The edge of neutral evolution in social dilemmas

New Journal of Physics

The edge of neutral evolution in social dilemmas

Jonas Cremer[1]**, Tobias Reichenbach**[2] **and Erwin Frey**[1]

[1] Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Department of Physics, Ludwig-Maximilians-Universität München, Theresienstrasse 37, D-80333 München, Germany
[2] Howard Hughes Medical Institute and Laboratory of Sensory Neuroscience, The Rockefeller University, 1230 York Avenue, New York, NY 10065 USA
E-mail: jonas.cremer@physik.uni-muenchen.de

Abstract. The functioning of animal as well as human societies fundamentally relies on cooperation. Yet, defection is often favorable for the selfish individual, and social dilemmas arise. Selection by individuals' fitness, usually the basic driving force of evolution, quickly eliminates cooperators. However, evolution is also governed by fluctuations that can be of greater importance than fitness differences, and can render evolution effectively neutral. Here, we investigate the effects of selection versus fluctuations in social dilemmas.

By studying the mean extinction times of cooperators and defectors, a variable sensitive to fluctuations, we are able to identify and quantify an emerging 'edge of neutral evolution' that delineates regimes of neutral and Darwinian evolution. Our results reveal that cooperation is significantly maintained in the neutral regimes. In contrast, the classical predictions of evolutionary game theory, where defectors beat cooperators, are recovered in the Darwinian regimes. Our studies demonstrate that fluctuations can provide a surprisingly simple way to partly resolve social dilemmas. Our methods are generally applicable to estimate the role of random drift in evolutionary dynamics.

IOP Institute of Physics ⚛ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Contents

1. Introduction

Individuals of ecological communities permanently face the choice of either cooperating with each other, or of cheating [1]–[5]. While cooperation is beneficial for the whole population and essential for its functioning, it often requires an investment by each agent. Cheating is then tempting, yielding social dilemmas where defection is the rational choice that would yet undermine the community and could even lead to ultimate self-destruction. However, bacteria or animals do not act rationally; instead, the fate of their populations is governed by an evolutionary process, through reproduction and death. The conditions under which cooperation can thereby evolve are subject of much contemporary, interdisciplinary research [2], [6]–[11].

Evolutionary processes possess two main competing aspects. The first one is selection by individuals' different fitness, which underlies adaptation [12]–[14] and is, by neo-Darwinists, viewed as the primary driving force of evolutionary change. In social dilemmas, defectors exploit cooperators rewarding them a higher fitness; selection therefore leads to fast extinction of cooperation, such that the fate of the community mimics the rational one. A large body of work is currently devoted to the identification of mechanisms that can reinforce cooperative behavior [19], e.g. kin selection [15, 16], reciprocal altruism [17, 18], or punishment [5, 8, 20]. However, the evolution of cooperation in Darwinian settings still poses major challenges. The second important aspect of evolution is random fluctuations that occur from the unavoidable stochasticity of birth and death events and the finiteness of populations. Neutral theories emphasize their influence which can, ignoring selection, explain many empirical signatures of ecological systems such as species–abundance relations as well as species–area relationships [21]–[25]. The importance of neutral evolution for the maintenance of cooperation has so far found surprisingly little attention [26]–[29].

In this paper, we introduce a general concept capable of investigating the effects of selection versus fluctuations by analyzing extinction events. We focus on social dilemmas,

IOP Institute of Physics **Φ** DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

i.e. we study the effects of Darwinian versus neutral evolution on cooperation[3]. For this purpose, we consider a population that initially displays coexistence of cooperators and defectors, i.e. cooperating and non-cooperating individuals. After some transient time, one of both 'species' will disappear [23], simply due to directed and stochastic effects in evolution and because extinction is irreversible: an extinct species cannot reappear again. The fundamental questions regarding cooperation are therefore: will cooperators eventually take over the whole population, and if not, for how long can a certain level of cooperation be maintained?

We show that the answers to these questions depend on the influence of stochasticity. For large fluctuations, evolution is effectively neutral, and cooperation maintained on a long time-scale, if not ultimately prevailing. In contrast, small stochastic effects render selection important, and cooperators die out quickly if disfavored. We demonstrate the emergence of an 'edge of neutral evolution' delineating both regimes.

2. Models and theory

2.1. Social dilemmas

Consider a population of N individuals that are either cooperators C or defectors D. We assume that individuals randomly engage in pairwise interactions, whereby cooperators and defectors behave distinctly differently and thereby gain different fitnesses. The population then evolves by fitness-dependent reproduction and random death, i.e. a generalized Moran process [2, 30], which we describe in detail in the next subsection. Here, we present the different possible fitness gains of cooperators and defectors.

In the *prisoner's dilemma* a cooperator provides a benefit b to another individual, at a cost c to itself (with the cost falling short of the benefit). In contrast, a defector refuses to provide any benefit and hence does not pay any costs. For the selfish individual, irrespective of whether the partner cooperates or defects, defection is favorable, as it avoids the cost of cooperation, exploits cooperators and ensures not to become exploited. However, if all individuals act rationally and defect, everybody is, with a gain of 0, worse off compared to universal cooperation, where a net gain of $b - c$ would be achieved. The prisoner's dilemma therefore describes, in its most basic form, the fundamental problem of establishing cooperation.

We can generalize the above scheme to include other basic types of social dilemmas [31]. Namely, two cooperators that meet are both rewarded a payoff $\mathcal{R}$, while two defectors obtain a punishment $\mathcal{P}$. When a defector encounters a cooperator, the first exploits the second, gaining the temptation $\mathcal{T}$, while the cooperator only gets the sucker's payoff $\mathcal{S}$. Social dilemmas occur when $\mathcal{R} > \mathcal{P}$, such that cooperation is favorable, in principle, while temptation to defect is large: $\mathcal{T} > \mathcal{S}$ and $\mathcal{T} > \mathcal{P}$. These interactions may be summarized by the payoff matrix

$$
\begin{array}{c|cc}
 & C & D \\
\hline
C & \mathcal{R} & \mathcal{S} \\
D & \mathcal{T} & \mathcal{P}
\end{array}
$$

Hereby, the entries in the upper row describe the payoff that a cooperator obtains when encountering a cooperator C or a defector D, and the entries in the lower row contain the payoffs for a defector.

---

[3] Within this paper, we use the term 'Darwinian' to signify evolutionary dynamics mainly driven by selection, as assumed within the modern synthesis of evolution.

