J. Appl. Prob. **54,** 522–539 (2017)
*doi:*10.1017/jpr.2017.15

A SPATIAL MODEL FOR SELECTION AND COOPERATION

PETER CZUPPON * AND
PETER PFAFFELHUBER,** *University of Freiburg*

Abstract

We study the evolution of cooperation in an interacting particle system with two types. The model we investigate is an extension of a two-type biased voter model. One type (called defector) has a (positive) bias α with respect to the other type (called cooperator). However, a cooperator helps a neighbor (either defector or cooperator) to reproduce at rate γ. We prove that the one-dimensional nearest-neighbor interacting dynamical system exhibits a phase transition at $\alpha = \gamma$. A special choice of interaction kernels yield that for $\alpha > \gamma$ cooperators always die out, but if $\gamma > \alpha$, cooperation is the winning strategy.

Keywords: Interacting particle system; voter model; cooperation; phase transition; extinction; clustering

2010 Mathematics Subject Classification: Primary 60K35
Secondary 82C22; 92D15

1. Introduction

In nature cooperative behavior amongst individuals is widespread. It is observed in animals, e.g. [5], [11], as well as in microorganisms, e.g. [6], [19]. In an attempt to understand this phenomenon in terms of models, theoretical approaches have introduced different interpretations and forms of cooperation, mostly within the area of game theory [15]. In all such approaches, a defector (or selfish) type tends to have more offspring, but there are cases when it is outcompeted by the cooperator type under some circumstances. Although, in all the models describing cooperation, the question of extinction and survival of a type or the coexistence of several types are the main subjects of the mathematical analysis, the frameworks for the theoretical studies may vary. While (stochastic) differential equations are mainly used for nonspatial systems (see, e.g. [1], [12]), the theory of interacting particle systems provides a suitable setup for the analysis of models with local interactions between the particles, [2], [10], [17]. In this paper we define a model using the latter structure and terminology.

Investigations of models incorporating cooperation are interesting because of the following dichotomy: in nonspatial (well-mixed) situations, the whole population benefits from the cooperative behavior. If defectors have a higher fitness than cooperators, defectors always outcompete cooperators in the long run. However, if the system is truly spatial, cooperators can form clusters and then use their cooperative behavior in order to defend themselves against defectors, even though those might have higher (individual) fitness. This heuristic suggests that only structured models can help to understand cooperative behavior in nature. For the model studied in the present paper, we will make it precise in Proposition 2 for extinction of

Received 8 March 2016; revision received 2 November 2016.
* Current address: Max-Planck-Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Plön, Germany. Email address: czuppon@evolbio.mpg.de
** Postal address: Department of Mathematical Stochastics, University of Freiburg, 79104 Freiburg, Germany.

cooperators in a nonspatial system and in Theorem 1 for extinction of defectors in a spatial system, if cooperation is strong enough.

Due to the variety of interpretations of cooperative behavior there are different ways of implementing these mechanisms in a spatial context. In the field of population dynamics, Sturm and Swart [17] studied an interacting particle system containing a cooperative-branching mechanism which can be understood as a sexual reproduction event. In [2], Blath and Kurt studied a branching-annihilating random walk and again, a cooperation mechanism is interpreted as sexual reproduction. In contrast, the model introduced by Evilsizor and Lanchier in [10] originates from the game-theoretical study of a two-player game with different strategies, where the strategies can be altruistic or selfish. Here, the altruistic strategies represent the cooperator type. We discuss the findings of these models to our results in Section 4.

Various interacting particle systems that appear in the literature are attractive, i.e. two versions of the system, which start in configurations where one dominates the other, can be coupled such that this property holds for all times; see, e.g. [17] for an attractive model mentioned above. For such processes, there exist several general results (see [13]) which provide some useful techniques to help in the analysis. However, cooperation often leads to nonattractive interacting particle systems; see [2], [10], and the one presented here. The reason here is that cooperators (or altruists) do not distinguish between noncooperators and their own type, which usually contradicts attractiveness.

The motivation for the present paper came from studies of bacterial cells in the context of public good dilemmas, e.g. [4], [8]. The idea is that there are two types (defector $= 0$, cooperator $= 1$), where only cooperators produce some public good that helps neighboring cells to reproduce. However, this production is costly which means that defectors will have a selective advantage over the cooperator type. The resulting model is a biased voter model with an additional cooperation mechanism. The main objective of our paper is to study the long-time behavior of such a model dependent on the parameters of the system.

In particular, we prove, for our main model in one dimension from Definition 1, that the system clusters independently of the parameter configuration. When starting in a translation invariant configuration, for $\alpha > \gamma$, defectors take over the population, whereas for $\gamma > \alpha$ cooperators win; see Theorem 1. Additionally, in higher dimensions, at least we can show that the parameter region where the defecting particle win is larger than for $d = 1$; see Theorem 2. We also show that a finite number of cooperators die out if $\alpha > \gamma$, but may survive if $\gamma > \alpha$. The converse holds for the defectors; see Theorem 3. What remains to be seen is if there are parameter combinations such that cooperators win also in higher dimensions. Some preliminary results in the limit of small parameters α and γ can be found in [7].

The paper is structured as follows. First, we give a general definition of the model in Section 2. After the definition we derive some properties of the model, show its existence, and consider some special cases and related systems. In Section 3 we state limit results for the main model and its derivatives, mainly restricted to the one-dimensional lattice. Subsequently, in Section 4, we compare our results with those obtained in similar models, e.g. from [2] and [10]. The rest of the paper is devoted to the proofs of the theorems.

2. The model and first results

2.1. The model

Let V be a countable vertex set, and $(a(u, v))_{u,v\in V}$ be a (not necessarily symmetric) Markov kernel from V to V. Additionally, $(b(u, (v, w)))_{u\in V,\, (v,w)\in V\times V}$ is a second Markov kernel from V to $V \times V$. We study an interacting particle system $X = ((X_t(u))_{u \in V})_{t \geq 0}$ with state space $\{0, 1\}^V$, where $X_t(u) \in \{0, 1\}$ is the type at site u at time t. A particle in state 0 is called a *defector* and a particle in state 1 is called a *cooperator*. The dynamics of the interacting particle system, which is a Markov process, is (informally) as follows, for some $\alpha, \gamma \geq 0$.

- *Reproduction.* A particle at site $u \in V$ reproduces with rate $a(u, v)$ to site v, i.e. $X(v)$ changes to $X(u)$. (This mechanism is well known from the voter model.)
- *Selection.* If $X(u) = 0$ (i.e. there is a *defector* at site $u \in V$), it reproduces with additional rate $\alpha a(u, v)$ to site v, i.e. $X(v)$ changes to 0. (A defector has a fitness advantage over the cooperators by this additional chance to reproduce. This mechanism is well known from the biased voter model.)
- *Cooperation.* If $X(u) = 1$ (i.e. there is a *cooperator* at site $u \in V$), the individual at site v (no matter which state it has) reproduces to site w at rate $\gamma b(u, (v, w)) \geq 0$. (A cooperator at site u helps an individual at site v to reproduce to site w.)

Remark 1. (*Interpretation*) (i) *Selection.* Since cooperation imposes an energetic cost on cooperators, the noncooperating individuals can use these free resources for reproduction processes. This leads to a fitness advantage that we describe with the parameter α.

(ii) *Cooperation.* The idea of the cooperation mechanism in our model is that each cooperator supports a neighboring individual, independent of its type, to reproduce to another site according to the Markov kernel b. A biological interpretation for this supportive interaction is a common good produced by cooperators and released to the environment, helping the colony to expand. The corresponding interaction parameter is γ.

