Advances in Complex Systems
Vol. 16, Nos. 4 & 5 (2013) 1350017 (22 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0219525913500173

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

HOW CAN SOCIAL HERDING ENHANCE COOPERATION?

FRANK SCHWEITZER*, PAVLIN MAVRODIEV
and CLAUDIO J. TESSONE

Chair of Systems Design, ETH Zurich,
Weinbergstrasse 58, 8092 Zurich, Switzerland
**fschweitzer@ethz.ch*

Received 5 November 2012
Revised 8 February 2013
Accepted 8 March 2013
Published 30 April 2013

We study a system in which N agents have to decide between two strategies θ_i ($i \in 1 \dots N$), for defection or cooperation, when interacting with other n agents (either spatial neighbors or randomly chosen ones). After each round, they update their strategy responding nonlinearly to two different information sources: (i) the payoff $a_i(\theta_i, f_i)$ received from the strategic interaction with their n counterparts, (ii) the fraction f_i of cooperators in this interaction. For the latter response, we assume social herding, i.e., agents adopt their strategy based on the frequencies of the different strategies in their neighborhood, *without* taking into account the consequences of this decision. We note that f_i already determines the payoff, so there is *no additional* information assumed. A parameter ζ defines to what level agents take the two different information sources into account. For the strategic interaction, we assume a Prisoner's Dilemma game, i.e., one in which defection is the evolutionary stable strategy. However, if the additional dimension of social herding is taken into account, we find instead a stable outcome where cooperators are the majority. By means of agent-based computer simulations and analytical investigations, we evaluate the critical conditions for this transition toward cooperation. We find that, in addition to a high degree of social herding, there has to be a *nonlinear* response to the fraction of cooperators. We argue that the transition to cooperation in our model is based on *less* information, i.e., on agents which are not informed about the payoff matrix, and therefore rely on just observing the strategy of others, to adopt it. By designing the right mechanisms to respond to this information, the transition to cooperation can be remarkably enhanced. Our results are obtained in an evolutionary PD game with fixed payoffs and a fixed four-player neighborhood, where agents follow a stochastic better response dynamics.

Keywords: Prisoner's dilemma; social influence; mechanism design; nonlinear voter model.

1. Introduction

Cooperation is an abundant phenomenon in biological and social systems, but in most game-theoretical approaches *defection* should be the rational strategy to

*Corresponding author.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

choose. In order to solve this paradox, a vast number of literature has proposed modifications to the classical approach. They can be categorized along different directions:

- *changes of the payoff structure*: lowering the costs of cooperation to make it more attractive in the first place is another form of "buying cooperation" [24].
- *extension of the time horizon*: considering either repeated interaction, a memory for the strategy of the counterparts, calculating payoffs over a longer time interval, anticipating the future response to the own action [2].
- *considering spatial interaction*: the threshold for the outbreak of cooperation is lowered if agents' interaction is constrained to their nearest or second-nearest neighbors (as opposed to randomly chosen agents), or if agents can migrate between different spatial domains [23].
- *co-evolution of strategies and interaction*: agents can adopt the strategy of their best-performing neighbor and are allowed to break links with defective players, which can lead to high levels of cooperation [41].
- *introduction of additional strategies*: instead of unconditional cooperation/defection, strategies can allow for retaliation against defectors, e.g., TIT-FOR-TAT, or voluntary participation [8, 11].

We note that, particularly for biological systems, other additional mechanisms have been considered [19], such as altruism, the role of kinship relations, selection mechanisms on the group level, etc.

In this paper, we add a new element to the discussion: social herding, i.e., a mechanism that does *not* take strategic considerations into account. Agents can observe the actions of others *without* knowing their consequence. In a game-theoretical setting this means they cannot adopt a certain strategy based on payoff considerations because the payoff structure is not known to them. Thus, agents are just left with knowing the frequency of strategies either globally or in their neighborhood, and they choose their own strategy only based on the information about the frequency of these strategies. In our model, we assume that any agent can consider both the payoff-related and the frequency-related information and weight their influence by a parameter ζ, which is assumed to be the level of social herding. Precisely, $\zeta \to 0$ results in purely payoff-driven decisions, $\zeta \to 1$ in pure social herding.

The case where social herding is dominant has been widely studied in binary opinion dynamics models [5, 13, 18, 25, 27, 33, 36, 39] where opinions are not necessarily related to payoff but rather to social norms. Thus, agents may adopt the opinion of a majority in order to minimize social conflicts, but they may not have a utility-based preference for either of these opinions. Instead their opinion results from a frequency-dependent decision. The so called *linear voter model*, where the probability to choose a particular opinion is directly proportional to its frequency is a very common example for this. For homogeneous networks and finite system sizes it is known to result in consensus, i.e., the existence of only one opinion, asymptotically, but the outcome which opinion will dominate is not determined. In the

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

mean-field limit, this model always results in consensus of either of the two opinions. Starting e.g., with the frequency f_1 of opinions $\theta = 1$ and $f_0 = 1 - f_1$ with $\theta = 0$, the probability that the final consensus state is $\theta_i = 1$ for all i is f_1 [12]. Hence, a simple majority rule of social herding, as expressed in the linear voter model, may not improve the situation for cooperation. If the network of interaction is heterogeneous, e.g., in a complex network where agents have different degrees, the probability to reach each of the ordered configurations is given by the initial fractions weighted by the degree [31, 40]. Furthermore, the network topology can slow down the ordering dynamics and may even lead to a disordered system where no consensus is reached in the thermodynamic limit [28].

Therefore, we turn to the class of *nonlinear voter models* [22] in Sec. 2. As we will also show analytically in Sec. 3, a nonlinear social herding by itself will not lead to a transition toward cooperation. Instead, it is needed the right level of social herding in combination with the right nonlinearity, to enhance cooperation.

What do we gain from such insights? First of all, a better understanding of the fact that more information does not necessarily lead to a better outcome (in this case, to cooperation). Common wisdom would suggest that it is always better to have more information, e.g., to choose among more alternatives, to determine their consequences in advance, and thus to *reduce the risk* associated with making the wrong decision. What seems to be an optimal strategy on the individual level, turns out to lead to the lock-in into unfavorable situations on the global level. For example, in experiments on the *wisdom of crowd* effect, it was shown that more information about the guesses of other agents, combined with social influence, leads to a failure in the predictions [16]. Also, in a network formation model of agents sharing knowledge it was shown that *best response*, i.e., the choice of partners based on knowing all alternatives, resulted in a worse global performance as compared to a situation where just the next best partner was accepted [15]. As we point out with this work, to leave the trap of defection also crucially depends on using less of the available information, or to have a considerable fraction of less informed agents.

Second, from our insights we can derive mechanisms to improve the outcome in systems of strategically interacting agents. Mechanism design can be seen as the engineering part of economics. It allows to propose rules, or algorithms, for interactions that avoid the system getting trapped in suboptimal states. Some of these algorithms, such as the nowadays famous “Gale-Shapley” algorithm [7], are basically related to combinatorial optimization problems. That is, they propose a solution *for* the agents *without* involving the agents in finding it, themselves. Systems design, the way we see it, aims instead at proposing new ways of *interaction* at the agent level, in order to arrive at more favorable solutions at the system’s level. Our paper gives a lucid example of this kind of systems design, by proposing a different way of combining information an individual agent already has. This still leaves room for the forces of self-organization to act, but restricts the possible negative outcome.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

2. Basic Model

2.1. *Combining social herding and strategic interaction*

We consider a system with N agents. Each agent $i \in 1 \dots N$ is characterized by two individual variables which may change over time: $\theta_i(t)$ shall describe the agent's strategic behavior when interacting with other agents, whereas $\zeta_i(t)$ shall describe how much the agent is prone to social influence. We adopt the definition of *social influence* as the psychological tendency of individuals to adhere to and behave according to the expectations of its local neighborhood [14]. In this sense, our approach belongs to a wider class of models which do not restrict herding behavior to perfectly rational agents [20].

In an economic context, θ_i refers to the strategy of a utility maximizing agent, chosen from a (discrete) set σ of possible strategies. We use the standard game theoretical setting of a Prisoner's Dilemma (PD) game, i.e., $\sigma \in \{0, 1\}$, where the strategic behavior $\sigma = 0$ refers to *defection* (D) and $\sigma = 1$ to *cooperation* (C).