Table 1. Different types of social dilemmas. We consider a population of cooperators and defectors, and describe their interactions in terms of four parameters $\mathcal{T}$, $\mathcal{R}$, $\mathcal{S}$ and $\mathcal{P}$, see text. Depending on the payoff-differences $\mathcal{S} - \mathcal{P}$ and $\mathcal{T} - \mathcal{R}$, four qualitatively different scenarios arise.

	$\mathcal{S} - \mathcal{P} < 0$	$\mathcal{S} - \mathcal{P} > 0$
$\mathcal{T} - \mathcal{R} < 0$	Coordination game	By-product mutualism
$\mathcal{T} - \mathcal{R} > 0$	Prisoner's dilemma	Snowdrift game

Variation of the parameters $\mathcal{T}$, $\mathcal{P}$, $\mathcal{R}$ and $\mathcal{S}$ yields four principally different types of games, see table 1 and figure 3. The *prisoner's dilemma* as introduced above arises if the temptation $\mathcal{T}$ to defect is larger than the reward $\mathcal{R}$, and if the punishment $\mathcal{P}$ is larger than the sucker's payoff $\mathcal{S}$, e.g. $\mathcal{R} = b - c$, $\mathcal{T} = b$, $\mathcal{S} = -c$ and $\mathcal{P} = 0$. As we have already seen above, in this case, defection is the best strategy for the selfish player. Within the three other types of games, defectors are not always better off. For the *snowdrift game* the temptation $\mathcal{T}$ is still higher than the reward $\mathcal{R}$ but the sucker's payoff $\mathcal{S}$ is larger than the punishment $\mathcal{P}$. Therefore, cooperation is favorable when meeting a defector, but defection pays off when encountering a cooperator, and a rational strategy consists of a mixture of cooperation and defection. Another scenario is the *coordination game*, where mutual agreement is preferred: either all individuals cooperate or defect as the reward $\mathcal{R}$ is higher than the temptation $\mathcal{T}$ and the punishment $\mathcal{P}$ is higher than the sucker's payoff $\mathcal{S}$. Lastly, the scenario of *by-product mutualism* yields cooperators fully dominating defectors since the reward $\mathcal{R}$ is higher than the temptation $\mathcal{T}$ and the sucker's payoff S is higher than the punishment $\mathcal{P}$. All four situations and the corresponding ranking of the payoff values are depicted in table 1 and figure 3.

2.2. The evolutionary dynamics

We describe the evolution by a generalized Moran process [2, 30], where the population size N remains constant and reproduction is fitness-dependent, followed by a random death event.

Let us denote the number of cooperators by N_{C}; the number of defectors then reads $N - N_{\mathrm{C}}$. The individuals' fitness are given by a constant background fitness, set to 1, plus the payoffs obtained from social interactions. The fitness of cooperators and defectors thus read $f_{\mathrm{C}} = 1 + \mathcal{R}(N_{\mathrm{C}} - 1)/(N - 1) + \mathcal{S}(N - N_{\mathrm{C}})/(N - 1)$ and $f_{\mathrm{D}} = 1 + \mathcal{T}N_{\mathrm{C}}/(N - 1) + \mathcal{P}(N - 1 - N_{\mathrm{C}})/(N - 1)$, respectively. In the following, we assume weak selection, i.e. the payoff coefficients are small compared to the background fitness. Note that within this limit, the self interactions of individuals are only of minor relevance. More important, in the case of weak selection, the evolutionary dynamics of the game depends only on the payoff differences $\mathcal{T} - \mathcal{R}$ and $\mathcal{S} - \mathcal{P}$. The different types of social dilemmas arising from these two parameters are listed in table 1.

In the Moran process, reproduction of individuals occurs proportional to their fitness, and each reproduction event is accompanied by death of a randomly chosen individual. As an example, the rate for reproduction of a defector and corresponding death of a cooperator reads

$$\Gamma_{\mathrm{C} \to \mathrm{D}} = \frac{f_{\mathrm{D}}}{\langle f \rangle} \frac{N_{\mathrm{C}}}{N} \frac{N - N_{\mathrm{C}}}{N}, \tag{1}$$

IOP Institute of Physics ⏀ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT
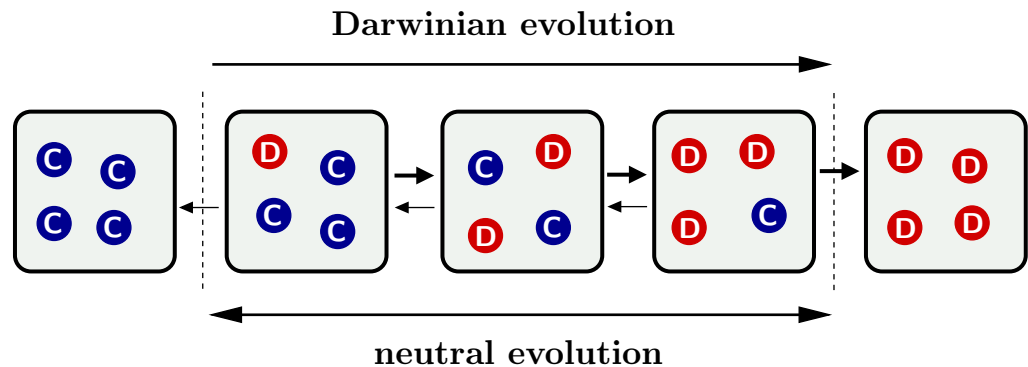
Darwinian evolution

Figure 1. Phase space exemplified for the prisoner's dilemma. The evolutionary dynamics consists of a Darwinian, directed part caused by selection of defectors (D) against cooperators (C), and a neutral, undirected part due to fluctuations. Eventually, only one species survives.

whereby $\langle f \rangle = f_C N_C / N + f_D (1 - N_C / N)$ denotes the average fitness. The timescale is such that an average number of N reproduction and death events occur in one time step.

2.3. Distinguishing Darwinian from neutral evolution: extinction times

The evolutionary dynamics is intrinsically stochastic. Although defectors may have a fitness advantage compared to cooperators, the latter also have a certain probability to increase. This situation is illustrated in figure 1 for a population of four individuals and the dynamics of the prisoner's dilemma. Darwinian evolution, through selection by individuals' fitness, points to the 'rational' state of only defectors, while fluctuations oppose this dynamics and can lead to a state of only cooperators. In any case, upon reaching overall defection or cooperation, the temporal development comes to an end. One species therefore eventually dies out.