We will deal with two situations, depending on whether $b(u, (v, u)) > 0$ or $b(u, (v, u)) = 0$. In the former case, we speak of an altruistic system, since a cooperator at site u can help the particle at site v to kill it. In the latter case, we speak of a cooperative system.

In order to uniquely define a Markov process, we will need the following assumption.

Assumption 1. (Markov kernels.) *The Markov kernels $a(\cdot, \cdot)$ and $b(\cdot, (\cdot, \cdot))$ satisfy*

$$\sum_{u \in V} a(u, v) < \infty \quad \textit{for all } v \in V \tag{1}$$

and

$$\sum_{u, v \in V} b(u, (v, w)) < \infty \quad \textit{for all } w \in V. \tag{2}$$

Remark 2. (*Some special cases.*) A special case is

$$b(u, (v, w)) = a(u, v) a(v, w) \quad \text{for all } u, v, w \in V. \tag{3}$$

Then, (2) is implied by the assumption

$$\sup_{v \in V} \sum_{u \in V} a(u, v) < \infty,$$

which is stronger than (1). We will also deal with a similar case by setting $b(u, (v, u)) = 0$, which means that u cannot help v to replace u. To be more precise, we set

$$b(u, (v, w)) = \begin{cases} a(u, v) \dfrac{a(v, w) \mathbf{1}_{\{w \neq u\}}}{\sum_{w' \neq u} a(v, w')} & \text{if } a(v, u) < 1, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for all } u, v, w \in V. \tag{4}$$

The normalizing sum in the denominator emerges from the exclusion of self-replacement, i.e. (4) is the two-step transition kernel of a self-avoiding random walk.

2.2. Existence and uniqueness of the process

We now become more formal and define the (pre)generator of the process X via its transition rates. Given $X \in \{0, 1\}^V$, the rate of change $c(u, X)$ from X to

$$X^u(v) = \begin{cases} X(v), & v \in V \setminus \{u\}, \\ 1 - X(u), & v = u, \end{cases}$$

is as follows. If $X(u) = 0$ then

$$c(u, X) = \sum_v a(v, u)X(v) + \gamma \sum_v X(v) \sum_w X(w)b(w, (v, u)). \tag{5}$$

If $X(u) = 1$ then

$$c(u, X) = (1 + \alpha) \sum_v a(v, u)(1 - X(v)) + \gamma \sum_v (1 - X(v)) \sum_w X(w)b(w, (v, u)). \tag{6}$$

Here, the first sum in $c(u, X)$ represents the rates triggered by reproduction and selection, whereas the last terms emerge from the cooperation mechanism.

The existence of a unique Markov process corresponding to the transition rates $c(u, X)$ satisfying Assumption 1 is guaranteed by standard theory; see, e.g. [13, Chapter 1]. Precisely, we define the *(pre)generator* Ω of the process through

$$(\Omega f)(X) = \sum_{u \in G} c(u, X)(f(X^u) - f(X)),$$

where $f \in \mathcal{D}(\Omega)$, the domain of Ω, is given by

$$\mathcal{D}(\Omega) := \{f : \{0, 1\}^V \to \mathbb{R} \text{ depends only on finitely many coordinates}\}.$$

We note that $\mathcal{D}(\Omega)$ is dense in $C_b(\{0, 1\}^V)$, the set of bounded continuous functions on $\{0, 1\}^V$, because of the Stone–Weierstrass theorem. We find the following general statement.

Proposition 1. (Existence of a unique Markov process.) *If Assumption 1 holds, the transition rates $c(\cdot, \cdot)$ given in (5) and (6) define a unique Markov process X on $\{0, 1\}^V$. Moreover, the closure $\bar{\Omega}$ of Ω is the generator of X.*

Proof. We need to show that the closure of Ω in $C(\{0, 1\}^V)$ is a generator of a semi-group which then uniquely defines a Markov process (see, e.g. [13, Theorem 1.1.5]). In order to show this, we follow [13, Theorem 1.3.9] and check the following two conditions:

$$\sup_{u \in V} \sup_{X \in \{0,1\}^V} c(u, X) < \infty, \tag{7}$$

$$\sup_{u \in V} \sum_{v \neq u} \widetilde{c}_u(v) < \infty, \tag{8}$$

where $\widetilde{c}_u(v) := \sup\{\|c(u, X_1) - c(u, X_2)\|_{\mathrm{T}} : X_1(w) = X_2(w)$ for all $w \neq v$ measures the dependence of the transition rate $c(u, X)$ of the site $v \in V$ and $\| \cdot \|_{\mathrm{T}}$ denotes the total variation norm.

Both inequalities follow from Assumption 1 and the definition of the transition rates $c(\cdot, \cdot)$. Using these we obtain, for any $X \in \{0, 1\}^V$ and $u \in V$,

$$c(u, X) \le (1+\alpha) \sum_{v \in V} a(v, u) + \gamma \sum_{v, w \in V} b(w, (v, u)) < \infty$$

proving (7). For (8), we note that $\widetilde{c}_u(v) \neq 0$ only when either $a(v, u) > 0$ or $b(w, (v, u)) > 0$ or $b(v, (w, u)) > 0$ for some $w \in V$. Hence, for all $u \in V$, we obtain

$$\begin{aligned} \sum_{v \neq u} \widetilde{c}_u(v) &\le \sum_{v \neq u} \bigg((1+\alpha) a(v, u) + \gamma \sum_{w \in V} b(w, (v, u)) + b(v, (w, u)) \bigg) \\ &\le \sum_{v \in V} (1+\alpha) a(v, u) + 2\gamma \sum_{v, w \in V} b(v, (w, u)) \\ &< \infty, \end{aligned}$$

where we used the inequalities (1) and (2), again proving (8).

Now, using [13, Theorem 1.3.9] we see that the closure of Ω in $C(\{0, 1\}^V)$ is a Markov generator of a Markov semigroup. This completes the proof. □

We can now define the voter model with bias and cooperation.

Definition 1. (*Cooperative/altruistic voter model with bias and cooperation.*) Let $a(\cdot, \cdot)$ be a Markov kernel from V to V satisfying (1) and $b(\cdot, (\cdot, \cdot))$ be a Markov kernel from V to $V \times V$ satisfying (2).

- The (unique) Markov process with transition rates given by (5) and (6) is called the *voter model with bias and cooperation* (VMBC).
- If (3) holds, the VMBC is called the *altruistic voter model with bias and cooperation* (aVMBC).
- If (4) holds, the VMBC is called the *cooperative voter model with bias and cooperation* (cVMBC).

2.3. Unstructured populations

As a first result, we show that the probability for cooperators to die out on a large, complete graph tends to 1 (for $\alpha > 0$). We consider the special case of an *unstructured population*. Therefore, let V^N be the vertex set of a graph with $|V^N| = N$ and

$$a^N(u, v) = \frac{1}{N-1}$$

for $u, v \in V^N$ with $u \neq v$. Due to the global neighborhood it is equally likely to find configurations of the form '101' and '110'. Hence, cooperation events favoring a defector or a cooperator happen with the same rate and, thus, cancel out when looking at the mean field behavior of the system. We will show that defectors always take over the system for large N. It can easily be seen that the aVMBC is dominated by the cVMBC, so it suffices to show extinction of cooperators for the cVMBC, i.e. we have

$$b^N(u, (v, w)) = \frac{\mathbf{1}_{\{u \neq v\}}}{N-1} \frac{\mathbf{1}_{\{v \neq w\}} \mathbf{1}_{\{w \neq u\}}}{(N-1)[(N-2)/(N-1)]} = \frac{1}{(N-1)(N-2)} \mathbf{1}_{\{u, v, w \text{ different}\}}.$$

We prove that in the limit for large N, the frequency of cooperators follows a logistic equation with negative drift, hence cooperators die out. See also [9, Chapter 11].