We assume that each agent plays a 2-person (noniterated) game with n other agents which are located in its neighborhood. The completion of these n games is called a round. From each of these interactions the agent receives a payoff which depends both on the strategic behavior of the agent itself and on the opponents'. The game structure describing a single interaction between two agents can be summarized by the standard payoff matrix of a 2-person game:

	$\theta_j = 1$	$\theta_j = 0$
$\theta_i = 1$	R/R	S/T
$\theta_i = 0$	T/S	P/P

Suppose, agent i has chosen to cooperate, then its payoff is R if the other agent j has also chosen to cooperate (without knowing about the decision of agent i), but S if agent j defects. On the other hand, if agent i has chosen to defect, then it will receive the payoff T if agent j cooperates, while it will receive P if agent j defects.

In this paper, we will restrict the discussion to the PD game, but we note that our investigations can be extended to other games that result from different values of R, S, T and P [24]. For the particular case of the PD game, the payoffs have to fulfill the following two inequalities:

$$T > R > P > S; \quad 2R > S + T. \tag{1}$$

The known standard values are $T = 5$, $R = 3$, $P = 1$, $S = 0$. This implies that, in a so-called one-shot game (no repeated interaction), defection $\sigma = 0$, is the rational strategy because it rewards the higher payoff for an agent i no matter whether the opponent chooses C or D. As this argument applies to both agents, one can expect that on the system level a global defective behavior emerges. Because of this, the PD game has become a paradigmatic model to study different mechanisms of transition toward a global cooperative behavior [1, 32], a question that has puzzled the scientific community for decades.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

Let us define the degree of cooperation on the system's level by the total number of cooperating agents, $N_1(t)$ relative to the total population N. Since the number of agents is constant, the global frequencies f_σ of cooperating and defecting agents are given by

$$\begin{aligned} N &= \sum_\sigma N_\sigma = N_0 + N_1 = \text{const.}; \quad \sigma \in \{0,1\}, \\ f_\sigma &= \frac{N_\sigma}{N}; \quad f \equiv f_1 = 1 - f_0. \end{aligned} \tag{2}$$

In the following, the variable f shall refer to the global frequency of cooperators.

The interaction of each agent with n other agents in a 2-person game results in $\binom{N}{n}$ different possibilities to choose a partner. As the result of these interactions that may occur *independently*, but *simultaneously* [9, 24], agent i receives a total payoff $A_i(\theta_i)$ which depends both on its own strategy θ_i and the strategies of the n different partners. Let us assume that n_0 of these partners have chosen to defect, whereas $n_1 = n - n_0$ partners have chosen to cooperate. Then the total payoff from these n interactions reads:

$$A_i(\theta_i) = \delta_{1,\theta_i}[n_1 R + n_0 S] + \delta_{0,\theta_i}[n_1 T + n_0 P], \tag{3}$$

where $\delta_{x,y}$ means the Kronecker delta, which is 1 only for $x = y$ and zero otherwise. Dividing by n gives the scaled total payoff:

$$a_i(\theta_i, f_i) = \frac{A_i(\theta_i)}{n} = \delta_{1,\theta_i}[f_i R + (1 - f_i)S] + \delta_{0,\theta_i}[f_i T + (1 - f_i)P], \tag{4}$$

where $f_i = n_1/n = 1 - n_0/n$ gives the fraction of cooperating agents agent i interacts with. Assuming e.g., that agent i interacts with its neighbors, f_i gives the *local* frequency of cooperators. If on the other hand agent i interacts with n randomly chosen agents, the probability to choose a cooperator is directly proportional to the global fraction f. That is, in the so-call *mean-field approach* we set $f_i \equiv f$.

Strategic considerations implies that agent i pays attention to the scaled payoff $a_i(\theta_i, f_i)$ expected from the interaction with f_i cooperators, which of course also depends on its own strategy θ_i. A nonlinear function $\mathcal{G}(a_i)$ shall consider the way agent i combines the information about the different payoffs $a_i(\theta_i, f_i)$ and $a_i(1 - \theta_i, f_i)$ resulting from its possible strategic choice. This shall be used below to define the transition rate for an agent to change between strategies, therefore we conveniently normalize $\mathcal{G}(a_i)$ to one. In a very general way, we assume:

$$\mathcal{G}(a_i) = \frac{\exp[\beta_i a_i(\theta_i, f_i)]}{\exp[\beta_i a_i(\theta_i, f_i)] + \exp[\beta_i a_i(1 - \theta_i, f_i)]}. \tag{5}$$

Equation (5) has the form of a logit-function well established in decision theory [4, 17, 37]. The parameter β_i allows agents to individually weight differences between the payoffs. $\beta_i \to 0$ represents the limit of random choice between strategies, $\mathcal{G}(a_i) \to 1/2$, whereas $\beta_i \to \infty$ means that even small differences in payoff lead to an immediate switch between $\mathcal{G}(a_i) = 0$ and $\mathcal{G}(a_i) = 1$. For small values of

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

β_i, the $\mathcal{G}(a_i)$ tends to one if the expected payoff times the $a_i(\theta_i, f_i)$ from strategy θ_i is much larger than the expected payoff $a_i(1-\theta_i, f_i)$ from the opposite strategy $1-\theta_i$ and it tends to zero in the opposite case. If both payoffs become comparable, $G(a_i)$ is about $1/2$. Intermediate values of β_i allow for a smooth transition between the two strategic cases.

We note that for sufficiently small values of β_i, Eq. (5) can be approximated by the linear function

$$\mathcal{G}(a_i) \approx \frac{1}{2}\left[1 + \frac{\beta_i}{2}\{a_i(\theta_i, f_i) - a_i(1-\theta_i, f_i)\}\right], \tag{6}$$

i.e., agents pay attention to the *difference* between the two possible payoffs.

The situation becomes different if the agent is unable to calculate the expected payoff. In our model, we assume that the agent then rather pays attention to the action of the majority and tends to imitate this without knowing about the consequences. Thus, agent i only responds to the information associated with the frequency which shall be described by a logit-function similar to Eq. (5):

$$\mathcal{F}(f_{\theta_i}) = \frac{\exp[2\beta_i\kappa_i(f_{\theta_i})f_{\theta_i}]}{\exp[2\beta_i\kappa_i(f_{\theta_i})f_{\theta_i}] + \exp[-2\beta_i\kappa_i(f_{\theta_i})f_{\theta_i}]} - \frac{1}{2} \tag{7}$$

f_{θ_i} describes the local frequency of agents playing strategy θ_i in the neighborhood of agent i, and $f_{1-\theta_i} = 1 - f_{\theta_i}$ is the local frequency of agents playing the opposite strategy. Both frequencies being equal, $\mathcal{F}(f_{\theta_i}) = \mathcal{F}(f_{1-\theta_i}) = 1/2$. Again, for sufficiently small β_i, from a linear approximation in Eq. (7) we find,

$$\mathcal{F}(f_{\theta_i}) \approx \beta_i\kappa_i(f_{\theta_i})f_{\theta_i}. \tag{8}$$

$\kappa_i(f_{\theta_i})$ is a nonlinear response function to consider a weighted influence of the frequency [22] as we will investigate below. $\kappa_i(f_{\theta_i})$ may also depend on the time an agent has kept its current strategy, or opinion [29, 30]. We emphasize that for the so-called linear voter model, $\kappa_i(f_{\theta_i})$ is simply a constant κ that does not depend on the frequency. So $\beta_i\kappa$ can be scaled to one, which means that for the linear voter model we simply arrive at $\mathcal{F}(f_{\theta_i}) = f_{\theta_i}$. Thus, the response of agent i is directly proportional to the local frequency of agents playing strategy θ_i.