The mean extinction time, i.e. the mean time it takes a population where different species coexist to eventually become uniform, allows us to distinguish Darwinian from neutral evolution. Consider the dependence of the mean extinction time T on the system size N. Selection, as a result of some interactions within a finite population, can either stabilize or destabilize a species' coexistence with others as compared to neutral interactions, thereby altering the mean time until extinction occurs. Instability leads to steady decay of a species, and therefore to fast extinction [28, 29, 32]: the mean extinction time T increases only logarithmically in the population size N, $T \sim \ln N$, and a larger system size does not ensure much longer coexistence. This behavior can be understood by noting that a species disfavored by selection decreases by a constant rate. Consequently, its population size decays exponentially in time, leading to a logarithmic dependence of the extinction time on the initial population size. In contrast, stable existence of a species induces $T \sim \exp N$, such that extinction takes an astronomically long time for large populations [28, 29, 33]. In this regime, extinction only stems from large fluctuations that are able to cause sufficient deviation from the (deterministically) stable coexistence. These large deviations are exponentially suppressed and hence the time until a rare extinction event occurs scales exponentially in the system size N.

An intermediate situation, i.e. when T has a power-law dependence on N, $T \sim N^\gamma$, signals dominant influences of stochastic effects and corresponds to neutral evolution [34]–[36]. Here

the extinction time grows considerably, though not exponentially, in increasing population size. Large N therefore clearly prolongs coexistence of species but can still allow for extinction within biologically reasonable timescales. A typical neutral regime is characterized by $\gamma = 1$, such that T scales linearly in the system size N. This corresponds to the case where the dynamics yields an essentially unbiased random walk in state space. The mean-square displacement grows linearly in time, with a diffusion constant proportional to N. The absorbing boundary is thus reached after a time proportional to the system size N. Other values of γ can occur as well. For example, and as shown later, $\gamma = 1/2$ can occur in social dilemmas (regimes (2) in figure 3).

To summarize, the mean extinction time T can be used to classify evolutionary dynamics into a few fundamental regimes. Darwinian evolution can yield stable and unstable coexistence, characterized by $T \sim \log N$ and $T \sim \exp N$, respectively. Power law dependences, $T \sim N^\gamma$, indicate neutral evolution. Transitions between these regimes can occur and manifest as crossovers in the functional relation $T(N)$.

2.4. Analytical description

An approximate analytical description, valid for a large number N of interacting individuals, is possible. The quantity of interest is thereby the probability $P(N_C, t)$ of having N_C cooperators at time t. Its time evolution is described by a master equation specified by transition rates such as (1). For large population sizes N, the master equation can be approximately described within a generalized diffusion approach, where the fraction $x = N_C/N$ of cooperators is considered as a continuous variable. The temporal development of $P(x, t)$ is then described by a Fokker–Planck equation [27, 37, 38],

$$\frac{\partial}{\partial t} P(x, t) = -\frac{\partial}{\partial x} [\alpha(x) P(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [\beta(x) P(x, t)]. \tag{2}$$

Hereby, $\alpha(x)$ describes the Darwinian of the evolution, due to selection by fitness differences, and corresponds to the deterministic dynamics $\frac{d}{dt} x = \alpha(x)$. The second part, which involves the diffusion term $\beta(x)$, accounts for fluctuations (to leading order) and thereby describes undirected random drift. $\beta(x)$ decreases like $1/N$ with increasing population size. For the social dilemmas that we study in this article, α and β are given by

$$\begin{aligned}
\alpha(x) &= x(1-x)\,[(\mathcal{S} - \mathcal{P})(1-x) - (\mathcal{T} - \mathcal{R})x], \\[6pt]
\beta(x) &= \frac{1}{N} x(1-x)\,[2 + (\mathcal{S} - \mathcal{P})(1-x) + (\mathcal{T} - \mathcal{R})x] \\[6pt]
&\approx \frac{2}{N} x(1-x).
\end{aligned} \tag{3}$$

Here, the approximation of β given in the last line is valid since weak selection is assumed.

The prisoner's dilemma, specified by $\mathcal{T} - \mathcal{R} = \mathcal{P} - \mathcal{S} \equiv c > 0$ describes the situation where defectors have a frequency independent fitness advantage $f_D - f_C = c$ as compared to cooperators. This scenario is frequently studied in population genetics [37]; we briefly discuss it in the following. The directed part and diffusion coefficients are given by

$$\begin{aligned}
\alpha(x) &= -cx(1-x), \\[6pt]
\beta(x) &= \frac{1}{N} x(1-x)\,[2 - c(1-2x)] \approx \frac{2}{N} x(1-x).
\end{aligned} \tag{4}$$

With these one can calculate the fixation probability $P_{\text{fix,C}}$ to end up with only cooperators if starting with an equal fraction of cooperators and defectors. It has already been calculated in previous work [37, 38] and reads,

$$P_{\text{fix,C}} = \frac{e^{-Nc/2} - e^{-Nc}}{1 - e^{-Nc}}. \tag{5}$$

The probability for fixation of defectors follows as $P_{\text{fix,D}} = 1 - P_{\text{fix,C}}$. Within the Darwinian regime ($Nc \to \infty$) defectors fixate ($P_{\text{fix,D}} = 1$), whereas for the neutral regime ($Nc \to 0$) both strategies have the same chance of prevailing ($P_{\text{fix,C}} = P_{\text{fix,D}} = 1/2$).

The fixation probability gives no information about the typical time needed for extinction of one of the two species. However, this time is important to determine whether extinction happens within the timescale of observation. We turn to this question in the following.

2.5. Analytical calculation of mean extinction times

The above analytical description, in the form of the Fokker–Planck equation (2), can be employed for computing the mean extinction time $T(x)$. The latter refers to the mean time required for a population initially consisting of a fraction $x = N_C/N$ of cooperators to reach a uniform state (only either cooperators or defectors). It is given as a solution to the corresponding backward Kolmogorov equation [39, 40],

$$\left[\alpha(x) \frac{\partial}{\partial x} + \frac{1}{2} \beta(x) \frac{\partial^2}{\partial x^2} \right] T(x) = -1, \tag{6}$$

with appropriate boundary conditions. This equation can be solved by iterative integration [39]. In detail, the mean extinction time, $T = T(x = 1/2)$, if starting with an equal fraction of cooperators $x = 1/2$ is given by

$$T = 2 \left[\left(\int_0^{1/2} du / \Psi(u) \right) \int_{1/2}^1 dy / \Psi(y) \int_0^y dz \, \Psi(z) / \beta(z) \right.$$
$$\left. - \left(\int_{1/2}^1 du / \Psi(u) \right) \int_0^{1/2} dy / \Psi(y) \int_0^y dz \, \Psi(z) / \beta(z) \right] \left[\int_0^1 du / \Psi(u) \right]^{-1}, \tag{7}$$

where $\Psi(x)$ is given by $\Psi(x) = \exp(\int_0^x dy \, 2\alpha(y)/\beta(y))$. We have performed these integrals for the general Moran process and show the results in the following.