Proposition 2. (Convergence in the unstructured case.) *Let X^N be a cVMBC on V^N and $S^N := (1/N)\sum_u X^N(u)$ the frequency of cooperators. Then, if $S_0^N \Longrightarrow s_0$ as $N \to \infty$ then*

$$S^N \Longrightarrow S \quad \textit{as } N \to \infty,$$

where S solves the ordinary differential equation

$$\mathrm{d}S = -\alpha S(1-S)$$

with $S_0 = s_0$, independently of γ.

Proof. In order to prove the limiting behavior for $N \to \infty$, we observe that S^N is a Markov process. A calculation of the generator Ω^N applied to some smooth function f yields

$$\begin{aligned}\Omega^N f(s) &= Ns\frac{1-s}{1-1/N}\left(f\left(s+\frac{1}{N}\right)-f(s)\right)\\ &\quad + (1+\alpha)N(1-s)\frac{s}{1-1/N}\left(f\left(s-\frac{1}{N}\right)-f(s)\right)\\ &\quad + \gamma Ns\frac{s-1/N}{1-1/N}\frac{1-s}{1-2/N}\left(f\left(s+\frac{1}{N}\right)-f(s)\right)\\ &\quad + \gamma Ns\frac{1-s}{1-1/N}\frac{s-1/N}{1-2/N}\left(f\left(s\left(-\frac{1}{N}\right)\right)-f(s)\right)\\ &\to -\alpha s(1-s)f'(s) \quad \text{as } N\to\infty.\end{aligned}$$

Applying standard weak convergence results, see, e.g. [9, Theorem 4.8.2], shows the claimed convergence. □

3. Results on the long-time behavior for $V = \mathbb{Z}^d$

Our main goal is to derive the long-time behavior of the VMBC with $V = \mathbb{Z}^d$. In spin-flip systems, results on the ergodic behavior can be obtained by general principles if the process is *attractive*. Thereby, a spin system is called attractive if for two configurations $X, Y \in \{0,1\}^V$ with $X \le Y$ componentwise, the corresponding transition rates c satisfy the following two relations for all $u \in V$:

$$\begin{aligned} X(u) = Y(u) = 0 \quad &\Longrightarrow \quad c(u,X) \le c(u,Y),\\ X(u) = Y(u) = 1 \quad &\Longrightarrow \quad c(u,X) \ge c(u,Y). \end{aligned} \tag{9}$$

However, the VMBC is not attractive for $\gamma > 0$. Indeed, consider the simple case when $V = \{u, v, w\}$ with Markov kernels

$$a(u,v) = a(v,w) = a(w,u) = 1 \quad \text{and} \quad b(u,(v,w)) = a(u,v)a(v,w).$$

Then, let $X = (001)$ and $Y = (101)$ (i.e. $X(u) = 0$, $Y(u) = 1$, $X(v) = Y(v) = 0$, $X(w) = Y(w) = 1$) and note that $X \le Y$, but

$$c(w,X) = 1+\alpha < 1+\alpha+\gamma = c(w,Y).$$

This shows that (9) is not satisfied at $w \in V$. Hence, proofs for the long-time behavior require other strategies that do not rely on attractiveness of the process.

Before we state our main results we define what we mean by extinction and clustering.

Definition 2. (*Extinction and clustering.*) (i) We say that in the VMBC-process $(X_t)_{t\geq 0}$ type $i \in \{0, 1\}$ *dies out* if

$$\mathbb{P}\Big(\lim_{t\to\infty} X_t(u) = 1 - i\Big) = 1 \quad \text{for all } u \in V.$$

(ii) We say that the VMBC process *clusters* if, for all $u, v \in V$,

$$\lim_{t\to\infty} \mathbb{P}(X_t(u) = X_t(v)) = 1.$$

We will use $V = \mathbb{Z}^d$ and nearest neighbor interaction via the kernels a and b. In this case, we have, for all $u, v, w \in \mathbb{Z}^d$ with $|u - v| = |w - v| = 1$,

$$a(u, v) = \frac{1}{2d}, \qquad b(u, (v, w)) = \frac{1}{(2d)^2}, \quad \text{for the aVMBC,} \tag{10}$$

and

$$a(u, v) = \frac{1}{2d}, \qquad b(u, (v, w)) = \frac{1}{2d(2d-1)} \mathbf{1}_{\{u\neq w\}}, \quad \text{for the cVMBC.}$$

We say that (the distribution of) a $\{0, 1\}^{\mathbb{Z}^d}$-valued random configuration X is *nontrivial* if $\mathbb{P}(X(u) = 0 \text{ for all } u), \mathbb{P}(X(u) = 1 \text{ for all } u) < 1$. Furthermore, we call X *translation invariant* if $(X(u_1), \dots, X(u_n)) \stackrel{\mathrm{D}}{=} (X(u_1 + v), \dots, X(u_n + v))$ for all $n \in \mathbb{N}, u_1, \dots, u_n, v \in \mathbb{Z}^d$, where '$\stackrel{\mathrm{D}}{=}$' denotes equality in distribution. If the VMBC model is started in a translation invariant configuration $X_0 \in \{0, 1\}^{\mathbb{Z}^d}$, the configuration X_t is translation invariant due to the homogeneous model dynamics.

Now we can state our main results. For cVMBC, we distinguish between the case $\alpha > \gamma$ where we can state a convergence result in all dimensions $d \geq 1$, the case $\gamma > \alpha$, and the case $\gamma = \alpha$. In the last two cases, the method of proof is applicable only in dimension $d = 1$.

Theorem 1. (cVMBC limits.) *Let $V = \mathbb{Z}^d$ and $a(\cdot, \cdot)$ be the nearest neighbor random walk kernel, and X be the cVMBC with $\alpha, \gamma \geq 0$ starting in some nontrivial translation invariant initial configuration.*

(i) *If $d \geq 1$ and $\alpha > \gamma$, the cooperators die out.*

(ii) *If $d = 1$ and $\gamma > \alpha$, the defectors die out.*

(iii) *If $d = 1$ and $\gamma = \alpha$, the process clusters.*

The proof of Theorem 1 can be found in Section 6. Briefly, for $\alpha > \gamma$, we will use a comparison argument with a biased voter model; see Definition 3. For $\gamma > \alpha$ and $d = 1$, however, we prove the convergence result with the help of a cluster-size process that takes the special form of a one-dimensional jump process. As we will see, for $\gamma > \alpha$, a cluster of cooperators has a positive probability to survive and expand to ∞, which will then yield the result. Unfortunately, due to the simple description of such a cluster in one dimension, this argument cannot be extended to higher dimensions. However, resorting to some simulation results for $d = 2$ and $d = 3$, we see a similar behavior (with a different threshold) such as in $d = 1$; see Figure 1. For higher dimensions, spatial correlations between sites are weaker, reducing the impact of clusters on the evolution of the system. This, in turn, leads to a reduced chance of survival of cooperators.

For the aVMBC, we can state a threshold only when cooperators die out.