After having defined the agent's response to strategic information and to social herding, we use the individual parameter ζ_i to weight these two different influences. Specifically, we define the transition rate for agent i to switch from strategy $(1-\theta_i)$ to the opposite strategy θ_i as follows:

$$w(\theta_i|(1-\theta_i), f_i, \zeta_i) = (1-\zeta_i)\mathcal{G}(a_i) + \zeta_i\mathcal{F}(f_{\theta_i}). \tag{9}$$

For $\zeta_i \to 0$, we cover the limit case of strategic interaction in PD game, for $\zeta_i \to 1$, we arrive at the limit case of pure social herding, i.e., imitation behavior without calculating the resulting consequences.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

2.2. *Specifying the transition rates*

Before describing the system's dynamics by means of a master equation in the following section, it will be handy to write down the transition rates of Eq. (9) more specifically. The transition rates apply for a *frequency dependent* process, i.e., they do not depend on the specific sequence of interaction. In this paper, we fix the number of independent, but simultaneous 2-person games to $n = 4$, which is convenient to compare random interactions with local ones on a regular lattice. Hence, the relevant frequencies have only discrete values $f_i \equiv k_i/n$ where $k_i \equiv n_1 = 0, 1, 2, 3, 4$ is the actual number of cooperating agents, agent i is interacting with. On the other hand, random interactions can be approximated by the so-called mean field approximation, where $f_i = f$, the global fraction of cooperators.

Dropping the individual index i for the moment, we have to distinguish between two different transition rates, $c_k(\zeta) = w(1|0, k, \zeta)$, i.e., the transition from defection to cooperation dependent on k cooperating agents, and $d_k(\zeta) = w(0|1, k, \zeta)$, i.e., the transition from cooperation to defection under the same conditions. Both of these rates are comprised of two parts, one resulting from strategic behavior ($\tilde{c}_k$, $\tilde{d}_k$), the other one resulting from social herding ($\hat{c}_k$, $\hat{d}_k$),

$$c_k(\zeta) = (1-\zeta)\tilde{c}_k + \zeta\hat{c}_k; \quad d_k(\zeta) = (1-\zeta)\tilde{d}_k + \zeta\hat{d}_k. \tag{10}$$

For the terms ($\hat{c}_k$, $\hat{d}_k$) related to social herding, we use the linear approximation, Eq. (8), i.e., for the specified neighborhood $n = 4$,

$$\hat{c}_k = \frac{k}{4}\beta\kappa_k; \quad \hat{d}_k = 1 - \frac{n-k}{4}\beta\kappa_k. \tag{11}$$

Again, for the linear voter model with $\kappa_k \equiv \kappa$ and the ($\hat{c}_k$, $\hat{d}_k$) would simply result from the set of values {0, 1/4, 2/4, 3/4, 1}. In order to use nonlinearities in the frequency response, we rather prefer to specify the ($\hat{c}_k$, $\hat{d}_k$) by discrete values α_0, α_1, α_2 as shown in Table 1.

The parameter α_0 describes the transition of a cooperator (defector) toward defection (cooperation) if surrounded by cooperators (defectors) solely based on *social herding*. Because agents with such strategies to follow are absent in the neighborhood, α_0 should be consequently zero, even if there is a strong *strategic* incentive for a cooperator to switch toward defection if surrounded by cooperators.

Table 1. Nonlinear transition rates to cooperation dependent on the frequency of cooperating neighbours, f.

$f = k/n$	$\tilde{c}_k$	$\tilde{d}_k$	$\hat{c}_k$	$\hat{d}_k$
0	$\tilde{c}_0$	$\tilde{d}_0$	α_0	$1-\alpha_0$
1/4	$\tilde{c}_1$	$\tilde{d}_1$	α_1	$1-\alpha_1$
2/4	$\tilde{c}_2$	$\tilde{d}_2$	α_2	α_2
3/4	$\tilde{c}_3$	$\tilde{d}_3$	$1-\alpha_1$	α_1
1	$\tilde{c}_4$	$\tilde{d}_4$	$1-\alpha_0$	α_0

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

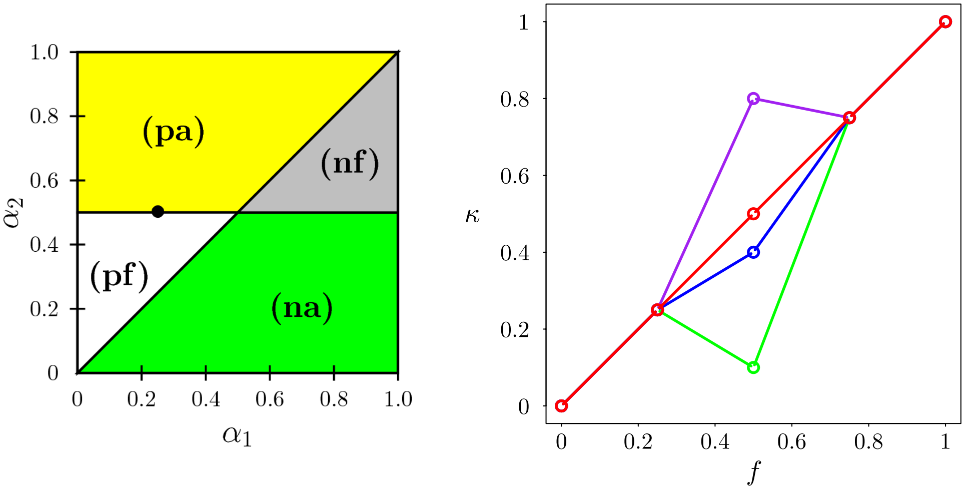

Fig. 1. (Color online) (left) Parameter space (α_1, α_2) to define the nonlinearity in social herding (see also Table 1). The different regions are explained in the text. We use the (pa) region, defined by Eq. (17). (right) Linear voter model (red line) and deviations controlled by α_2 at $f = 0.5$.

Hence, considering only social herding, pure cooperation and pure defection are "absorbing" states for the dynamics of the system. This can be avoided by choosing $\alpha_0 = \varepsilon$, a very small value that allows for occasional random changes of the strategies [22], but in this paper we choose $\alpha_0 = 0$.

Possible combinations of (α_1, α_2) define a parameter space to distinguish between different forms of social herding, as shown in Fig. 1(left). Positive frequency dependence (pf) means that the probability to change to the opposite strategy monotonously increases with the frequency of that strategy in the neighborhood, also known as "majority voting". Negative frequency dependence (nf) means the opposite, i.e., the probability monotonously decreases with the frequency, also known as "minority voting". On the other hand, (pa) and (na) define parameter regions with nonmonotonous dependence. For example, (pa) means an increase of the probability as long as the opposite strategy is not the majority, also known as voting against the trend, while (na) describes constellations with a strong amplification of minority strategies. We note that the so-called "voter point" that represents the linear voter model — where $\alpha_1 = 1/4$ and $\alpha_2 = 2\alpha_1 = 1/2$ are strictly proportional to k — is on the border between the (pf) and (pa) parameter regions. For our investigations, we will consider a scenario where the nonlinearity is only represented by α_2, whereas α_1 is chosen according to the linear voter model. Four possible cases which refer to the (pf), (pa), (na) and the linear voter model are shown in Fig. 1(right).

It remains to specify the payoff related terms $(\tilde{c}_k,\ \tilde{d}_k)$ which follow directly from Eq. (5). Here, we assume the deterministic limit $\beta_i \to 0$, for which we get $\mathcal{G}(a_i) = \Theta[a_i(\theta_i, f_i) - a_i(1-\theta_i, f_i)]$, where $\Theta[y]$ is the Heavyside function, which is one if $y > 0$ and zero otherwise. That is, $\mathcal{G}(a_i)$ is either one or zero depending on

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

whether the payoff for the changed strategy is larger or less than the payoff resulting from the current strategy. Taking into account the payoff relations, Eq. (1), we verify that the expected payoffs, Eq. (4), for defectors, $a(0, f)$, are *always* larger than the corresponding ones for cooperators, $a(1, f)$, regardless of the fraction of cooperators in the neighborhood. That is, in nonrepeated games as considered here, defection is an evolutionary stable strategy. Hence, in the deterministic limit of strategic interaction, we have always $\tilde{c}_k = 0$ and $\tilde{d}_k = 1$. This can be rightly assumed as the worst-case scenario because, considering only a strategic point of view, the system will always end up in pure defection. The most important thing is to identify conditions where an additional social herding allows not only to avoid this trap, but also to let the dynamics to converge to pure cooperation.

The observant reader may have noticed that we have interpreted β_i differently for social herding (where we assumed that it is just small) and for strategic interaction (where $\beta \to 0$ was assumed). This is not a contradiction. In fact, β quantifies the randomness in following the different information, and we can assume that the payoff related attention is much higher and less prone to errors than the response to the behavior of neighbors. In general, we may distinguish between $\tilde{\beta}_i$ and $\hat{\beta}_i$ for the different responses, but this is not applied here.