For the special case of the prisoner's dilemma, specified by $\mathcal{T} - \mathcal{R} = \mathcal{P} - \mathcal{S} \equiv c > 0$, (frequency independent fitness advantage), equation (7) can be solved exactly. The solution reads,

$$T = \frac{1}{N} \left[P_{\text{fix,C}} \left\{ -\ln(cN) - \gamma + \text{Ei}(cN/2) + e^{cN} \left[\text{Ei}(-cN) - \text{Ei}(-cN/2) \right] \right\} \right.$$
$$\left. + P_{\text{fix,D}} \left\{ \ln(cN) + \gamma - \text{Ei}(-cN/2) + e^{-cN} \left[\text{Ei}(cN/2) - \text{Ei}(cN) \right] \right\} \right], \tag{8}$$

where $\text{Ei}(x)$ denotes the exponential integral $\text{Ei}(x) = \int x^{-1} \exp(x) dx$ and $\gamma \approx 0.577$ is the Euler Mascheroni constant. $P_{\text{fix,C}}$ and $P_{\text{fix,D}}$ denote the fixation probabilities of cooperators and defectors, given by equation (5). The analytical solution of the mean extinction time as a function of N is shown and compared to stochastic simulations in figure 2. For a further discussion of $T(N)$ (equation (8)) and its impact on evolutionary dynamics we defer the
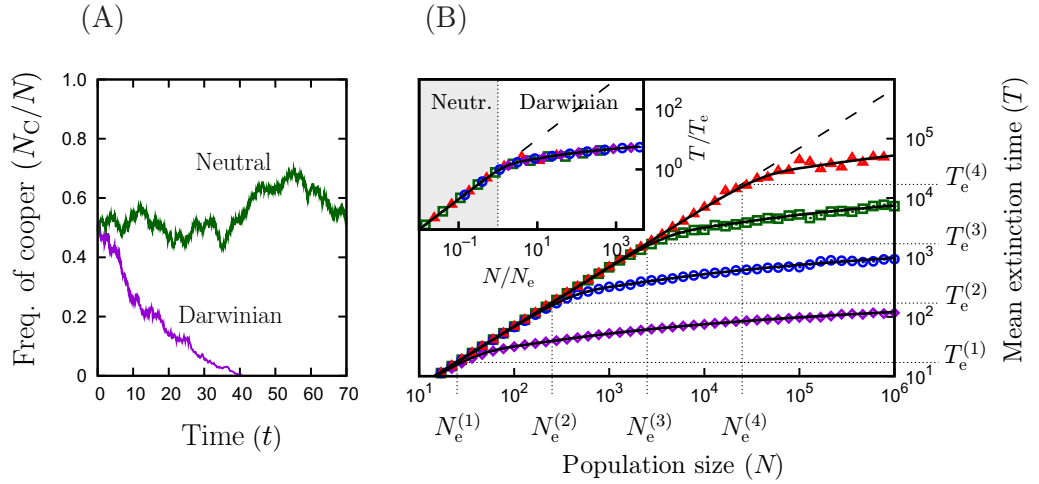
(A)   (B)

Figure 2. The prisoner's dilemma. Defectors save the cost c of cooperation and therefore have a fitness advantage of c compared to cooperators. (A) Exemplary evolutionary trajectories. A high selection strength, i.e. a high fitness difference $c = 0.1$ (purple), leads to Darwinian evolution and fast extinction of cooperators, while a small one, $c = 0.001$ (green), allows for dominant effects of fluctuations and maintenance of cooperation on long time-scales. We have used $N = 1000$ in both cases. (B) The dependence of the corresponding mean extinction time T on the system size N. We show data from stochastic simulations as well as analytical results (solid lines) for T, starting from equal abundances of both species, for different values of c (see text): $c_1 = 0.1$ ($\diamond$), $c_2 = 0.01$ ($\circ$), $c_3 = 0.001$ ($\square$) and $c_4 = 0.0001$ ($\triangle$). The transition from the neutral to the Darwinian regime occurs at population sizes $N_e^{(1)}$, $N_e^{(2)}$, $N_e^{(3)}$ and $N_e^{(4)}$. They scale as $1/c$: $N_e \approx 2.5/c$, as is confirmed by the rescaled plot where the data collapse onto the universal scaling function G, shown in the inset.

reader to section 3. Here, just note that the asymptotic behavior, of $\mathsf{Ei}(x)$ is given by $\mathsf{Ei}(x) \approx \log(|x|) + \gamma + x$ for $x \to 0$, and $\mathsf{Ei}(x) \approx \pm \log(|x|) + \exp(x)/x$ for $x \to \pm\infty$. With this, the well-known asymptotic solutions for high and low population size N, $T = \log(2)N$ and $T \sim \log N$ are obtained.

For general social dilemmas with arbitrary payoff values $\mathcal{T}, \mathcal{P}, \mathcal{R}, \mathcal{S}$, we need to rely on some approximations. Using the drift and diffusion coefficient given by equation (3) we now linearize the fraction α/β, i.e. we write $\alpha(x)/\beta(x) \approx g(x - x^*)$. Hereby $x^* = (\mathcal{S} - \mathcal{P})/(\mathcal{S} - \mathcal{P} + \mathcal{T} - \mathcal{R})$ denotes the fixed point of the deterministic dynamics, where $\alpha(x^*) = 0$ and $g = -N(\mathcal{S} - \mathcal{P} + \mathcal{T} - \mathcal{R})/2$. As an example, in the situation $\mathcal{S} - \mathcal{P} + \mathcal{T} - \mathcal{R} > 0$, $|x^*| \gg 1/\sqrt{g}$ and $|1 - x^*| \gg 1/\sqrt{g}$, we obtain the mean extinction time

$$
\begin{aligned}
T = \frac{N \ln(2)}{g} & [\mathsf{erfi}[\sqrt{g}\,(1 - x^*)] + \mathsf{erfi}(\sqrt{g}x^*)]^{-1} \\
& \times \Big\{ \mathsf{erfi}(-\sqrt{g}x^*)[-\mathcal{F}(g\,(1 - x^*)) + \mathcal{F}(g\,(1/2 - x^*))] \\
& - \mathsf{erfi}\left(\sqrt{g}\,(1 - x^*)\right)\left[\mathcal{F}(g\,(1/2 - x^*)) - \mathcal{F}(g\,(x^*))\right] \\
& + \mathsf{erfi}\left(\sqrt{g}\,(1/2 - x^*)\right)\left[\mathcal{F}(g\,(1 - x^*)) - \mathcal{F}(g\,(x^*))\right] \Big\}.
\end{aligned}
\tag{9}
$$

