|_u
$$

for $t\le T$. For ϵ_0 in Lemma 4, we let $\delta=\frac{1}{3}e^{-LT}\epsilon$ for $\epsilon<\epsilon_0$ and define

$$
\Omega_0 = \{\omega : \|\eta_0^n-\rho_0\|_u\le\delta\}, \qquad \Omega_1 = \Big\{\omega : \sup_{t\le T}\|M_t^n\|_u\le\delta\Big\}.
$$

Then, for $\omega \in \Omega_0 \cap \Omega_1$, we have $\sup_{\tau \leq T} \|\eta^n_\tau - \rho_\tau\|_u \leq 2\delta e^{LT}$ by the Gronwall lemma. Choose n_0 such that $\|\eta^n_0 - \rho_0\|_u \leq \delta$ for a.e. ω for $n \geq n_0$. Then for $\epsilon \leq \epsilon_0$ and $n \geq n_0$,

$$
\begin{aligned}
P\Big\{\sup_{\tau \leq T} \|\eta^n_\tau - \rho_n\| \geq \epsilon\Big\} &\leq P(\Omega^c_0) + P(\Omega^c_1) \\
&\leq P\{\omega : \|\eta^n_0 - \rho_0\|_u \geq \delta\} + P\Big\{\omega : \sup_{t \leq T} \|M^n_t\|_u \geq \delta\Big\} \\
&\leq 2|S| e^{-n^d \delta^2/(TC_0)} = 2|S| e^{-n^d \epsilon^2/(TC)},
\end{aligned}
$$

where $C := 9C_0 e^{2LT}$. □

A.5 *Solutions to Linear IDE's*

Applying the Fourier transform to (29) on an element-by-element basis, we obtain

$$
\frac{\partial \hat{Z}(k)}{\partial t} = (D_r\Phi(f_0, f_0)\hat{\mathcal{J}}(k) + D_s\Phi(f_0, f_0))\hat{Z}(k) \tag{47}
$$

for each $k \in \mathbb{Z}^d$ and $\hat{Z}(k) \in \mathbb{C}^{|S|}$. By solving the ODE system (47) for each k and using the inverse formula, we obtain

$$
D(x, t) = \sum_{k \in \mathbb{Z}} e^{(D_r\Phi(f_0, f_0)\hat{\mathcal{J}}(k) + D_s\Phi(f_0, f_0))t} \hat{g}(k) e^{2\pi i x \cdot k},
$$

where $e^{(D_r\Phi(f_0, f_0)\hat{\mathcal{J}}(k) + D_s\Phi(f_0, f_0))t}$ is an $|S| \times |S|$ matrix and $\hat{g}(k)$ is an $|S| \times 1$ vector.

A.6 *Proof of Proposition 4*

Proof. We first note that $p_1 > \zeta$, $p_2, p_3 < \zeta$. Then for $\alpha := a_{11} - a_{21} + a_{22} - a_{12}$,

$$
\begin{aligned}
&\beta\alpha\big(1 - l_\beta(\alpha(p_i - \zeta))\big) l_\beta(\alpha(p_i - \zeta)) < 1 \quad \text{for } i = 1, 3 \\
&\beta\alpha\big(1 - l_\beta(\alpha(p_i - \zeta))\big) l_\beta(\alpha(p_i - \zeta)) > 1 \quad \text{for } i = 2.
\end{aligned}
$$

Suppose that $\beta > \beta_C$ and consider p_1. Because $l_\beta(\alpha(p_1 - \zeta)) = p_1$, we have $\beta\alpha(1 - p_1)p_1 < 1$. Then, because $\hat{\mathcal{J}}(k) \leq 1$ for all k, we have

$$
\lambda_L(k) = \beta\alpha(1 - p_1)p_1\hat{\mathcal{J}}(k) - 1 < \beta\alpha(1 - p_1)p_1 - 1 < 0.
$$