2.3. *Dynamics to change the strategy*

In the previous section, we have defined the "rules" for agents to change their strategy dependent on both strategic information and social herding. Most agent-based models, at this point, would continue with extensive computer simulations to probe the parameter space for some nontrivial results. We will certainly follow with computer simulations as well, however we are also interested in some analytical insights into the model which would allow us to predict the system's dynamics without testing every possible parameter combination. For this reason, we need to specify the dynamics of agents in a more formal way, on two different levels, (i) on the micro level of the individual agent, and (ii) on the macro level, describing the fraction of cooperators in the system.

For the micro level, we use a stochastic approach, i.e., we deal with the probability $p_i(\theta_i, t)$ that agent i uses strategy θ_i at time t. As explained before, this probability depends on the strategies of agents in the neighborhood of agent i expressed by the vector $\underline{\theta}_i = \{\theta_{i_1}, \theta_{i_2}, \ldots, \theta_{i_n}\}$. Hence, $p_i(\theta_i, t)$ is defined as the marginal distribution:

$$p_i(\theta_i, t) = \sum_{\underline{\theta}'_i} p(\theta_i, \underline{\theta}'_i, t). \tag{12}$$

The summation is over all possible distributions $\underline{\theta}'_i$. Specific realizations of these distributions shall be denoted as $\underline{\sigma}$. For $n = 4$, there are 2^n possible realizations. For the time-dependent change of $p_i(\theta_i, t)$ we assume the following master equation:

$$\frac{d}{dt} p_i(\theta_i, t) = \sum_{\underline{\theta}'_i} [w(\theta_i|(1 - \theta_i), \underline{\theta}'_i) p(1 - \theta_i, \underline{\theta}'_i, t) - w(1 - \theta_i|\theta_i, \underline{\theta}'_i) p(\theta_i, \underline{\theta}'_i, t)]. \tag{13}$$

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

This equation considers all possible processes that may lead to an increase or decrease in the probability that agent i uses strategy θ_i given the neighborhood distribution $\underline{\theta}_i$, with the transition rates $w(\theta_i|(1-\theta_i),\underline{\theta}'_i)$, $w(1-\theta_i|\theta_i,\underline{\theta}_i)$. Note that these are not the transition rates defined in Eq. (9), which only depend on the local frequency f_i, but not on the neighborhood distribution $\underline{\theta}_i$. In order to map the two, we have to consider how many specific realizations of the distribution $\underline{\theta}_i$ may lead to the same f_i. Taking the example $\underline{\sigma} = \{0010\}$, there are exactly $\binom{4}{1}$ different possibilities to realize $f_i = 1/n$. Hence, transforming the master equation (13) that depends on the neighborhood distribution $\underline{\theta}_i$ into one that only contains the respective local frequency f_i results in a combinatorial prefactor of $\binom{n}{k}$. Using again the specific notations c_k, d_k, Eq. (10) for the transition rates, we can rewrite the master equation (13) now as

$$\frac{d}{dt}p_i(1,\zeta,t) = \sum_{k=0}^{n}\binom{n}{k}[c_k(\zeta)p(0,k/n,\zeta,t) - d_k(\zeta)p(1,k/n,\zeta,t)]. \tag{14}$$

The corresponding master equation for $p_i(0,\zeta,t) = 1 - p_i(1,\zeta,t)$ follows likewise. Note that in Eq. (14) we have chosen the individual parameter ζ_i to be a constant ζ. That is, whereas the local frequency $f_i = k/n$ changes over time because of concurrent decisions of neighboring agents about their strategies. ζ is, in this paper, assumed to be a global control parameter the impact of which will be discussed together with the computer simulations.

With this, we have a bottom-up description of the system's dynamics given by N stochastic equations, Eq. (14), which are coupled because of the overlapping neighborhoods of agents, expressed in terms of f_i. On the other hand, on the macroscopic level we have to deal with the probability $P(f,\zeta,t)$ to find a given fraction of cooperators, f, at time t, assuming the social herding factor ζ. The dynamics can again be specified by a stochastic equation:

$$\frac{d}{dt}P(f,\zeta,t) = \sum_{f'}[W(f|f',\zeta)P(f',\zeta,t) - W(f'|f,\zeta)P(f,\zeta,t)]. \tag{15}$$

f' denotes all possible deviations from a given value f that can be reached during one time step by means of the transition rates $W(f'|f,\zeta)$. These are not identical with the individual transition rates, Eq. (9), but aggregated rates that take into account all possible ways to change f. The smallest change of $f \equiv N_1/N$, Eq. (2), is the addition or subtraction of a single cooperator, i.e., $f' \in \{(N_1+1)/N; (N_1-1)/N\}$. The individual equivalent for such processes is given by Eq. (10), where the terms $c_k(\zeta)$ describe the transition of a single defector into a cooperator, and the $d_k(\zeta)$ the opposite transition. Hence, we find for the aggregated transition rates

$$\begin{aligned} W(f+1/N|f,\zeta) &\equiv W_+(f,\zeta) = \sum_{k=0}^{n}\binom{n}{k}f^k(1-f)^{n-k}c_k(\zeta) \\ W(f-1/N|f,\zeta) &\equiv W_-(f,\zeta) = \sum_{k=0}^{n}\binom{n}{k}f^{n-k}(1-f)^k d_{n-k}(\zeta). \end{aligned} \tag{16}$$

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

The combinatorial prefactors preceding the $c_k(\zeta)$ and $d_k(\zeta)$ result from the various ways to choose agents with $n = 4$ neighbors, k of which could be cooperators given the global fraction of cooperators f. Here, we have used the so-called mean-field assumption that replaces the frequencies f_i of the individual neighborhoods by the global value f. With the specific values for $c_k(\zeta)$ and $d_k(\zeta)$ given by Eqs. (10) and (11), the dynamics on the systemic level is also completely specified. In the following, we will use the dynamics on the micro level for carrying out computer simulations, while the dynamics on the macro level will be used for analytical investigations.

3. Results of Computer Simulations

We now use the dynamics specified in Eq. (14) to run agent-based computer simulations for different sets of parameters. According to Eqs. (10), (11), we only need to vary the weight $0 \leq \zeta \leq 1$ and the parameters $0 \leq (\alpha_1, \alpha_2) \leq 1$ assigned to the social herding of the agents. Regarding their strategic decision, everything is already defined, and with $\tilde{c}_k = 0$, $\tilde{d}_k = 1$ defection remains the only choice. This "worst case scenario" can be only changed because of a considerable amount of social herding, in which agents copy the strategy of their neighbors regardless of the payoff assigned to it. This is shown in Fig. 2. Below a critical level for social herding, $\zeta \approx 0.7$, only defection remains. For $\zeta > 0.7$, we observe different levels of cooperation which depend on the combination of ζ and α_2. If $\zeta > 0.8$, cooperation even becomes the majority, i.e., $f > 0.5$, but only for large values of ζ and α_2 full cooperation, $f \to 1$, is reached. This issue is further investigated below.

The role of the nonlinearity in social herding, expressed in terms of α_1, α_2, is further investigated in Fig. 3, given a supercritical level of social herding. We see

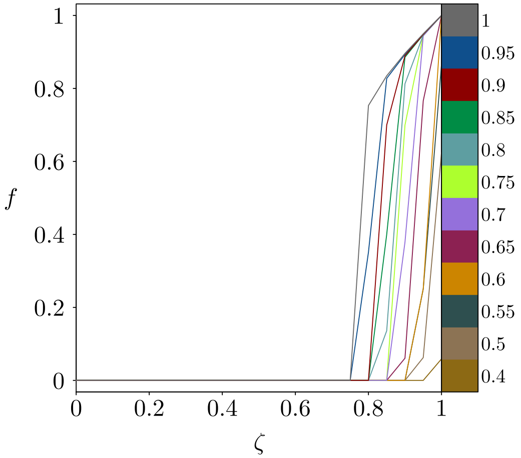

Fig. 2. (Color online) Global fraction of cooperation f dependent on the level of social herding ζ. $\alpha_1 = 0.25$ is fixed, α_2 varies between 0.4 and 1.0 according to the color scale. System size $N = 400$.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