Theorem 2. (aVMBC limits.) *Let* $V = \mathbb{Z}^d$ *and* $a(\cdot,\cdot)$ *be the nearest neighbor random walk kernel, and* X *be the aVMBC with* $\alpha, \gamma \geq 0$ *starting in some nontrivial translation invariant initial configuration.*

(i) *If* $d \geq 1$ *and* $\alpha > \gamma(d-1)/d$, *the cooperators die out. In particular, for* $d = 1$, *the cooperators die out if* $\alpha > 0$ *independently of* γ.

(ii) *If* $d = 1$, *the process is equal to the cVMBC with parameters* $\alpha + \gamma/2$ *and* $\gamma/2$ *in distribution. In particular, if* $\gamma > \alpha = 0$, *the process clusters.*

The proof of the Theorem can be found in Section 6. Again, for $\alpha > \gamma(d-1)/d$, we can use a comparison argument with the biased voter model. However, it remains an open question whether cooperators in the aVMBC have a positive probability of survival in any dimension. On the one hand, the difference between the aVMBC and the cVMBC becomes smaller in high dimensions and Figure 1 suggests survival of cooperators for the cVMBC in all dimensions, if γ is large enough. On the other hand, clustering is usually more difficult in higher dimensions but the cooperators can only survive due to clustering. First simulation results for $d = 2$ and $d = 3$ show that survival of cooperators is unlikely.

Remark 3. (*Cooperation only among cooperators.*) Another cooperation mechanism we might consider arises if cooperators help only other cooperators, i.e. the cells recognize related cells. In ecological literature, this behavior is called kin-recognition or kin-discrimination; see [16] for an overview. As to the theoretical behavior of the model, this changes the transition rate in (6), i.e. if $X(u) = 1$ then

$$c(u, X) = (1+\alpha)\sum_v a(v,u)(1 - X(v)).$$

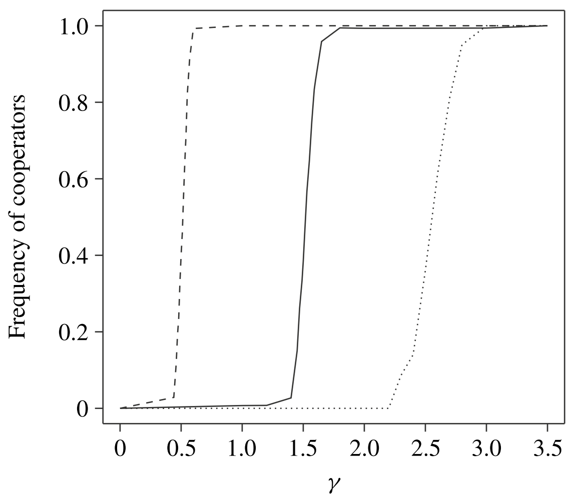

Figure 1: Relative frequencies of cooperators after 100,000 transitions of the cVMBC on a 1,000 sites torus in one dimension (*dashed line*), a 40×40 sites torus in two dimensions (*solid line*) and a $12 \times 12 \times 12$ sites torus in three dimensions (*dotted line*). The initial configuration was a Bernoulli-product measure with probability 0.5 and the selection rate α was set to 0.5. We suspect that the smaller slope in three dimensions is a finite-number effect.

Here, cooperators are less likely to die and, hence, this process dominates the cVMBC. In particular, for translation invariant initial conditions, defectors die out for $\gamma > \alpha$ in one dimension. Moreover, as can be seen from a calculation similar as in the proof of Lemma 1, a biased voter model, where type 0 is favored, still dominates this process for $\alpha > \gamma$. Hence, we also have that cooperators die out in this case and the same results as in Theorem 1 hold.

Since cooperators always die out in $d = 1$ for the aVMBC (as long as $\alpha > 0$), we focus on the cVMBC in the sequel. We state some results if the starting configuration is not translation invariant, but contains only a finite number of cooperators or defectors.

Theorem 3. (Finite initial configurations.) *Let $V = \mathbb{Z}$ and $a(\cdot, \cdot)$ be the nearest neighbor random walk kernel, and X be the cVMBC with $\alpha, \gamma \geq 0$. Let X_0 contain either finitely many defectors or finitely many cooperators (i.e. $X_0 = \mathbf{1}_A$ or $X_0 = 1 - \mathbf{1}_A$ for some finite $A \subseteq V$).*

(i) *The process clusters.*

(ii) *If $\alpha \geq \gamma$ and X_0 contains finitely many cooperators, the cooperators die out.*

(iii) *If $\gamma \geq \alpha$ and X_0 contains finitely many defectors, the defectors die out.*

Remark 4. (*Starting with a single particle.*) A particularly simple initial condition is given if $|A| = 1$. In the case that there is initially only a single cooperator, we note that the size of the cluster of cooperators $(C_t)_{t\geq 0}$ is a birth–death process which jumps from C to

$$C + 1 \text{ at rate } \mathbf{1}_{\{C>0\}} + \gamma\, \mathbf{1}_{\{C\geq 2\}}, \qquad C - 1 \text{ at rate } (1+\alpha)\, \mathbf{1}_{\{C>0\}}\,.$$

Conversely, if there is only a single defector, the size of the cluster of defectors $(D_t)_{t\geq 0}$ is a birth–death process which jumps from D to

$$D + 1 \text{ at rate } (1+\alpha)\, \mathbf{1}_{\{D>0\}} + \gamma\, \mathbf{1}_{\{D=1\}}, \qquad D - 1 \text{ at rate } (1+\gamma)\, \mathbf{1}_{\{D>0\}}\,.$$

Hence, either cooperators or defectors die out, depending on whether $(C_t)_{t\geq 0}$ (or $(D_t)_{t\geq 0}$) hits 0 or not.

The proof of Theorem 3 is given in Section 6. Note that the only situations where the process does not converge to a deterministic configuration in this setting are the cases where $\gamma > \alpha$ ($\alpha > \gamma$) and the process starts with finitely many cooperators (defectors). Here, the limit distribution is a linear combination of the invariant measures $\delta_{\underline{0}}$ and $\delta_{\underline{1}}$ which basically means that we observe clustering, which is statement (i) above.

4. Comparison to results from [2] and [10]

In this section we compare our results on the cVMBC to those obtained by Blath and Kurt in [2] and the system introduced by Evilsizor and Lanchier in [10]. We choose these two models since both have mechanisms favoring one type, while a second type is favored only if it occurs in a cluster.

4.1. Comparison to [2]

One model studied is the cooperative caring double-branching annihilating random walk (ccDBARW) on the integer lattice $\mathbb{Z}$. Particles migrate to neighboring sites with rate m and annihilate when meeting another particle. (Note that this mechanism favors the unoccupied state.)

The double-branching events happen with rate $1 - m$. Here, the authors restrict branching to particles with an occupied neighboring site and such particles branch to the next unoccupied site to the left and to the right. (That is, if a cluster of size ≥ 2 already exists, the branching mechanism extends the cluster.) Their result about this process, starting in a finite configuration (see Theorem 2.4 in the paper), states that for $m < \frac{1}{2}$, particles survive with positive probability, whereas for $m > \frac{2}{3}$, particles die out almost surely.

Although Blath and Kurt discuss only the case of a finite initial configuration, the results are in line with our findings. If the mechanism to favor enlargement of existing clusters (cooperation in our case and cooperative branching in their case) is too weak, type 0 (or the unoccupied state) wins. Importantly, in both models, enlargement of existing clusters can be strong enough in order to outcompete the beneficial (or unoccupied) type.