Here, $\mathsf{erfi}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}y \exp(y^2)$ denotes the complex error function and $\mathcal{F}(x) \equiv x \mathsf{F}_{1,1;3/2,2}(x)$ involves a generalized hypergeometric function. For graphical representation of equation (8) see figure 4(A, upper branch). As before, the correct asymptotic behavior can also be calculated for this case. Note that the asymptotic behavior of $\mathcal{F}(x)$ is given by $\mathcal{F}(x) \approx x$ for $x \to 0$ and $\mathcal{F}(x) \approx \mathsf{erfi}(\sqrt{x}) - \log(|x|)/2 - 1$ for $x \to \infty$. For small population size, the mean extinction time scales again like $T = \log(2)N$. For asymptotically large system sizes, the scaling depends on the value of the fixed point x^*. For an internal fixed point $x^* \in (0, 1)$, as arises in the snowdrift game, T scales as expected like $T \sim \exp(N)$.

In section 3, we analyze the properties of the analytical form of the mean extinction time, equations (8) and (9), together with numerical simulations, and demonstrate how it defines an emerging edge of neutral evolution.

2.6. Edges of neutral evolution

In section 3, we show that the mean extinction time, equation (9), exhibits different regimes of neutral and Darwinian dynamics. Here, we provide further information on how the boundaries between these regimes can be obtained analytically. For this purpose, we further approximate the dynamics. Let us, firstly, focus on the edge of the regime, where $T \sim N$ emerges. We note that, before unavoidable extinction of one species occurs, a quasi-stationary distribution may form around the fixed point x^*. Following the generic behavior of an Ornstein–Uhlenbeck process, its shape is approximately Gaussian [41]. Its width is given by $w \sim \sqrt{1/|g|}$. x^* and g are specified in the preceding section. Now, for small width, $w \ll 1$, the Darwinian evolution dominates the behavior, meaning $T \sim \ln(N)$ or $T \sim \exp(N)$. In contrast, if $w \gg 1$ the dynamics is essentially a random walk, and $T \sim N$ emerges. The edge of neutral evolution therefore arises at $w \sim 1$. Remembering that g is given by $g = -N(\mathcal{S} - \mathcal{P} + \mathcal{T} - \mathcal{R})/2$, it follows that the edge between both regimes for $\mathcal{S} - \mathcal{P}$ and $\mathcal{T} - \mathcal{R}$ is described by $(\mathcal{T} - \mathcal{R}) = d/N - (\mathcal{S} - \mathcal{P})$. Numerical simulations yield a good agreement with this prediction. As discussed later (see figure 2), they reveal that the crossover between the two regimes is remarkably sharp. The constant d which specifies the exact position of the crossover can therefore be estimated as $d \approx 5$. It follows that the regime of $T \sim N$ therefore corresponds to the square circumscribed by straight lines connecting the points $(\mathcal{T} - \mathcal{R}, \mathcal{S} - \mathcal{P}) = (5/N, 0), (0, -5/N), (-5/N, 0), (0, 5/N)$ as shown in figure 3.

A similar argument allows to determine the crossover from the other neutral regime, with $T \sim \sqrt{N}$, to the Darwinian regimes. The neutral regime emerges if the fixed point x^* is close to the boundaries, such that $w \sim |x^*|$ or $w \sim |1 - x^*|$ denotes the crossover to the Darwinian regimes. From these relations, it follows that the shapes of this second neutral regime are described by $\mathcal{T} - R \approx -(\mathcal{S} - \mathcal{P}) + (\mathcal{S} - \mathcal{P})^2 N$ and $\mathcal{S} - \mathcal{P} \approx -(\mathcal{T} - \mathcal{R}) + (\mathcal{T} - \mathcal{R})^2 N$. The proportionality constant has again been estimated from numerical simulations. From the latter, we have also found that the parabolic curves constitute a valid approximation to this second edge of neutral evolution.

3. Results

We employ the analytical expression, equations (8) and (9), for the mean extinction time, as well as computer simulations, to show how regimes of Darwinian and neutral evolution can be
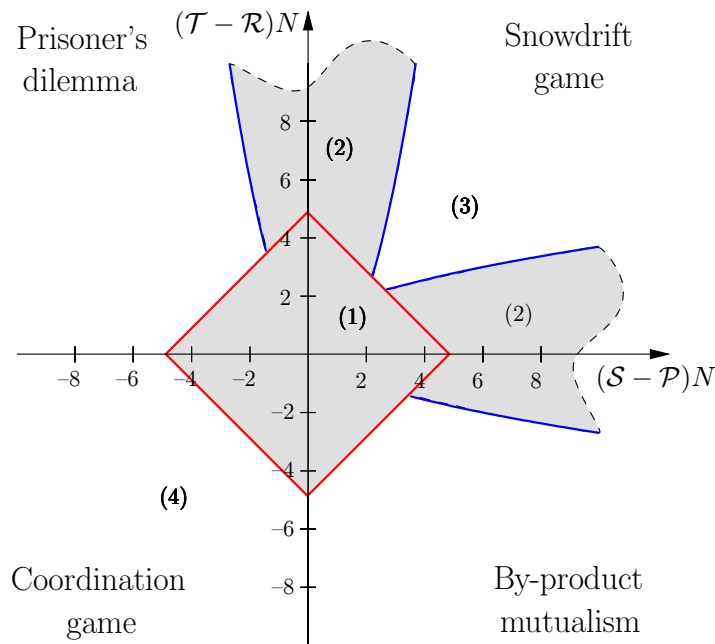
Figure 3. Social dilemmas. Depending on the sign of the payoff differences $\mathcal{T} - \mathcal{R}$ and $\mathcal{S} - \mathcal{P}$, a prisoner's dilemma, snowdrift game, by-product mutualism or coordination game arises. Two regimes of neutral evolution, (1) and (2) shown in grey, intervene two Darwinian regimes, (3) and (4), depicted in white. Coexistence of cooperators and defectors is lost after a mean time T which discriminates the distinct regimes: In (1), we encounter $T \sim N$, while $T \sim \sqrt{N}$ emerges in (2), $T \sim \exp N$ in (3), and $T \sim \ln N$ in (4). In the prisoner's dilemma and the coordination game, neutral evolution can thus maintain cooperation at a much longer time than Darwinian evolution. The edges of neutral evolution, red and blue curves, scale as $1/N$ (see text). We therefore show them depending on $(\mathcal{T} - \mathcal{R})N$ and $(\mathcal{S} - \mathcal{P})N$, where they adopt universal shapes.

distinguished. We demonstrate that neutral evolution can maintain cooperation on much longer timescales than Darwinian, even if cooperation has a fitness disadvantage.