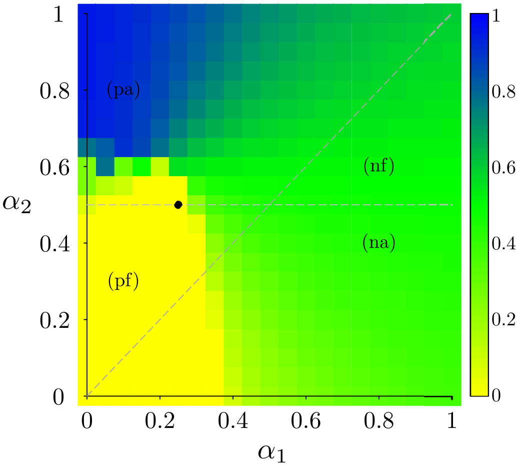

Fig. 3. (Color online) Fraction of cooperation (color scale) dependent on the nonlinearities in social herding, defined by α_1, α_2. Fixed level of social herding $\zeta = 0.95$. The four different areas are defined in Fig. 1(left). $\bullet$ indicates the linear voter model. System size $N = 400$.

that there is an *optimal nonlinearity* to enhance cooperation, i.e., α_1, α_2 have to be chosen such that they belong to the area of "positive allee" (pa) effects. This area is defined by the inequalities [see also Eq. (11)],

$$0 \leq \alpha_1 \leq \alpha_2; \quad (1 - \alpha_1) \leq \alpha_2 \leq 1. \tag{17}$$

It describes a response where the transition toward a given strategy *increases* with the frequency of that strategy as long as that strategy is *not* the majority, i.e., minority strategies are favored. A special case where α_1 is taken from the linear voter model, whereas α_2 is larger than 0.5 is shown in Fig. 1 (right). We note in particular that social herding according to the *linear* voter model will *not* allow the transition toward cooperation, which will be further substantiated by analytical results in the next section. Further, all forms of the transition rates that *monotonously* increase with the frequency, indicated by the (pf) area, will *not* lead to cooperation. Social herding in this case only amplifies defection.

In order to also estimate the impact of the *system size*, N, on the fraction of cooperation, we present two cuts through the parameter space of α_1, α_2 in Fig. 3, i.e., α_1 varies while α_2 is fixed to either a high or a low value. In Fig. 4, the fraction of cooperation is shown for four different values of N. Differences in the symbols are hardly noticeable, which means that f is largely independent of N for system sizes of 100 and larger. Finite size effects on the fraction of cooperation play a role only for rather small systems ($N < 50$), which we did not consider in the paper.

Assuming the right choice of parameters for the transition to cooperation, we can now take a look how the dynamics evolve in space. We have chosen a two-dimensional regular lattice with Von-Neumann neighborhood, where each agent interacts with $n = 4$ local neighbors. Initially, we assume a small cluster of

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

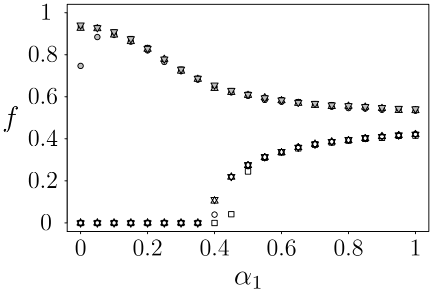

Fig. 4. Fraction of cooperation, f dependent on the nonlinearity α_1, with fixed $\alpha_2 = 0.75$ (upper curve, filled symbols) and $\alpha_2 = 0.25$ (lower curve, empty symbols). Different system sizes: $N = 100$ (□), $N = 400$ (◯), $N = 1000$ (△), $N = 10,000$ (▽). The fact that they are almost indistinguishable indicates that N does not have an impact. Other parameters: $\zeta = 0.95$.

cooperating agents. *Without* social herding, this cluster would immediately disappear in the next time step because all agents will choose defection, which is the rational choice to maximize their payoff. We observe instead a spreading of cooperation, i.e., an invasion of the cooperating strategy into the domain of defectors. The cooperating agents, however, do not form compact clusters. A minority fraction of defectors will always survive and their spatial distribution in small clusters across the domain of cooperators continues to change in time. That is, we never reach a stationary state in space, despite that the global fraction of both strategies, on average, reaches an equilibrium.

We further note that there is a critical size for the initial cluster of cooperators to grow. This has been already discussed in detail for pure PD games on a regular lattice [9, 24], and in opinion dynamics models [36]. Now, the addition of supercritical social herding of course reduces these requirements. Is it worth mentioning that, starting from random initial conditions in a spatially extended system, we find that a vanishingly small initial density of cooperators is enough to trigger the final state. The reason for this stems from the fact that, if the system is large enough, one cluster of cooperators larger than the critical size will appear by chance. This cluster will be sufficient to trigger the outbreak of cooperation. Here, however, we will not dig further on this discussion. Instead, the initial conditions and parameter constellations for the *outbreak of cooperation* will be further discussed for the mean-field case, in the next section.

4. Mean-Field Investigations

4.1. *Calculating the effort*

We verified by means of computer simulations that there is indeed a way of utilizing social herding to boost cooperation. Now, we try to illustrate this finding by some analytical considerations. As a first step, we want to calculate the "effort" to transfer the system into a majority of cooperators. Considering only the strategic dimension, this effort should be very high because there is a strong incentive to defect. On the

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

other hand, social herding may help in this situation because it neglects the payoff differences. So, it is particularly important in the first stage of the phase transition.

A formal approach to calculate the effort starts from the master equation (15) on the systemic level, in the mean-field limit. The detailed balance condition, which is a specific form of the equilibrium condition $dP(f,t)/dt = 0$, requires that the net probability fluxes are balanced, i.e.,

$$W(f\,|\,f-1/N,\zeta)P^0(f-1/N,\zeta) = W(f-1/N\,|\,f,\zeta)P^0(f,\zeta), \tag{18}$$

where $P^0(f,\zeta)$ denotes the equilibrium probability distribution which is independent of t. This equation is recursive and, using $f = N_1/N$, Eq. (2), can be reformulated as:

$$P^0(f,\zeta) = P^0(0,\zeta)\prod_{i=1}^{N_1}\frac{W\left(\frac{i}{N}\middle|\frac{i-1}{N},\zeta\right)}{W\left(\frac{i-1}{N}\middle|\frac{i}{N},\zeta\right)}. \tag{19}$$

The normalization $P^0(0,\zeta)$ can be found by enforcing $\sum_{i=0}^{N} P^0(i/N,\zeta) = 1$ and the transition rates are given by Eq. (16). We visualize the equilibrium probability distribution by means of a potential $\Omega(f,\zeta)$ that has its minimum where $P^0(f,\zeta)$ has its maximum, i.e., it represents the "effort" of reaching a given equilibrium state,

$$P^0(f,\zeta) = \exp\{-\Omega(f,\zeta)\}, \tag{20}$$

where Ω is given by

$$\Omega(f,\zeta) = -\ln P^0(0,\zeta) - \sum_{i=1}^{N_1}\ln\left[\frac{W\left(\frac{i}{N}\middle|\frac{i-1}{N},\zeta\right)}{W\left(\frac{i-1}{N}\middle|\frac{i}{N},\zeta\right)}\right]. \tag{21}$$

Figure 2 shows the effort $\Omega(f,\zeta)$ as a function of the global fraction of cooperators f and the level of social herding ζ, which acts as a control parameter. We observe that for very low values of ζ the effort is a monotonously increasing function of the frequency f. Given a fraction of cooperators, $f = 0.2$, and small ζ, it becomes more and more difficult, or unlikely, to find a larger fraction of cooperators (red line). Considering instead a high level of social herding, e.g., ζ about 0.85. There is a *monotonous decrease* of the effort with an increasing fraction of cooperators. That is, starting from a supercritical level of social herding, the outbreak and the increase of cooperation becomes very likely (green line).