4.2. Comparison to [10]

The model studied in [10]—called the death–birth updating process—emerges from a game theoretic model with two strategies. This means that transition rates are derived from a 2×2 payoff matrix with entries a_{ij} for $i, j \in \{1, 2\}$ representing the payoff obtained by a particle of type i due to interacting with a particle of type j. Now, a particle dies with rate 1 and is replaced by a particle in its neighborhood proportional to its fitness which is determined by the values of the payoff matrix. The neighborhood is given by blocks of radius R. The authors analyze this model in different settings. They call a strategy i selfish if $a_{ii} > a_{ji}$ for $j \neq i$ (i.e. the payoff having strategy i as opponent is larger if one has the same strategy i) and altruistic if $a_{ii} < a_{ji}$. Again, in a nonspatial version of this game, selfish strategies always outcompete altruistic strategies. Noting that selfish strategies seem to be fitter, altruistic strategies might become favorable if they form a large cluster since altruists might have a high payoff. As the results in [10] show (see their Figure 2), there are parameter regions—in particular in a spatial prisoner's dilemma—where altruists can outcompete selfish strategies.

Clearly, the cVMBC is a much simpler model than the death–birth updating process. This is seen in the results, since [10] show parameter combinations with coexistence for the death–birth updating process, but our results never show coexistence for the cVMBC. However, as in our findings for the cVMBC, [10] find that types unfavorable in a nonspatial context can indeed win in all dimensions. Unfortunately, they can give bounds only on the phase transition in their model, while we have seen that $\alpha = \gamma$ is a sharp threshold.

5. Preliminaries

Here, we provide some useful results for the proofs of our theorems. In particular, we provide a comparison with a biased voter model in Section 5.1 and a particular jump process in Section 5.2.

5.1. Comparison results

In cases where $\alpha > \gamma$, it is possible to prove a stochastic domination of the VMBC by a biased voter model. The precise statements will be given below. But first, we define this process, which was introduced by Williams and Bjernkes in [18] and first studied by Bramson and Griffeath in [3].

Definition 3. (*Biased voter model.*) The biased voter model with bias $\beta \ge -1$ and $\delta \ge -1$ is a spin system $\widetilde{X}$ with state space $\{0,1\}^V$ and transition rates as follows. If $\widetilde{X}(u) = 0$ then

$$\widetilde{c}(u, \widetilde{X}) = (1+\beta) \sum_v a(v,u) X(v).$$

If $\widetilde{X}(u) = 1$ then

$$\widetilde{c}(u, \widetilde{X}) = (1+\delta) \sum_v a(v,u)(1 - X(v)).$$

Remark 5. (*Long-time behavior of the biased voter model.*) The long-time behavior of the biased voter model is quite simple. In [3], the limit behavior of the biased voter model in $V = \mathbb{Z}^d$ with nearest neighbor interactions was studied. Generalizations to the case of d-regular trees for $d \ge 3$ can be found in [14]. We restate the results for $V = \mathbb{Z}^d$.

Let $\widetilde{X}$ be a biased voter model with bias $\beta > -1$ and $\delta > -1$ as introduced in Definition 3. For any configuration $X_0 \in \{0,1\}^{\mathbb{Z}^d}$ with infinitely particles of each type, it holds that the type with less bias dies out, i.e.

(i) if $\beta > \delta$, type 0 dies out (i.e. $\mathbb{P}(\lim_{t\to\infty} \widetilde{X}_t(u) = 1) = 1$ for all $u \in V$);

(ii) if $\delta > \beta$, type 1 dies out (i.e. $\mathbb{P}(\lim_{t\to\infty} \widetilde{X}_t(u) = 0) = 1$ for all $u \in V$).

Lemma 1. (cVMBC is equal to or less than the biased voter model.) *Let X be a cVMBC with bias α and cooperation coefficient γ, and $\widetilde{X}$ a biased voter model with bias γ and α. Then, if $b(\cdot, (\cdot,\cdot))$ satisfies $\sum_u b(u,(v,w)) \le a(v,w)$, and $X_0 \le \widetilde{X}_0$, it is possible to couple X and $\widetilde{X}$ such that $X_t \le \widetilde{X}_t$ for all $t \ge 0$.*

Proof. We need to show (see [13, Theorem 3.1.5]) that, for $X \le \widetilde{X}$,

$$\begin{aligned} &\text{if } X(u) = \widetilde{X}(u) = 0 \quad \text{then } c(u,X) \le \widetilde{c}(u,\widetilde{X}), \\ &\text{if } X(u) = \widetilde{X}(u) = 1 \quad \text{then } c(x,X) \ge \widetilde{c}(u,\widetilde{X}). \end{aligned} \tag{11}$$

We start with the first assertion and write

$$\begin{aligned} c(u,X) &= \sum_v a(v,u)X(v) + \gamma \sum_v X(v) \sum_w X(w) b(w,(v,u)) \\ &\le \sum_v a(v,u)X(v) + \gamma \sum_v X(v) a(v,u) \\ &\le (1+\gamma) \sum_v a(v,u)\widetilde{X}(v) \\ &= \widetilde{c}(u,\widetilde{X}), \end{aligned}$$

and for the second inequality, we have

$$\begin{aligned} c(u,X) &= (1+\alpha) \sum_v a(v,u)(1-X(v)) + \gamma \sum_v (1-X(v)) \sum_{w,v} X(w) b(w,(v,u)) \\ &\ge (1+\alpha) \sum_v a(v,u)(1-X(v)) \\ &\ge (1+\alpha) \sum_v a(v,u)(1-\widetilde{X}(v)) \\ &= \widetilde{c}(u,\widetilde{X}). \end{aligned}$$

□

Next, we focus on the aVMBC in the $V = \mathbb{Z}^d$ case and a symmetric, nearest-neighbor random walk kernel.

Lemma 2. (aVMBC is equal to or less than the biased voter model.) *Let $V = \mathbb{Z}^d$, $a(\cdot, \cdot)$ be the nearest-neighbor random walk kernel defined in (10), X be an aVMBC with bias α and cooperation coefficient γ, and $\widetilde{X}$ a biased voter model with bias $\gamma(2d-1)/(2d)$ and $\alpha+\gamma/(2d)$. Then, if $X_0 \le \widetilde{X}_0$, it is possible to couple X and $\widetilde{X}$ such that $X_t \le \widetilde{X}_t$ for all $t \ge 0$.*

Proof. Again, we need to show that for $X \le \widetilde{X}$ the inequalities in (11) hold. We start with the first assertion and write, using the fact that $X(u) = 0$,

$$\begin{aligned} c(u, X) &= \sum_v a(v, u)X(v) + \gamma \sum_v X(v) \sum_w X(w)a(v, w)a(v, u) \\ &\le \sum_v a(v, u)X(v) + \gamma \sum_v X(v)a(v, u) \sum_{w \ne u} a(v, w) \\ &\le \left(1 + \gamma \frac{2d-1}{2d}\right) \sum_v a(v, u)\widetilde{X}(v) \\ &= \widetilde{c}(u, \widetilde{X}), \end{aligned}$$

and for the second inequality, now using $X(u) = 1$, we have

$$\begin{aligned} c(u, X) &= (1+\alpha) \sum_v a(v, u)(1 - X(v)) + \gamma \sum_v (1 - X(v)) \sum_w X(w)a(w, v)a(v, u) \\ &\ge (1+\alpha) \sum_v a(v, u)(1 - X(v)) + \gamma \sum_v (1 - X(v))a(u, v)a(v, u) \\ &= \left(1 + \alpha + \frac{\gamma}{2d}\right) \sum_v a(v, u)(1 - X(v)) \\ &\ge \left(1 + \alpha + \frac{\gamma}{2d}\right) \sum_v a(v, u)(1 - \widetilde{X}(v)) \\ &= \widetilde{c}(u, \widetilde{X}). \end{aligned}$$

This yields the statement. □

5.2. A result on a jump process

In the proof of Theorem 1, we will use the dynamics of the size of a cluster of cooperators and rely on a comparison of this cluster-size process with a certain jump process (which jumps downward by at most one and upwards by at most two). The following proposition will be needed.