3.1. Prisoner's dilemma

We start with the special case of the prisoner's dilemma where defectors have a frequency independent fitness advantage c compared to cooperators. The fixation probabilities, equation (5), provides first insight into the dynamics. When the population size N is large and selection by fitness differences dominates the dynamics, i.e. when $cN \gg 1$, the probability that defectors ultimately take over the whole population tends to 1. Cooperators are guaranteed to eventually die out. This is the regime of *Darwinian* evolution; the resulting outcome equals the one of rational agents. However, in the situation of small populations and small fitness difference, i.e. $cN \ll 1$, both cooperators and defectors have an equal chance of $1/2$ of fixating. In this regime, fluctuations have an important influence and dominate the evolutionary dynamics, leaving fitness advantages without effect, evolution is *neutral*.

Further quantification of the regimes of Darwinian and neutral evolution is feasible by considering the mean extinction time, given by equation (8). It is compared to stochastic simulations in figure 2(B) for different costs (fitness advantages) c. The excellent agreement confirms the validity of our analytic approach. Regarding the dependence of T on the population size N and the fitness difference c, the mean extinction time can be cast into the form,

$$T(N, c) = T_\mathrm{e}\mathsf{G}(N/N_\mathrm{e}), \tag{10}$$

with a scaling function G. T_e and N_e are characteristic timescales and population sizes depending only on the selection strength c. Analyzing its properties, it turns out that G increases linearly in N for small argument $N/N_\mathrm{e} \ll 1$, such that $T \sim N$, cf figure 2(B). This is in line with our classification scheme and the expected behavior. It indicates [28, 29] that for small system sizes, $N \ll N_\mathrm{e}$, evolution is *neutral*. Fluctuations dominate the evolutionary dynamics, while the fitness advantage of defectors does not give them an edge, cf figure 2(A). Indeed, in this regime, cooperators and defectors have an equal chance of surviving, see equation (5). The $T \sim N$ behavior shows that the extinction time considerably grows with increasing population size; a larger system size proportionally extends the time cooperators and defectors coexist. As expected, a very different behavior emerges for large system sizes, $N/N_\mathrm{e} \gg 1$, where G increases only logarithmically in N, and therefore $T \sim \ln N$, again in correspondence with our classification scheme of the mean extinction time. The extinction time remains small even for large system sizes, and coexistence of cooperators and defectors is unstable. Indeed, in this regime, selection dominates over fluctuations in the stochastic time evolution and quickly drives the system to a state where only defectors remain, cf figure 2(A). The evolution is *Darwinian*.

As described above, the regimes of neutral and Darwinian evolution emerge for $N/N_\mathrm{e} \ll 1$ and $N/N_\mathrm{e} \gg 1$, respectively. The cross-over population size N_e delineates both scenarios. Further analyzing the universal scaling function G, as well as comparison with data from stochastic simulations, see figure 2(B), reveals that the transition at N_e is notably sharp. We therefore refer to it as the *edge of neutral evolution*.

The crossover time T_e and the crossover population size N_e, which define the edge of neutral evolution decrease as $1/c$ in increasing cost c. This can be understood by recalling that the cost c corresponds to the fitness advantage of defectors and can thus be viewed as the selection strength. The latter drives the Darwinian dynamics, which therefore intensifies when c grows, and the regime of neutral evolution diminishes. On the other hand, when the cost of cooperation vanishes, evolution becomes neutral also for large populations. Indeed, in this case, defectors do not have a fitness advantage compared to cooperators; both do equally well. Our approach now yields information about how large the cost may be until evolution changes from neutral to Darwinian. From numerical inspection of G, we find that neutral evolution is present for $cN < 2.5$, and Darwinian evolution takes over for $cN > 2.5$. This resembles a condition previously derived by Kimura, Ohta and others [38, 42, 43] for frequency independent fitness advantages. The edge of neutral evolution arises at $N_\mathrm{e} = 2.5/c$ and $T_\mathrm{e} = 2.5/c$.

As a consequence we note that, though selection pressure clearly disfavors cooperation, our results reveal that the ubiquitous presence of randomness (stochasticity) in any population dynamics opens a window of opportunity where cooperation is facilitated. In the regime of neutral evolution, for $cN < 2.5$, cooperators have a significant chance of taking over the whole population when initially present. Even if not, they remain on timescales proportional to the system size, $T \sim N$, and therefore considerably longer than in the regime of Darwinian evolution, where they become extinct after a short transient time, $T \sim \ln N$.
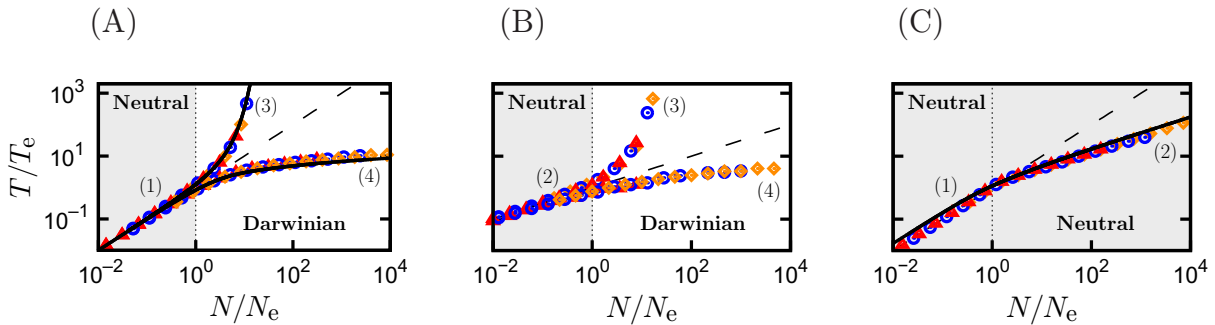
Figure 4. Transitions and universal scaling. We show the rescaled mean extinction time, T/T_e, depending on N/N_e, for different transitions emerging in social dilemmas (cf figure 3). (A) Transition from the neutral regime (1), where $T \sim N$ emerges, to the Darwinian regimes (3) ($T \sim \exp N$) as well as (4) ($T \sim \ln N$). (B) From neutral dynamics in regime (2) ($T \sim \sqrt{N}$) to the Darwinian regimes (3) ($T \sim \exp N$) and (4) ($T \sim \ln N$). (C) Transition between the two neutral regimes (1) ($T \sim N$) and (2) ($T \sim \sqrt{N}$). Analytical calculations are shown as black lines, and symbols have been obtained from stochastic simulations for large ($\triangle$), medium ($\circ$), and small ($\diamond$) values of $\mathcal{S} - \mathcal{P}$ and/or $\mathcal{T} - R$. The data collapse onto universal curves revealing the accuracy of the scaling laws. In (A), we have used $\mathcal{S} - \mathcal{P} = \mathcal{T} - \mathcal{R} \in \{-0.1, -0.01, -0.001, 0.001, 0.01, 0.1\}$, while $\mathcal{S} - \mathcal{P} \in \{-0.1, -0.01, -0.001, 0.001, 0.01, 0.1\}$, $\mathcal{T} - \mathcal{R} = 1$ in (B), and $\mathcal{S} - \mathcal{P} = 0$, $\mathcal{T} - \mathcal{R} \in \{0.001, 0.01, 0.1\}$ in (C).