Thus p_1 is linearly stable. A similar argument shows that p_3 is linearly stable. The case of the Hawk–Dove type game follows from $\alpha < 0$. □

References

Aldous, David (1978), "Stopping times and tightness." *The Annals of Probability*, 6, 335–340. [864]

Apesteguia, Jose, Steffen Huck, and Jorg Oechssler (2007), "Imitation—theory and experimental evidence." *Journal of Economic Theory*, 136, 217–235. [830]

Apesteguia, Jose, Steffen Huck, Jorg Oechssler, and Simon Weidenholzer (2010), "Imitation and the evolution of Walrasian behavior: Theoretically fragile but behaviorally robust." *Journal of Economic Theory*, 145, 1603–1617. [830]

Bates, Peter W., Paul C. Fife, Xiaofeng Ren, and Xuefeng Wang (1997), "Traveling waves in a convolution model for phase transitions." *Archive for Rational Mechanics and Analysis*, 138, 105–136. [850]

Benaim, Michel and Jorgen W. Weibull (2003), "Deterministic approximation of stochastic evolution in games." *Econometrica*, 71, 873–903. [831, 841, 843, 848, 849]

Bergin, James and Dan Bernhardt (2009), "Cooperation through imitation." *Games and Economic Behavior*, 67, 976–988. [830]

Billingsley, Patrick (1968), *Convergence of Probability Measures*. Wiley, New York. [864, 868]

Bisin, Alberto and Thierry Verider (2011), "The economics of cultural transmission and socialization." In *Handbook of Social Economics*, 339–416, Elsevier, San Diego. [830]

Blume, Lawrence E. (1993), "The statistical mechanics of strategic interaction." *Games and Economic Behavior*, 5, 387–424. [830, 833, 838, 840]

Blume, Lawrence E. (1995), "The statistical mechanics of best-response strategy revision." *Games and Economic Behavior*, 11, 111–145. [830, 838]

Blume, Lawrence E. (1997), "Population games." In *The Economy as a Complex Adaptive System II* (W. Brian Arthur, Steven N. Durlauf, and David A. Lane, eds.), 425–460, Addison-Wesley, Reading, Massachusetts. [834, 848]

Bowles, Samuel (2004), *Microeconomics: Behavior, Institutions and Evolution*. Princeton University Press, Princeton. [830]

Brock, William A. and Steven N. Durlauf (1999), "A formal model of theory choice in science." *Economic Theory*, 14, 113–130. [854]

Brock, William A. and Steven N. Durlauf (2001), "Discrete choice with social interactions." *Review of Economic Studies*, 68, 235–260. [854]

Carr, Jack and Robert L. Pego (1989), "Metastable patterns in solutions of $u_t = \epsilon^2 u_x x - f(u)$." *Communications on Pure and Applied Mathematics*, 42, 523–576. [855]

Chen, Xinfu (1997), "Existence, uniqueness, and asymptotic stability of traveling waves in nonlocal evolution equations." *Advances in Differential Equations*, 2, 125–160. [835, 836, 854]

Collet, Pierre and Jean Pierre Eckmann (1990), *Instabilities and Fronts in Extended Systems*. Princeton University Press, Princeton. [849]

Comets, Francis (1987), "Nucleation for a long range magnetic model." *Annales de l'Institut Henri Poincare*, 23, 135–178. [838, 847]

Darling, Richard and James Norris (2008), "Differential equation approximations for Markov chains." *Probability Survey*, 5, 37–79. [831, 861, 863]

De Masi, Anna, Thierry Gobron, and Errico Presutti (1995), "Traveling fronts in non-local evolution equations." *Archive for Rational Mechanics and Analysis*, 132, 143–205. [835, 854]

De Masi, Anna, Enza Orlandi, Errico Presutti, and Livio Triolo (1994), "Glauber evolution with Kac potentials: I. Mesoscopic and macroscopic limits, interface dynamics." *Nonlinearity*, 7, 633–696. [835, 838, 853]