Proposition 3. (A jump process.) *Let $(\mu(t))_{t\ge0}$, $(\lambda_1(t))_{t\ge0}$, and $(\lambda_2(t))_{t\ge0}$ be an $\mathbb{R}_+$-valued càdlàg stochastic processes, adapted to some filtration $(\mathcal{F}_t)_{t\ge0}$, which satisfy*

$$\begin{aligned} &\lambda_1(t) + 2\lambda_2(t) - \mu(t) > \varepsilon > 0 \quad \text{for some } \varepsilon, \\ &\lambda_1(t) + \lambda_2(t) + \mu(t) < C \quad \text{for some } C > 0. \end{aligned} \tag{12}$$

In addition, let $(C_t)_{t\ge0}$ be a $\mathbb{Z}$-valued $(\mathcal{F}_t)_{t\ge0}$ Markov jump process, which jumps at time t from x to

$$x - 1 \text{ at rate } \mu(t), \qquad x + 1 \text{ at rate } \lambda_1(t), \qquad x + 2 \text{ at rate } \lambda_2(t).$$

Then,

(i) $C_t \to \infty$ *as* $t \to \infty$ *almost surely, and*

(ii) $\mathbb{P}(T_1 = \infty) > 0$ *for* $C_0 = 2$ *and* $T_1 := \inf\{t : C_t = 1\}$.

Proof. In the case of time-homogeneous rates, i.e. constant μ, λ_1, and λ_2, the assertion is an immediate consequence of the law of large numbers. We prove the general case by using martingale theory. We assume without loss of generality that $\lambda_1(t) + \lambda_2(t) + \mu(t) = 1$ for all $t \geq 0$. (Otherwise, we use a time rescaling. Note that this rescaling is bounded by assumption (12) and, therefore, $C_t \to \infty$ as $t \to \infty$ holds if and only if it holds for the rescaled process.)

We start by showing that there exists $a_c > 0$ such that, for all $a \in (0, a_c)$, the process $(\exp(-aC_t))_{t\geq 0}$ is a positive $(\mathcal{F}_t)_{t\geq 0}$-supermartingale. For this, consider the (time-dependent) generator of the process $(C_t)_{t\geq 0}$ applied to the function $f(x) = \exp(-ax)$ which yields, at time t,

$$\begin{aligned}(G_t^{\mathcal{C}} f)(x) &= \lambda_1(t)\exp(-a(x+1)) + \lambda_2(t)\exp(-a(x+2)) \\ &\qquad + \mu(t)\exp(-a(x-1)) - \exp(-ax) \\ &= \exp(-ax)(\lambda_1(t)\exp(-a) + \lambda_2(t)\exp(-2a) + \mu(t)\exp(a) - 1) \\ &= \exp(-ax)g_t(a)\end{aligned}$$

for $g_t(a) := \lambda_1(t)\exp(-a) + \lambda_2(t)\exp(-2a) + \mu(t)\exp(a) - 1$. Noting that, for all t, we have $g_t(0) = 0$ and

$$\frac{\partial g_t}{\partial a}(0) = -\lambda_1(t) - 2\lambda_2(t) + \mu(t) < -\varepsilon$$

by (12), we find $a_c > 0$ such that $g_t(a) < 0$ for all $0 < a < a_c$ and all $t \geq 0$, which means that $(\exp(-aC_t))_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$-supermartingale. By the martingale convergence theorem, it converges almost surely and—since the sum of rates is bounded away from 0—the only possible almost sure limit is 0. Now, (i) follows since $C_t \to \infty$ as $t \to \infty$ if and only if $\exp(-aC_t) \to 0$ as $t \to \infty$ for some $a > 0$. For (ii), the process $(\exp(-aC_{t\wedge T_1}))_{t\geq 0}$ is a nonnegative supermartingale by optional stopping. Let us assume that $T_1 < \infty$ almost surely, which occurs if and only if $C_{t\wedge T_1} \to 1$ as $t \to \infty$ almost surely. Then, using the optional stopping theorem, we obtain, with $C_0 = 2$,

$$\begin{aligned}\exp(-2a) &= \mathbb{E}[\exp(-aC_0)] \\ &\geq \lim_{t\to\infty} \mathbb{E}[\exp(-aC_{t\wedge T_1})] \\ &= \mathbb{E}[\lim_{t\to\infty} \exp(-aC_{t\wedge T_1})] \\ &= \exp(-a),\end{aligned}$$

a contradiction since $a > 0$. Thus, we have $\mathbb{P}(T_1 = \infty) > 0$. □

6. Proofs

Here, we will show our main results.

Proof of Theorem 1. For (i), we have $\alpha > \gamma$. The assertion is a consequence of the coupling with the biased voter model from Lemma 1 (with bias γ and α). Since the biased voter model dominates the cVMBC and type 1 dies out in the biased voter model (5), the same holds for the cVMBC.

The proof of (ii) is more involved. We have to show that cooperators survive almost surely when started in a nontrivial translation invariant configuration. Therefore, we analyze an arbitrary cluster of cooperators and show that the size of such a cluster has a positive probability to diverge to ∞. Note that the flanking regions of a cluster of cooperators can have three different forms:

$$\begin{array}{ccc} \text{Case A} & \text{Case B} & \text{Case C} \\ 0\underbrace{01\cdots10}_{\text{Cluster of cooperators}}0 & 1\underbrace{01\cdots10}_{\text{Cluster of cooperators}}1 & 0\underbrace{01\cdots10}1 \text{ or } 1\underbrace{01\cdots10}0 \\ & & \text{Cluster of cooperators} \end{array}$$

These are the only possible environments a cluster of cooperators can encounter in one dimension. Note that a cluster can also only consist of a single cooperator. The dynamics of the cluster size depends on the environment. Precisely, by the dynamics of the process, we obtain the following. A cluster of size $x > 1$,

$$\begin{aligned} &\text{in case A, jumps to} \begin{cases} y = x+1 & \text{at rate } 1+\gamma, \\ y = x-1 & \text{at rate } 1+\alpha; \end{cases} \\ &\text{in case B, jumps to} \begin{cases} y \geq x+2 & \text{at rate at least } 2+\gamma, \\ y = x-1 & \text{at rate } 1+\alpha+\gamma; \end{cases} \\ &\text{in case C, jumps to} \begin{cases} y \geq x+2 & \text{at rate at least } 1+\tfrac{1}{2}\gamma, \\ y = x+1 & \text{at rate } \tfrac{1}{2}(1+\gamma), \\ y = x-1 & \text{at rate } 1+\alpha+\tfrac{1}{2}\gamma. \end{cases} \end{aligned} \tag{13}$$