The observant reader will notice in Fig. 5 for large ζ the *nonmonotonous dependence* of the effort on the fraction of cooperators. That is, there is a critical region around of $f \approx 0.2$ below which defection becomes the most probable state. This relates to the critical cluster size of cooperators in Fig. 6 to allow the transition toward cooperation. However, there is a noticeable difference underlying both results. Figure 5 is based on the mean-field limit, i.e., there is no spatial correlation between interacting agents, whereas Fig. 6 assumes a spatial neighborhood defined by the regular lattice. In fact, it is known that spatial interaction enhances cooperation [21, 24, 26]. Already small, randomly formed clusters of cooperators are

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

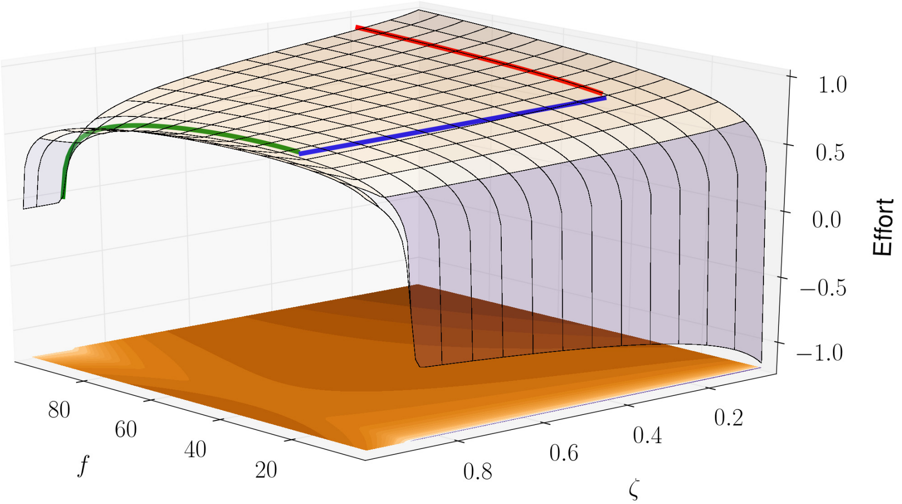

Fig. 5. (Color online) Effort $\Omega(f,\zeta)$, Eq. (21) dependent on the global fraction of cooperators f and the level of social herding ζ. The nonlinearity is specified by $\alpha_1 = 0.25$, $\alpha_2 = 0.85$.

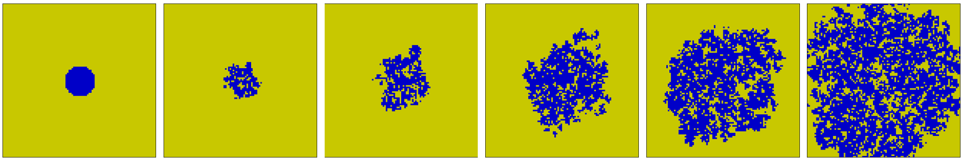

Fig. 6. (Color online) Snapshots of the transition toward cooperation at times $t = 0, 10, 20, 50, 150, 500$. $N = 10^4$ agents are placed on a regular lattice and interact each with their $n = 4$ spatial neighbors. Dark color (blue) indicates cooperators, light color (yellow) defectors. Parameters $\alpha_1 = 0.25$, $\alpha_2 = 0.7$, $\zeta = 0.95$.

sufficient for the outbreak of cooperation, whereas random interaction results in a much larger threshold.

4.2. *Competition dynamics*

Eventually, we can also derive a deterministic dynamics for the global fraction of cooperators, $f(t)$, in the mean-field limit. Basically, there are two ways of deriving this. One starts from the stochastic dynamics on the microscopic level, $p_i(\theta_i, t)$, Eq. (13) and is discussed in detail in [22]. The other one starts from the stochastic dynamics on the macroscopic level, $P(f, \zeta, t)$, Eq. (15). The expected value for the global fraction of cooperators then follows from

$$\langle f(\zeta, t)\rangle = \sum_{f'} P(f', \zeta, t), \tag{22}$$

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

where f' denote all possible realizations of f. Using the master equation (15), we arrive at the deterministic dynamics

$$\frac{d\,\langle f(\zeta,t)\rangle}{dt} = W_+(f,\zeta)(1-\langle f\rangle) - W_-(f,\zeta)\,\langle f\rangle\,, \tag{23}$$

where the aggregated transition rates $W_+(f,\zeta)$, $W_-(f,\zeta)$ are given by Eq. (16). Assuming a narrow probability distribution in equilibrium, $P^0(f,\zeta)$, the expected value $\langle f^0(\zeta)\rangle$ can be approximated by the maxima of $P^0(f,\zeta)$. In particular, the deterministic dynamics will converge to those areas where $P^0(f,\zeta)$ is largest, or where $\Omega(f,\zeta)$ has its minima, shown in Fig. 5. While we do not argue about the specific global dynamics at intermediate times (which can be governed by stochastic influences in particular in early stages), we can see the late stage of the dynamics as a "quasi-stationary" motion along the valley in the potential landscape shown in Fig. 5, provided ζ chosen large enough.

We can rewrite Eq. (23) which basically describes the "replication" of cooperators at the global scale, to make it more alike to the known replicator equation,

$$\frac{d\,\langle f(\zeta,t)\rangle}{dt} = \langle f\rangle\,(1-\langle f\rangle)[E_1(f,\zeta) - E_0(f,\zeta)]. \tag{24}$$

The two terms E_1 and E_0 are the fitness values associated with the two different strategies. The fraction of cooperation will grow if the fitness of cooperation $E_1(f,\zeta)$ is larger than the fitness of defection $E_0(f,\zeta)$, which both depend on the global level of cooperation and the level of social herding,

$$E_1(f,\zeta) = \frac{W_+(f,\zeta)}{f}; \quad E_0(f,\zeta) = \frac{W_-(f,\zeta)}{1-f}. \tag{25}$$

To evaluate the fitness values, one should note the strictly nonlinear dependence of the transition rates on f, (16). Figure 7 shows the difference $E_1 - E_0$ on the whole range of f and ζ. We emphasize that this graph holds for fixed values of the nonlinearity parameters α_1, α_2, i.e., it adds another dimension to Fig. 3, which was obtained for a fixed herding level ζ. Figure 7 also clearly shows the influence of the initial fraction of cooperators, $f(0)$, for the mean-field case. Assuming e.g., a fixed value of $\zeta = 0.85$, we see that the fraction of cooperators $f(t)$ can be increased in time only if $f(0)$ is between 0.15 and 0.6. While the lower bound has an intuitive meaning as the minimum threshold to start cooperation, the upper bound is less obvious. It results indeed from the influence of the *nonlinear* social herding, which does not simply support cooperation if that is the strategy of the majority. We recall that social herding does not assume any "value" related to the strategies. Hence, for the example considered, the maximum fraction of cooperators is given by $f = 0.6$. A higher level of social herding, or different values for the nonlinearities, may increase this fraction up to about one, i.e., full cooperation.

Adv. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

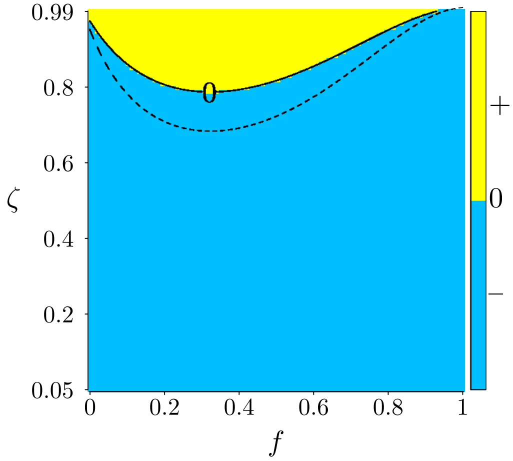

Fig. 7. (Color online) Difference of the fitness values $E_1(f,\zeta)-E_0(f,\zeta)$ dependent on the fraction of cooperators, $f\in[0.02,0.99]$, and the level of social herding, $\zeta\in[0.05,0.99]$. Dashed-line denotes the minimum level of social herding required for a given initial fraction, $f(0)$, such that the growth of cooperation starts. Contour line at 0 denotes the minimum level of social herding that ensures positive fraction of cooperators in the long term. Nonlinearity parameters: $\alpha_1=0.25$, $\alpha_2=0.8$.

Another way of expressing the dynamics of Eq. (24) is through

$$\begin{aligned}\frac{d\langle f(\zeta,t)\rangle}{dt} &= \langle f(\zeta,y)\rangle\,(E_1-\langle E\rangle);\\ \langle E\rangle &= \textstyle\sum_\sigma E_\sigma\,\langle f_\sigma\rangle = E_1\,\langle f\rangle + E_0(1-\langle f\rangle).\end{aligned}\tag{26}$$

As long as E_1 is larger than the average fitness, $\langle E\rangle$, the fraction of cooperators in the system is able to grow, but one has to recognize that, because of the time dependence of $\langle f(t)\rangle$ and its implicit feedback on E_σ, $\langle E(t)\rangle$ evolves over time as well. Hence, Eq. (26) describes a nonlinear selection process for each of the strategies dependent on the parameters describing strategic interaction and social herding.