3.2. General social dilemmas

Let us now consider the influence of fluctuations within the more general form of social dilemmas, given by the parameters $\mathcal{T}, \mathcal{P}, \mathcal{R}$ and $\mathcal{S}$. We employ the analytical form of the mean extinction time, equation (9), as well as results from stochastic simulations. Examples for different paths in parameter space are shown in figure 4. Again, the approximative analytical results agree excellently with numerics.

Concerning the dependence of the mean extinction time on the population size, different behaviors emerge, reflecting the different regimes of evolutionary dynamics. Two regimes of Darwinian evolution form, depicted white in figure 3. The first one occurs within the snowdrift game, where the extinction time increases exponentially in the population size, $T \sim \exp N$, and coexistence of cooperators and defectors is stable. The second regime comprises parts of the prisoner's dilemma, the coordination game and by-product mutualism. There, either defectors or cooperators eventually survive, and the mean extinction time of the other strategy is small, and obeys a logarithmic law $T \sim \ln N$. We have encountered this regime already in the particular case of the prisoner's dilemma specified by $\mathcal{T} - R = P - S \equiv c > 0$. These two Darwinian regimes are separated by two regimes of neutral evolution, shown in grey in figure 3. Firstly, for small N and small differences in the payoffs (i.e. around the point where the four types of games coincide) a $T \sim N$ behavior emerges. Secondly, at the lines where the snowdrift game turns into the prisoner's dilemma, respectively, by-product mutualism, the mean extinction time increases as a square-root in the population size, $T \sim \sqrt{N}$.

Similar to the prisoner's dilemma, we now aim at identifying the edge of neutral evolution, i.e. the crossover from the Darwinian regimes to the regimes of neutral evolution. We have calculated the boundaries of both neutral regimes, $T \sim N$ and $T \sim \sqrt{N}$ analytically, see section 2.6. They are described by straight lines for the first one and by parabola-shaped lines for the second one, see figure 3.

Both edges of neutral evolution scale proportional to the system size N. Therefore, while increasing the system size changes the payoff parameters where the crossovers appear, the shape and relations of the different regimes are not altered. Concerning the dependence of the edges of neutral evolution on the characteristic strength of selection s, meaning the average contribution of the fitness-dependent payoff to the overall fitness, different scaling laws arise. For the crossover from the neutral regime $T \sim N$ to the other regimes, T_e and N_e scale as $1/s$. In contrast, a scaling law $N_e \sim 1/s^2$ for crossovers between the neutral regime with $T \sim \sqrt{N}$ and the Darwinian regimes emerges. This different scaling behavior arises, for example, for $\mathcal{T} - \mathcal{R} = 1$ and varying $s = \mathcal{S} - \mathcal{P}$ as shown in figure 4(B).

4. Discussion

Cooperation is often threatened by exploitation and therefore, although beneficial, vulnerable to extinction. In evolutionary dynamics, this mechanism comes in through selection by individuals' fitness, the driving force of Darwinian evolution. However, evolution also possesses stochastic aspects. Employing a standard formulation of social dilemmas, we have shown that fluctuations can support cooperation in two distinct ways. Firstly, they can lead cooperators to fully take over the population. Secondly, neutral evolution considerably increases the time at which cooperators and defectors coexist, i.e. at which a certain level of cooperation is maintained. To emphasize the importance of the second point, we note that in real ecological systems the rules of the dynamics themselves change due to external [44] or internal [45] influences, setting an upper limit to the timescales at which evolution with constant payoffs, as we study here, applies. In particular, these times can be shorter than the times that would be needed for extinction of either cooperators or defectors, such that it may be less important to look at which of both would ultimately remain, but what the timescales for extinction are.

Quantitatively, we have shown the emergence of different Darwinian and neutral regimes. In the Darwinian regime of the prisoner's dilemma, cooperators are guaranteed to become extinct; the same is true for the second neutral regime, where $T \sim \sqrt{N}$. However, in the other neutral regime, with $T \sim N$, a random process determines whether cooperators or defectors prevail. Cooperators may therefore take over due to essentially neutral evolution. Moreover, even if cooperators eventually disappear, they remain for a considerably longer time in the neutral regimes than in the Darwinian regime. Indeed, in the regimes of neutral evolution, coexistence of cooperators and defectors is maintained for a mean time T obeying $T \sim N$, respectively, $T \sim \sqrt{N}$. For medium and large population sizes, this time exceeds by far the time $T \sim \ln N$ at which cooperation disappears in the Darwinian regimes of the prisoner's dilemma or of the coordination game (if defectors happen to dominate in the latter case). Neutral evolution can therefore maintain cooperation on a much longer timescale than Darwinian evolution. This effect is relevant as the neutral regimes considerably extend into the prisoner's dilemma as well as the cooperation game region. There, a form of neutrally maintained cooperation evolves.

IOP Institute of Physics 🜨 DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Our results have been obtained by applying a general concept based on extinction times that allows us to classify evolutionary dynamics into regimes of Darwinian and neutral character, separated by an emerging edge of neutral evolution. Apart from the social dilemmas under consideration here, we believe that our quantitative analytical approach can be versatilely applied to disentangle the effects of selection and fluctuations in various ecological situations where different species coexist [46]–[50]. Encouraged by our findings, we expect such studies to reveal further unexpected effects of fluctuations on ecology and evolution.