Under the assumptions of Theorem 1, let $(V_t)_{t\geq 0}$ be a stochastic process representing the cluster of cooperators that is closest to the origin and contains at least two cooperators. (If there is no such cluster at time 0, wait for some time $\varepsilon > 0$ and pick the cluster then.) We will show that

$$\mathbb{P}(V_t \uparrow \mathbb{Z}) > 0. \tag{14}$$

For this, we compare $|V| = (|V_t|)_{t\geq 0}$ with a jump process $(\widetilde{V}_t)_{t\geq 0}$ as in Proposition 3, where the jump rates at times t are given as follows:

$$\begin{aligned} &\text{in case A,} \begin{cases} \lambda_1(t) = 1+\gamma, \\ \lambda_2(t) = 0, \\ \mu(t) = 1+\alpha; \end{cases} \\ &\text{in case B,} \begin{cases} \lambda_1(t) = 0, \\ \lambda_2(t) = 2+\gamma, \\ \mu(t) = 1+\alpha+\gamma; \end{cases} \\ &\text{in case C,} \begin{cases} \lambda_1(t) = \tfrac{1}{2}(1+\gamma), \\ \lambda_2(t) = 1+\tfrac{1}{2}\gamma, \\ \mu(t) = 1+\alpha+\tfrac{1}{2}\gamma. \end{cases} \end{aligned}$$

Moreover, this process is stopped when reaching 1. By the comparison in (13), we see that we can couple $|V|$ and $\widetilde{V}$ such that $\widetilde{V} \leq |V|$, at least until $\widetilde{V}$ reaches 1. Since the jump rates of $\widetilde{V}$ indeed satisfy $2\lambda_2(t) + \lambda_1(t) - \mu(t) > \varepsilon > 0$ for all times $t \geq 0$, we find $\widetilde{V}_t \to \infty$ as $t \to \infty$ with positive probability which implies that $\mathbb{P}(|V_t| \to \infty) > 0$ as $t \to \infty$ holds as well. Still, we need to make sure that the cluster does not wander to $\pm\infty$. For this, consider both boundaries of the cluster if it has grown to a large extent. The right boundary is again bounded from below by a jump process of the form as in Proposition 3 with $\lambda_1(t) = \frac{1}{2}(1+\gamma), 0$; $\lambda_2(t) = 0, 1 + \frac{1}{2}\gamma$ and $\mu(t) = \frac{1}{2}(1+\alpha), \frac{1}{2}(1+\alpha+\gamma)$ for the cases A and B (note that the right boundary alone of case C is already captured by the right boundaries of the cases A and B). So, again, we see from Proposition 3 that the right border of the cluster goes to ∞ with positive probability. The same holds for the left border of the cluster which tends to $-\infty$. Therefore, we have shown (14).

Now we use (14) to show that defectors indeed go extinct. Note that, from the argument given above, the probability of survival of a cluster of cooperators depends on the environment, but is bounded away from 0 by some $p > 0$. We start at time 0 with a cluster of cooperators which has at least probability p to survive as proved above. In the case that it survives we are done, otherwise it goes extinct in finite time and has at most merged with finitely many other cooperating clusters until then. Thus, at this extinction time we can choose another cluster of cooperators that exists due to the translation invariance of the starting configuration. This cluster again has a probability of survival of at least p, independently of the history of the interacting particle system. This allows for a Borel–Cantelli argument showing that when repeating these steps arbitrarily often eventually one of the cooperating clusters survives. This happens at the latest after a geometrically distributed number of attempts and, thus, in finite time. Hence, we have $\mathbb{P}(\lim_{t\to\infty} X_t(u) = 1) = 1$ for all u and we are done.

For (iii), in order to prove clustering in the $\alpha = \gamma > 0$ case, there are actually two proofs. One relies on the dual lattice and the study of the process of cluster interfaces, which performs annihilating random walks. This technique would actually show clustering for all parameters α and γ. However, we show clustering by studying the probability of finding a cluster edge in the special case $\alpha = \gamma$. Clustering for the other parameter configurations was already shown in (i) and (ii) since extinction also implies clustering of the process.

For our method, we write $p_t(i_0 \cdots i_k) := \mathbb{P}(X_t(0) = i_0 \cdots X_t(k) = i_k)$ for $i_0, \cdots, i_k \in \{0, 1\}$ and $k = 0, 1, 2, \ldots$. We have to show that

$$p_t(10) \to 0 \quad \text{as } t \to \infty, \qquad p_t(01) \to 0 \quad \text{as } t \to \infty \tag{15}$$

since then—by translation invariance—every configuration carrying both types has vanishing probability for $t \to \infty$.

We start with the dynamics of $p_t(1)$, which reads (recall that $\alpha = \gamma$)

$$\begin{aligned}
\frac{\partial p_t(1)}{\partial t} &= \frac{1}{2}(p_t(10) + p_t(01)) + \frac{\gamma}{2}(p_t(110) + p_t(011)) \\
&\quad - \frac{1+\alpha}{2}(p_t(10) + p_t(01)) - \gamma p_t(101) \\
&= -\frac{\alpha}{2}(p_t(10) + p_t(01)) + \frac{\alpha}{2}(p_t(10) + p_t(01) - 2p_t(010)) - \gamma p_t(101) \\
&= -\gamma(p_t(101) + p_t(010)) \\
&\leq 0.
\end{aligned}$$

Since $p_t(1) \in [0, 1]$, this probability has to converge for $t \to \infty$, hence $\partial p_t(1)/\partial t \to 0$ as $t \to \infty$, and, therefore,

$$p_t(101) \to 0 \quad \text{as } t \to \infty, \qquad p_t(010) \to 0 \quad \text{as } t \to \infty. \tag{16}$$

Now, consider the dynamics of $p_t(11)$, which is

$$\begin{aligned}\frac{\partial p_t(11)}{\partial t} &= p_t(101) + \frac{\gamma}{2}(p_t(1101) + p_t(1011)) \\ &\quad - \frac{1+\alpha}{2}(p_t(110) + p_t(011)) - \frac{\gamma}{2}(p_t(1011) + p_t(1101)) \\ &= p_t(101) - \frac{1+\alpha}{2}(p_t(110) + p_t(011)).\end{aligned}$$

Since we know that $p_t(101) \to 0$ as $t \to \infty$ by (16), and since $p_t(11) \in [0, 1]$, we also have

$$p_t(110) \to 0 \quad \text{as } t \to \infty, \qquad p_t(011) \to 0 \quad \text{as } t \to \infty.$$

We now conclude with

$$\begin{aligned}p_t(10) &= p_t(010) + p_t(110) \to 0 \quad \text{as } t \to \infty, \\ p_t(01) &= p_t(010) + p_t(011) \to 0 \quad \text{as } t \to \infty,\end{aligned}$$

which shows (15). □

Proof of Theorem 2. (i) We use the comparison with the biased voter model from Lemma 2. Therefore, we have $\alpha > \gamma(d-1)/d$ if and only if $\alpha + \gamma/(2d) > \gamma(2d-1)/(2d)$. Since, for this choice of parameters, type 1 goes extinct in the biased voter model which dominates the aVMBC and we are done.