For some special cases, we are able to derive closed form solutions of the competition dynamics expressed by Eqs. (23)–(26). In the absence of any social herding, $\zeta=0$, we just have to count in the transition rates from strategic interaction, which are $\tilde{c}_k=0$, $\tilde{d}_k=1$. This results in $E_1(f,\zeta=0)=0$ and $E_0(f,\zeta=0)=1$, i.e., the dynamics reads $\langle f(t)\rangle=f(0)\exp\{-t\}$, which means that cooperation dies out, exponentially. In the opposite case, $\zeta=1$, i.e., absence of any strategic interaction, Eq. (24) can be solved for the case of the linear voter model, which implies $\hat{c}_k=k/4$ and $\hat{d}_k=1-(k/4)$, Eq. (11). We then find $E_1(f,\zeta=1)=E_0(f,\zeta=1)$, i.e., the fitness of both strategies, which are actually mere labels without any payoff assigned, is the same. This results in the dynamics $\langle f(t)\rangle=f(0)$, i.e., a *conservation* of the initial fraction of cooperators, on average. This is known as one of the puzzles associated with the linear voter model, i.e., individual realizations of the dynamics, e.g., using stochastic simulations, always lead to convergence with $f\to 0$ or $f\to 1$,

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

but averaging over many runs reveals that the frequency at which cooperators or defectors dominate is equal to their initial fraction $f(0)$.

These two limiting cases allow us to position the dynamics if $0 < \zeta < 1$, i.e., the influence of both strategic interaction and social herding at the same time. For social herding, let us first assume the case of the linear voter model as described above. We can then verify that the closed solution for the dynamics of cooperators is given as:

$$\langle f(\zeta, t)\rangle = f(0)\exp\{(\zeta - 1)t\}, \tag{27}$$

which is similar to the case of only strategic interaction, except that the time scale for the extinction of cooperators is stretched by the factor $(1 - \zeta)$. This is an important result because it demonstrates that *linear* social herding will *not* prevent the extinction of cooperation, not even for large ζ. Hence, in order to turn defection into cooperation, we essentially need a *high* level of *nonlinear* social herding, i.e., the right ζ and α_2 values.

Considering a nonlinearity where $\alpha_1 = 1/4$ but $\alpha_2 \neq 2/4$, we find from Eq. (23)

$$\frac{d\langle f(\zeta, t)\rangle}{dt} = \langle f\rangle \{\zeta[1 + 3\langle f\rangle (1 - \langle f\rangle)^2(2\alpha_2 - 1)] - 1\}. \tag{28}$$

For $\alpha_2 = 2/4$, the solution reduces to Eq. (27), whereas for $\zeta = 1$ we arrive at the mean-field equation for the nonlinear voter model, only [22]. In order to make cooperation, $\langle f\rangle = 1$, a stable fixed point for the full dynamics, the following condition for α_2 has to be met:

$$\frac{1}{2} + \frac{1 - \zeta}{6\zeta\langle f\rangle (1 - \langle f\rangle)^2} < \alpha_2 \leq 1, \tag{29}$$

which implies $1/[1 + 3\langle f\rangle (1 - \langle f\rangle)^2] < \zeta < 1$. This inequality can be only met for a considerable high level of social herding. The feasible range of (f, ζ) values that is consistent with a given value of α_2, e.g., $\alpha_2 = 0.8$, is shown in Fig. 7. The maximum range resulting from $\alpha_2 = 1$ is also shown in the same figure by dashed line. We note again that, even if Eq. (29) is fulfilled, the dynamics does not necessarily converge to $f \to 1$. Dependening on the parameters $\{\zeta, \alpha_1, \alpha_2\}$ also lower equilibrium fractions of cooperators may be reached, i.e., we find a *coexistence of cooperation and defection.*

5. Conclusion

In this paper, we have explored a new route toward cooperation. This route differs from many other attempts, most of which are rooted in traditional or evolutionary game theory, where the transition toward cooperation is induced by specific neighborhood relations, repeated interactions, discounted payoffs over long time horizons, indirect reciprocity, favorable strategy mutations, the enforcement of social norms, etc. [10, 23, 32, 34, 35]. All of these propositions either improve the payoff of the cooperating strategy or provide, in one or another way, *additional information* agents may consider when making a strategic decision.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

Our approach is much simpler, by not changing payoffs at all, but only counting on the information agents already have if they simultaneously play a 2-Person PD game with their n neighbors (which can be spatial neighbors, or randomly chosen). This information is the local fraction of cooperators, $f_i = n_1/n$, and defectors, $(1 - f_i)$, of an agent, that also enters the calculation of the payoff, Eq. (4). That means there is *no* additional information assumed. We argue instead that agents, at the same time, respond to this information in two different ways, as summarized in Eq. (9). In a strategic interaction, they choose the strategy θ_i that will lead to the highest payoff $a_i(\theta_i, f_i)$, whereas in the case of social herding they simply respond to the local frequency of each strategy in a nonlinear manner, $\mathcal{F}(f_{\theta_i})$. In some sense, the second way assumes *less* information because no payoff matrix needs to be known. This implies that both strategies are seen as equally valuable.

The parameter ζ_i gives a weight to these two different ways of utilizing the information associated with f_i. In Eq. (9) we have assumed ζ_i to be an individual parameter, which means that agents dependent on their internal preferences or access to knowledge (such as a known payoff) can give different weights to these two responses. In this paper, however, we did not further explore this source of heterogeneity, but kept it as a global parameter, constant and the same for all agents. This limit case is equivalent of assuming a population of agents, a fraction ζ of which *only* follows social herding, whereas a fraction $(1 - \zeta)$ *only* considers strategic interactions. This allows to interpret our main result about a critical ζ to turn a population of defectors into cooperators in a more general manner: ζ can be seen as the minimal fraction of agents following only social herding, to enable the transition to cooperation. With respect to the access to information, we can interpret this finding as follows: if the information about the payoff matrix is known to all agents, they will — in the given Prisoner's Dilemma setting — collectively choose defection (which is the suboptimal state). However, if only a small fraction of agents (about 20) respond to the decision of others by means of *nonlinear* social herding, this can drive the system toward a state where cooperation is the dominant strategy. To put it succinctly: *less* information (or a larger fraction of uninformed agents) will lead to *more* cooperation.

This interesting and important conclusion is derived for the case of an evolutionary PD game with fixed payoffs and a fixed four player neighborhood, where agents follow a stochastic better response dynamics. The transition toward cooperation relies on choosing the right nonlinear social herding in response to the local (or global) fraction of cooperators. We have demonstrated that the *linear* response, where the probability to choose a strategy is directly proportional to the fraction of that strategy in the neighborhood (or the population), *fails* to enhance cooperation. Instead, we have to choose a nonlinearity, expressed in terms of the parameters α_1, α_2, from the region of *positive allee* (pa) effects (Fig. 2). As a minimal condition for the transition toward cooperation, all transition rates can be (but not necessarily have to be) chosen according to the linear voter model, except α_2, which has to be above the critical value 0.5 to break the tie in case of an equal fraction of cooperators

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

and defectors. Further, the combination of ζ and α_2 also determines the maximum level of cooperation that can be reached using the two different responses.

Our finding tells that social herding matters most in tie situations, which is also similar to another class of group decision models [6]. To design a mechanism that influences social herding only in this situation also provides a quite "cost-efficient" solution in that we will not need to enforce a decision against the majority, to allow for the transition toward cooperation. Agents can still follow the strategy of the majority — just in the undecided case, we need to ensure that the symmetry is broken into the "right" direction.

Eventually, we wish to point out that in this paper we have discussed a kind of worst-case scenario where, in the absence of social herding, defection is the only stable state for the system. Even for this case, our proposed mechanism excels in transferring defectors into cooperators, on the population level. We can leverage other model ingredients to further facilitate this transition. For example, we could count in stochastic changes of the strategy as already considered in the strategic component $\mathcal{G}(a_i)$, Eq. (5), which would support random cooperation. We can further allow for repeated interaction or "the shadow of the future" which are already known to foster cooperation [2, 3, 38]. The important message here is that, even under worst conditions there *is* a way to reach cooperation in a game-theoretical setting by means of social herding, i.e., by pure social influence. Including this additional dimension into strategic interaction avoids the lock-in into pure defection, which is the suboptimal state compared to pure cooperation. The mechanism we have proposed here does not rely on additional information, in fact it uses less of the available information, in particular no information about the payoff structure and no comparison of alternative strategies. Further, we emphasize again the "cost efficiency" of the mechanism proposed in that it does not enforce decisions against the majority, but influences the decisions of agents only in tie situations.