De Masi, Anna, Enza Orlandi, Errico Presutti, and Livio Triolo (1996a), "Glauber evolution with Kac potentials: II. Fluctuations." *Nonlinearity*, 9, 27–51. [837]

De Masi, Anna, Enza Orlandi, Errico Presutti, and Livio Triolo (1996b), "Glauber evolution with Kac potentials: III. Spinodal decomposition." *Nonlinearity*, 9, 53–114. [837]

Durrett, Rick (1999), "Stochastic spatial models." *SIAM Review*, 41, 677–718. [830, 832]

Durrett, Rick (2002), *Mutual Invadability Implies Coexistence in Spatial Models*. American Mathematical Society, Providence. [836]

Eisele, Theodor and Richard S. Ellis (1983), "Symmetry breaking and random waves for magnetic systems on a circle." *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 63, 297–348. [837]

Ellison, Glenn (1993), "Learning, local interaction, and coordination." *Econometrica*, 61, 1047–1071. [830, 836]

Ellison, Glenn (2000), "Basins of attraction, long-run stochastic stability, and the speed of step-by-step by evolution." *Review of Economic Studies*, 67, 17–45. [830, 836]

Ethier, Stewart N. and Thomas G. Kurtz (1986), *Markov Processes: Characterization and Convergence*. Wiley, New York. [839, 846, 861]

Fife, Paul C. (1979), *Mathematical Aspects of Reaction and Diffusing Systems*. Springer-Verlag, Berlin. [849]

Fisher, Arthur and Jerrold E. Marsden (1975), "Linearization stability of nonlinear partial differential equations." *Proceedings of Symposia in Pure Mathematics*, 27, 219–263. [849]

Freidlin, Mark I. and Alexander D. Wentzell (1998), *Random Perturbations of Dynamical Systems*, second edition. New York: Springer. [837]

Fudenberg, Drew and David K. Levine (1998), *The Theory of Learning in Games*. MIT Press, Cambridge, Massachusetts. [831]

Grauwin, Florence Goffette-Nago Sebastian and Pablo Jensen (2012), "Dynamic models of residential segregation: An analytical solution." *Journal of Public Economics*, 96, 124–141. [830]

Harsanyi, John C. and Reinhard Selten (1988), *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge, Massachusetts. [834]

Hofbauer, Josef (1997), "Equilibrium selection via traveling waves." In *Game Theory, Experience, Rationality: Foundations of Social Sciences, Economics and Ethics* (Werner Leinfellner and Eckehart Kohler, eds.), 245–249, Vienna Circle Institute Yearbook 5, Kluwer Academic, Dordrecht. [832, 854]

Hofbauer, Josef (1999), "The spatially dominant equilibrium of a game." *Annals of Operations Research*, 89, 233–251. [832, 854]

Hofbauer, Josef, Vivian Hutson, and Glenn T. Vickers (1997), "Travelling waves for games in economics and biology." *Nonlinear Analysis*, 30, 1235–1244. [832, 854]

Hofbauer, Josef and Karl Sigmund (1998), *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge. [845]

Hofbauer, Josef and Karl Sigmund (2003), "Evolutionary game dynamics." *Bulletin of American Mathematical Society*, 40, 479–519. [831, 841]

Hopkins, Ed (1999), "A note on best response dynamics." *Games and Economic Behavior*, 29, 138–150. [853]

Hutson, Vivian and Glenn T. Vickers (1992), "Travelling waves and dominance of ESS's." *Journal of Mathematical Biology*, 30, 457–471. [832, 854]

Josephson, Jens and Alexander Matros (2004), "Stochastic imitation in finite games." *Games and Economic Behavior*, 49, 244–259. [830]

Katsoulakis, Markos A., Petr Plechac, and Dimitrios K. Tsagkarogiannis (2005), "Mesoscopic modeling for continuous spin lattice systems: Model problems and micromagnetic applications." *Journal of Statistical Physics*, 119, 347–389. [838, 847]