(ii) For $d = 1$ and the nearest neighbor random walk, the altruistic mechanism is such that a configuration 01 (or 10) turns into 00 at rate $\alpha/2 + \gamma/4$. The same holds for the cVMBC with selection rate $\alpha + \gamma/2$. In addition, 110 (or 011) turns to 111 at rate $\gamma/2$, which is the same as for the cVMBC with cooperation parameter γ. This shows that the transition rates for the altruistic process $\widetilde{X}$ satisfy the following. If $\widetilde{X}(u) = 0$ then

$$c(u, \widetilde{X}) = \frac{1}{2} \sum_{v:\, |v-u|=1} \widetilde{X}(v) + \frac{\gamma}{4} \sum_{v:\, |v-u|=1} \widetilde{X}(v) \sum_{w:\, |w-v|=1, w\neq u} \widetilde{X}(w).$$

If $\widetilde{X}(u) = 1$ then

$$c(u, \widetilde{X}) = \frac{1+\alpha+\gamma/2}{2} \sum_{v:\, |v-u|=1} (1-\widetilde{X}(v)) + \frac{\gamma}{4} \sum_{v:\, |v-u|=1} (1-\widetilde{X}(v)) \sum_{w:\, |w-v|=1, w\neq u} \widetilde{X}(w).$$

These resemble the transition rates of a cVMBC with selection rate $\alpha + \frac{1}{2}\gamma$ and cooperation rate $\frac{1}{2}\gamma$; see also (5) and (6). In particular, clustering follows from Theorem 1(iii). □

Proof of Theorem 3. At time t, let N_t be the number of finite clusters in X_t with sizes $C_t^1, \dots, C_t^{N_t}$. If the process starts with finitely many defectors (cooperators), $C_t^1, C_t^3, C_t^5, \dots$ are sizes of clusters of defectors (cooperators), and $C_t^2, C_t^4, \dots$ are sizes of clusters of cooperators (defectors). Note that $(N_t, C_t^1, \dots, C_t^{N_t})_{t\geq 0}$ is a Markov process. We will show that the following hold.

(a) Either $N_t \to 0$ as $t \to \infty$ or $N_t \to 1$ as $t \to \infty$.

(b) In Theorems 3(ii) and 3(iii), $N_t \to 0$ as $t \to \infty$.

(c) If $N_t \to 1$ as $t \to \infty$ then $C_t^1 \to \infty$ as $t \to \infty$.

Note that (a) and (c) together imply Theorem 3(i), i.e. X clusters in all cases. Of course, (b) implies Theorems 3(ii) and 3(iii).

(a) The process $N = (N_t)_{t\geq 0}$ is nonincreasing and bounded from below by 0, so convergence of N is certain. We assume that $N_0 = n \geq 3$. Note that N_0 is odd and remains so until it hits 1 from where it may or may not jump to 0. In order to prove the claim, we show that the hitting time $\inf\{s\colon N_s < n\}$ is finite almost surely. For this, it suffices to show that

$$T := \inf\{s\colon C_s^k = 1 \text{ for some } 1 \leq k \leq N_s\} < \infty \tag{17}$$

almost surely, since by time T, some cluster has size 1 and there is a positive chance that N decreases at the next transition. If N does not decrease, there is the next chance after another finite time and eventually N will decrease.

If $\alpha \geq \gamma$, consider the size C_t of a cluster of cooperators. Before time T, all clusters have size at least 2, so C_t jumps

$$\text{from } C \text{ to } C+1 \text{ at rate } 1+\gamma, \qquad \text{from } C \text{ to } C-1 \text{ at rate } 1+\alpha,$$

hence, $(C_{t\wedge T})_{t\geq 0}$ is dominated by a symmetric random walk with jump rate $1+\alpha$, stopped when hitting 1, which implies that $T < \infty$ almost surely due to the recurrence of the symmetric random walk in one dimension. If $\gamma \geq \alpha$, the same argument shows that $T < \infty$ if the role of cooperators and defectors is exchanged. Hence, we have proved (17) and (a) is shown.

(b) If $N_t \to 1$ as $t \to \infty$ and $\alpha \geq \gamma$, the remaining finite cluster must consist of defectors (since the argument used in (a) shows that a finite cluster of cooperators would die out). Therefore, in Theorem 3(ii), we must have $N_t \to 0$ as $t \to \infty$. If $\gamma \geq \alpha$, the remaining finite cluster consists of cooperators for the same reason. Hence, in Theorem 3(iii), we must have $N_t \to 0$ as $t \to \infty$. Thus, we have shown (b).

(c) As just argued in (a) and (b), if $N_t \to 1$ as $t \to \infty$, the remaining finite cluster must contain the stronger type, i.e. defectors for $\alpha > \gamma$ and cooperators for $\gamma > \alpha$. The size of the remaining finite cluster therefore is a biased random walk which goes to ∞ on $\{N_t \to 1\}$ as $t \to \infty$ and the result follows. □

Acknowledgements

We would like to thank two anonymous referees for a thorough reading and useful comments which improved the manuscript. This research was supported by the DFG through grant Pf-672/5-1.

References

[1] Archetti, M. and Scheuring, I. (2012). Review: game theory of public goods in one-shot social dilemmas without assortment. *J. Theoret. Biol.* **299,** 9–20.

[2] Blath, J. and Kurt, N. (2011). Survival and extinction of caring double-branching annihilating random walk. *Electron. Commun. Prob.* **16,** 271–282.

[3] Bramson, M. and Griffeath, D. (1981). On the Williams-Bjerknes tumour growth model I. *Ann. Prob.* **9,** 173–185.

[4] Brockhurst, M. A., Buckling, A., Racey, D. and Gardner, A. (2008). Resource supply and the evolution of public-goods cooperation in bacteria. *BMC Biology* **6,** 6pp.

[5] Clutton-Brock, T. (2009). Cooperation between non-kin in animal societies. *Nature* **462,** 51–57.

[6] Crespi, B. J. (2001). The evolution of social behavior in microorganisms. *Trends Ecol. Evol.* **16,** 178–183.

[7] Czuppon, P. (2016). Phenotypic heterogeneity in bacterial populations – a mathematical study. Doctoral Thesis, University of Freiburg.

[8] Drescher, K. *et al.* (2014). Solutions to the public good dilemma in bacterial biofilms. *Current Biol.* **24,** 50–55.

[9] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley, New York.
[10] Evilsizor, S. and Lanchier, N. (2016). Evolutionary games on the lattice: death-birth updating process. *Electron. J. Prob.* **21,** 29pp.
[11] Griffin, A. S. and West, S. A. (2003). Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science* **302,** 634–636.
[12] Hutzenthaler, M., Jordan, F. and Metzler, D. (2015). Altruistic defense traits in structured populations. Preprint. Available at https://arxiv.org/abs/1505.02154v1.
[13] Liggett, T. M. (1985). *Interacting Particle Systems*. Springer, New York.
[14] Louidor, O., Tessler, R. and Vandenberg-Rodes, A. (2014). The Williams–Bjerknes model on regular trees. *Ann. Appl. Prob.* **24,** 1889–1917.
[15] Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science* **314,** 1560–1563.
[16] Penn, D. J. and Frommen, J. G. (2010). Kin recognition: an overview of conceptual issues, mechanisms and evolutionary theory. In *Animal Behaviour: Evolution and Mechanisms*, Springer, Heidelberg, pp. 55–85.
[17] Sturm, A. and Swart, J. M. (2015). A particle system with cooperative branching and coalescence. *Ann. Appl. Prob.* **25,** 1616–1649.
[18] Williams, T. and Bjerknes, R. (1972). Stochastic model for abnormal clone spread through epithelial basal layer. *Nature* **236,** 19–21.
[19] Wingreen, N. S. and Levin, S. A. (2006). Cooperation among microorganisms. *Plos Biol.* **4,** 3pp.