Summing up, adding social herding to strategic interactions is a way to substantially increase the level of cooperation with *less*, *not more*: simple rules instead of far-reaching regulations to enforce cooperation, no additional information as assumed e.g., in success driven mechanisms, no additional costs as in other incentive schemes. Just social herding, the right (nonlinear) way.

References

[1] Axelrod, R., *The Evolution of Cooperation: Revised Edition* (Basic books, 2006).

[2] Axelrod, R. and Dion, D., The further evolution of cooperation, *Science* **242** (1988) 1385–1390.

[3] Binmore, K. G. and Samuelson, L., Evolutionary stability in repeated games played by finite automata, *J. Econ. Theory* **57** (1992) 278–305.

[4] Blume, L. E., The statistical mechanics of strategic interaction, *Games Econ. Behav.* **5** (1993) 387–424.

[5] Castellano, C., Fortunato, S. and Loreto, V., Statistical physics of social dynamics, *Rev. Mod. Phys.* **81** (2009) 591–646.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.

[6] Galam, S., Sociophysics: A review of galam models, *Int. J. Mod. Phys. C* **19** (2008) 409–440.

[7] Gale, D. and Shapley, L., College admissions and the stability of marriage, *Am. Math. Mon.* **69** (1962) 9–15.

[8] Hauert, C., De Monte, S., Hofbauer, J. and Sigmund, K., Volunteering as red queen mechanism for cooperation in public goods games, *Science* (*New York, N.Y.*) **296** (2002) 1129–32.

[9] Hauert, Ch., Fundamental clusters in spatial 2x2 games, *Proc. R. Soc. Lond. Ser. B, Biol. Sci.* **268** (2001) 761–769.

[10] Helbing, D. and Yu, W., Migration as a mechanism to promote cooperation, *Adv. Complex Syst.* **11** (2008) 641–652.

[11] Hirshleifer, J., Carlos, J. and Coll, M., What strategies can support the evolutionary emergence of cooperation? *J. Conflict Resolution* **32** (1988) 367–398.

[12] Holley, R. A. and Liggett, T. M., Ergodic theorems for weakly interacting infinite systems and the voter model, *Ann. Probab.* **3** (1975) 643–663.

[13] Hølyst, J. A., Kacperski, K. and Schweitzer, F., Social impact models of opinion dynamics, in *Annual Reviews of Computational Physics*, Vol. 9 (World Scientific, 2001), pp. 253–272.

[14] Kahan, D. M., Social influence, social meaning, and deterrence, *Virginia Law Rev.* **83** (1997) 349–395.

[15] König, M., Battiston, S. and Schweitzer, F., Modeling evolving innovation networks, in *Innovation Networks: New Approaches in Modelling and Analyzing*, eds. Pyka, A. and Scharnhorst, A. (Springer, Berlin, 2009), pp. 189–269.

[16] Lorenz, J., Rauhut, H., Schweitzer, F. and Helbing, D., How social influence can undermine the wisdom of crowd effect, *Proc. Nat. Acad. Sci.* **108** (2011) 9020–9025.

[17] McKelvey, R. D. and Palfrey, T. R., Quantal response equilibria for normal form games, *Games Econ. Behav.* **10** (1995) 6–38.

[18] Miguel, M. S., Eguíluz, V. M., Toral, R. and Klemm, K., Binary and multivariate stochastic models of consensus formation, *Comput. Sci. Eng.* **7** (2005) 67–73.

[19] Nowak, M., Five rules for the evolution of cooperation, *Science* **314** (2006) 1560–1563.

[20] Parker, W. D. and Prechter, R. R., Herding: An Interdisciplinary Integrative Review From A Socionomic Perspective (2005), available at SSRN 2009898.

[21] Roca, C., Cuesta, J. and Sánchez, A., Effect of spatial structure on the evolution of cooperation, *Phys. Rev. E* **80** (2009) 046106.

[22] Schweitzer, F. and Behera, L., Nonlinear voter models: The transition from invasion to coexistence, *Eur. Phys. J. B* **67** (2009) 301–318.

[23] Schweitzer, F. and Behera, L., Optimal migration promotes the outbreak of cooperation in heterogeneous populations, *Adv. Complex Syst.* **15** (2012) 1250059.

[24] Schweitzer, F., Behera, L. and Mühlenbein, H., Evolution of cooperation in a spatial prisoner's dilemma, *Adv. Complex Syst.* **5** (2002) 269–299.

[25] Schweitzer, F. and Holyst, J. A., Modelling collective opinion formation by means of active Brownian particles, *Eur. Phys. J. B* **15** (2000) 723–732.

[26] Schweitzer, F., Mach, R. and Mühlenbein, H., Agents with heterogeneous strategies interacting in a spatial IPD, in *Nonlinear Dynamics and Heterogenous Interacting Agents*, eds. Lux, T., Reitz, S. and Samanidou, E., *Lecture Notes in Economics and Mathematical Systems*, Vol. 550 (Springer, 2005), pp. 87–102.

[27] Schweitzer, F., Zimmermann, J. and Mühlenbein, H., Coordination of decisions in a spatial agent model, *Physica A* **303** (2002) 189–216.

[28] Sood, V. and Redner, S., Voter model on heterogeneous graphs, *Phys. Rev. Lett.* **94** (2005) 178701.

[29] Stark, H.-U., Tessone, C. J. and Schweitzer, F., Decelerating microdynamics can accelerate macrodynamics in the voter model, *Phys. Rev. Lett.* **101** (2008) 018701.

[30] Stark, H.-U., Tessone, C. J. and Schweitzer, F., Slower is faster: Fostering consensus formation by heterogeneous inertia, *Adv. Complex Syst.* **11** (2008) 551–563.

[31] Suchecki, K., Eguíluz, V. M. and Miguel, M. S., Conservation laws for the voter model in complex networks, *Europhys. Lett.* **69** (2005) 228.

[32] Szabó, G. and Fath, G., Evolutionary games on graphs, *Phys. Reports* **446** (2007) 97–216.

[33] Sznajd-Weron, K. and Sznajd, J., Opinion evolution in closed community, *Int. J. Mod. Phys. C* **11** (2000) 228–234.

[34] Szolnoki, A., Perc, M., Szabó, G. and Stark, H., Impact of aging on the evolution of cooperation in the spatial prisoner's dilemma game, *Phys. Rev. E* **80** (2009) 021901.

[35] Tessone, C. J., Sánchez, A. and Schweitzer, F., Diversity-induced resonance in the response to social norms (2012), to appear in *Phys. Rev. E.* arXiv:1209.5518.

[36] Tessone, C. J., Toral, R., Amengual, P., Wio, H. S. and San Miguel, M., Neighborhood models of minority opinion spreading, *Eur. Phys. J. B* **39** (2004) 535–544.

[37] Thurstone, L., A Law of Comparative Judgement, *Psychol. Rev.* **101** (1994) 226–270.

[38] Trivers, R. L., The evolution of reciprocal altruism, *Q. Rev. Biol.* **46** (1971) 35–57.

[39] Vazquez, F., González-Avella, J. C., Equíluz, V. and Miguel, M. S., Collective phenomena in complex social networks, in *Applications of Nonlinear Dynamics Model and Design of Complex Systems*, eds. In, V., Longhini, P. and Palacios, A. (Springer-Verlag, 2009).

[40] Wu, F., Huberman, B. A., Adamic, L. A. and Tyler, J. R., Information flow in social groups, *Physica A, Stat. Mech. Appl.* **337** (2004) 327–335.

[41] Zimmermann, M. G., Eguíluz, V. M. and San Miguel, M., Coevolution of dynamical states and interactions in dynamic networks, *Phys. Rev. E* **69** (2004) 065102.

Advs. Complex Syst. 2013.16. Downloaded from www.worldscientific.com
by SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH (ETH) on 08/18/14. For personal use only.