References

[1] Axelrod R 1984 *The Evolution of Cooperation* (New York: Basic Books)
[2] Nowak M A 2006 *Evolutionary Dynamics: Exploring the equations of Life* (Cambridge: Belknap Press)
[3] Ratnieks F L W, Foster K R and Wenseleers T 2006 Conflict resolution in insect societies *Annu. Rev. Entomol.* **51** 581–608
[4] Diggle S P, Griffin A S, Campbell G S and West S A 2007 Cooperation and conflict in quorum-sensing bacterial populations *Nature* **450** 411–4
[5] Fehr E and Fischbacher U 2003 The nature of human altruism *Nature* **425** 785–91
[6] Axelrod R and Hamilton W D 1981 The evolution of cooperation *Science* **211** 1390–6
[7] Milinski M 1987 Tit for tat in sticklebacks and the evolution of cooperation *Nature* **325** 433–5
[8] Rockenbach B and Milinski M 2006 The efficient interaction of indirect reciprocity and costly punishment *Nature* **444** 718–23
[9] Traulsen A and Nowak M A 2006 Evolution of cooperation by multilevel selection *Proc. Natl Acad. Sci. USA* **103** 10952–5
[10] Szabó G, Vukov J. and Szolnoki A 2005 Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices *Phys. Rev.* E **72** 047107
[11] Chuang J S, Rivoire O and Leibler S 2009 Simpson's paradox in a synthetic microbial system *Science* **323** 272–5
[12] Fisher R A 1930 *The Genetical Theory of Natural Selection* (Oxford: Oxford University Press)
[13] Wright S 1931 Evolution in Mendelian populations *Genetics* **16** 97–159
[14] Maynard Smith J 1982 *Evolution and the Theory of Games* (Cambridge: Cambridge University Press)
[15] Hamilton W D 1964 The genetical evolution of social behaviour. I+II *J. Theor. Biol.* **7** 1–52
[16] Hamilton W D 1996 *Narrow Roads of Gene Land: Evolution of Social Behaviour* (Oxford: Oxford University Press)
[17] Trivers R L 1971 The evolution of reciprocal altruism *Q. Rev. Biol.* **46** 35
[18] Nowak M A and Sigmund K 1992 Tit for tat in heterogeneous populations *Nature* **335** 250–3
[19] Nowak M A 2006 Five rules for the evolution of cooperation *Science* **314** 1560
[20] Hauert C, Traulsen A, Brandt H, Nowak M A and Sigmund K 2007 Via freedom to coercion: the emergence of costly punishment *Science* **316** 1905–7
[21] Volkov I, Banavar J R, Hubbell S P and Maritan A 2007 Patterns of relative species abundance in rainforests and coral reefs *Nature* **450** 45–9
[22] Bell G 2001 Neutral macroecology *Science* **293** 2413–8

[23] Hubbell S P 2001 *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton, NJ: Princeton University Press)

[24] Muneepeerakul R, Bertuzzo E, Lynch H J, Fagan W F, Rinaldo A and Rodriguez-Iturbe I 2008 Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin *Nature* **453** 220–3

[25] Muneepeerakul R, Weitz J S, Levin S A, Rinaldo A and Rodriguez-Iturbe I 2007 A neutral metapopulation model of biodiversity in river networks *J. Theor. Biol.* **245** 351–63

[26] Nowak M A, Sasaki A, Taylor C and Fudenberg D 2004 Emergence of cooperation and evolutionary stability in finite populations *Nature* **428** 646–50

[27] Traulsen A, Claussen J C and Hauert C 2005 Coevolutionary dynamics: from finite to infinite populations *Phys. Rev. Lett.* **95** 238701

[28] Antal T and Scheuring I 2006 Fixation of strategies for an evolutionary game in finite populations *Bull. Math. Biol.* **68** 1923

[29] Taylor C, Iwasa Y and Nowak M A 2006 A symmetry of fixation times in evolutionary dynamics *J. Theor. Biol.* **243** 245–51

[30] Moran P A 1964 *The Statistical Processes of Evolutionary Theory* (Oxford: Clarendon)

[31] Dawes R M 1980 Social dilemmas *Annu. Rev. Psychol.* **31** 169–93

[32] Reichenbach T, Mobilia M and Frey E 2007 Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games *Nature* **448** 1046–9

[33] Cremer J, Reichenbach T and Frey E 2008 Anomalous finite-size effects in the battle of the sexes *Eur. Phys. J.* B **63** 373–80

[34] Reichenbach T, Mobilia M and Frey E 2006 Coexistence versus extinction in the stochastic cyclic Lotka-Volterra model *Phys. Rev.* E **74** 051907

[35] Berr M, Reichenbach T, Schottenloher M and Frey E 2009 Zero-one survival behavior of cyclically competing species *Phys. Rev. Lett.* **102** 048102

[36] Altrock P M and Traulsen A 2009 Fixation times in evolutionary games under weak selection *New J. Phys.* **11** 013012

[37] Ewens W J 2004 *Mathematical Population Genetics* 2nd edn (New York: Springer)

[38] Kimura M 1983 *The Neutral Theory of Molecular Evolution* (Cambridge: Cambridge University Press)

[39] Gardiner C W 2007 *Handbook of Stochastic Methods* (Berlin: Springer)

[40] Risken H 1984 *The Fokker–Planck Equation—Methods of Solution and Applications* (Heidelberg: Springer)

[41] Van Kampen N G 2001 *Stochastic Processes in Physics and Chetry* 2nd edn (Amsterdam: North-Holland)

[42] Kimura M 1969 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles *Genet. Res., Camb.* **11** 247–69

[43] Ohta T and Gillespie J H 1996 Development of neutral and nearly neutral theories *Theor. Pop. Biol.* **49** 128–42

[44] Raup D M 1994 The role of extinction in evolution *Proc. Natl Acad. Sci. USA* **91** 6758–63

[45] Worden L and Levin S A 2007 Evolutionary escape from the prisoner's dilemma *J. Theor. Biol.* **245** 411–22

[46] Szabó G and Fáth G 2007 Evolutionary games on graphs *Phys. Rep.* **446** 97–216

[47] Reichenbach T, Mobilia M and Frey E 2007 Noise and correlations in a spatial population model with cyclic competition *Phys. Rev. Lett.* **99** 238105

[48] Reichenbach T, Mobilia M and Frey E 2008 Self-organization of mobile populations in cyclic competition *J. Theor. Biol.* **254** 368–83

[49] Reichenbach T and Frey E 2008 Instability of spatial patterns and its ambiguous impact on species diversity *Phys. Rev. Lett.* **101** 058102

[50] May R M 1974 *Stability and Complexity in Model Ecosystems* 2nd edn (Princeton, NJ: Princeton University Press)