Katsoulakis, Markos A. and Panagiotis E. Souganidis (1997), "Stochastic ising models and anisotropic front propagation." *Journal of Statistical Physics*, 87, 63–89. [839, 853, 855]

Kipnis, Claude and Claudio Landim (1999), *Scaling Limits of Interacting Particle Systems*. Springer-Verlag, Berlin. [842, 847, 864, 866]

Kreindler, Gabriel H. and H. Peyton Young (2013), "Fast convergence in evolutionary equilibrium selection," *Games and Economic Behavior*, 80, 39–67. [836]

Kurtz, Thomas G. (1970), "Solutions of ordinary differential equations as limits of pure jump Markov process." *Journal of Applied Probability*, 7, 49–58. [831]

Lebowitz, Joel L. and Oliver Penrose (1966), "Rigorous treatment of the van der Waals–Maxwell theory of the liquid vapour transition." *Journal of Mathematical Physics*, 7, 98–113. [838]

Levine, David K. and Wolfgang Pesendorfer (2007), "The evolution of cooperation through imitation." *Games and Economic Behavior*, 58, 293–315. [830]

Ligget, Thomas (1985), *Interacting Particle Systems*. Springer-Verlag, New York. [839]

Kandori, Michihiro, George J. Mailath, and Rafael Rob (1993), "Learning, mutation, and long-run equilibria in games." *Econometrica*, 61, 29–56. [830]

Murray, James D. (1989), *Mathematical Biology*. Springer, Berlin. [849, 850, 857]

Orlandi, Enza and Livio Triolo (1997), "Traveling fronts in nonlocal models for phase separation in external field." *Proceedings of the Royal Society*, 127 A, 823–835. [835, 854, 855]

Presutti, Errico (2009), *Scaling Limits in Statistical Mechanics and Microstructures in Continuum Mechanics*. Springer, Berlin. [837, 838, 853]

Robson, Arthur J. and Fernando Vega-Redondo (1996), "Efficient equilibrium selection in evolutionary games with random matching." *Journal of Economic Theory*, 70, 65–92. [830]

Sandholm, William H. (2010a), "Pairwise comparison dynamics and evolutionary foundations for Nash equilibrium." *Games*, 1, 3–17. [840]

Sandholm, William H. (2010b), *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, Massachusetts. [831, 834, 839, 845, 853]

Schelling, Thomas C. (1971), "Dynamic models of segregation." *Journal of Mathematical Sociology*, 1, 143–186. [830]

Smith, Michael J. (1984), "The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov." *Transportation Science*, 18, 245–252. [840]

Stein, Elias M. and Rami Shakarchi (2003), *Fourier Analysis: An Introduction*. Princeton University Press, Princeton. [851]

Szabo, Gyorgy and Gabor Fath (2007), "Evolutionary games on graphs." *Physics Report*, 446, 97–216. [830, 838, 839]

Vickers, Glenn, Vivian Hutson, and Chris J. Budd (1993), "Spatial patterns in population conflicts." *Journal of Mathematical Biology*, 31, 411–430. [832, 850]

Volpert, Aizik I., Vitaly A. Volpert, and Vladimir A. Volpert (1994), *Traveling Wave Solutions of Parabolic Systems*. American Mathematical Society, Providence. [850]

Weibull, Jorgen W. (1995), *Evolutionary Games Theory*. MIT Press, Cambridge, Massachusetts. [831, 841, 845]

Young, H. Peyton (1993), "The evolution of conventions." *Econometrica*, 61, 57–84. [830]

Young, H. Peyton (1998), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton. [830, 833, 834, 838]

Young, H. Peyton and Mary A. Burke (2001), "Competition and custom in economic contracts: A case study of illinois agriculture." *American Economic Review*, 91, 559–573. [830]

Submitted 2010-7-2. Final version accepted 2012-10-14. Available online 2012-10-14.