Impact of fraud on the mean-field dynamics of cooperative social systems

Torsten Röhl,[1] Claudia Röhl,[2] Heinz Georg Schuster,[1] and Arne Traulsen[3]
[1]*Institute of Theoretical Physics and Astrophysics, University of Kiel, Leibnizstraße 15, D-24098 Kiel, Germany*
[2]*Institute of Anatomy, University of Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany*
[3]*Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138, USA*
(Received 12 March 2007; published 27 August 2007; corrected 6 September 2007)

The evolution of costly cooperation between selfish individuals seems to contradict Darwinian selection, as it reduces the fitness of a cooperating individual. However, several mechanisms such as repeated interactions or spatial structure can lead to the evolution of cooperation. One such mechanism for the evolution of cooperation, in particular among humans, is indirect reciprocity, in which individuals base their decision to cooperate on the reputation of the potential receiver, which has been established in previous interactions. Cooperation can evolve in these systems if individuals preferably cooperate with those that have shown to be cooperative in the past. We analyze the impact of fake reputations or fraud on the dynamics of reputation and on the success of the reputation system itself, using a mean-field description for evolutionary games given by the replicator equation. This allows us to classify the qualitative dynamics of our model analytically. Our results show that cooperation based on indirect reciprocity is robust with respect to fake reputations and can even be enhanced by them. We conclude that fraud *per se* does not necessarily have a detrimental effect on social systems.

DOI: 10.1103/PhysRevE.76.026114 PACS number(s): 89.65.−s, 87.23.Ge, 05.45.−a

I. INTRODUCTION

Evolution is a constant struggle for survival. Individuals compete continuously. Natural selection implies that individuals should not support others at a cost to themselves. Thus, it is surprising that individuals are often willing to forgo some of their reproductive potential and support others instead. While such an action reduces fitness in the short run, it might increase the survival chances in the long run. This phenomenon has been explained in terms of mechanisms for the evolution of cooperation, which include kin selection [1,2], group selection [3–5], network reciprocity [6–16], direct reciprocity [17–19], and indirect reciprocity [20–28]; see [29] for a recent review. Among humans, a particularly interesting mechanism is indirect reciprocity, which "presumably may distinguish us humans from all other living species on earth" [30].

Here, we model indirect reciprocity, where cooperation is based on a status or reputation that individuals obtained in previous interactions. Indirect reciprocity is frequently applied to improve trade on online platforms [31–33]. Recently, it has been shown that it is also a potential mechanism that can help to address such global cooperation problems as climate preservation [34]. Theoretical work on reputation systems has considered the influence of unintended errors [27,28]: A "trembling hand" can lead to the wrong action and a "fuzzy mind" might lead to a wrong assessment and hence a wrong reputation. However, a possible manipulation of reputation by the players is usually not taken into account. Furthermore, these models do not involve any time delays and typically assume that an action is immediately assessed [20,35]. In reality, this is not always the case: We discover that we have been ripped off when it is too late, read the fine print of a contract when we have already signed, or discover that public funds are missing when it is not longer possible to backtrack the scammers. Although their behavior is "bad," these scammers can maintain a "good" reputation for some time. Indeed, in human social systems generally a small fraction of people exist who break the rules for their own advantage and are sanctioned by the community [36]. As this kind of "fraud" is fake cooperation, it is different from public defection and also the effect of punishing this action can deviate from the conventional ones [37,38]. So far, it is unclear why sanctions by the community have not eliminated fraud from social systems. Also the impact of fraud on the evolution of cooperation has not been analyzed yet. We address these questions by extending the image scoring framework described by Nowak and Sigmund [20]. Image scoring is a particular simple moral system assigning a reputation to an action and determining the choice of an action based on such reputation. As image scoring bases reputation only on actions and is independent of previous reputations, it is especially robust if information on previous encounters is not reliable. Theoreticians have criticized this framework, as more sophisticated moral systems avoid the problem that a good individual who refuses to cooperate with a bad individual immediately becomes bad [36,39,40]. Moreover, image scoring is not evolutionarily stable and does not belong to the "leading eight" moral systems of Ohtsuki and Iwasa, who analyzed 4096 moral systems and identified the eight most cooperative of them [21]. However, behavioral experiments have shown that image scoring is a plausible mechanism for cooperation among humans and is preferred under many circumstances over more sophisticated moral systems [41–43]. It also requires only minimal information. Thus, image scoring seems to be a reasonable starting point to address the problem of fraud. We base our model on the replicator equation, which provides a mean-field description of the dynamics of game-theoretic systems [44,45]. This allows us to classify the different dynamical regimes of the system analytically.

II. THE MODEL

Here, we first recall the image scoring framework and discuss the intuitive meaning of the parameters. The detailed mathematical implementation of the process is described in Secs. II A and II B. In the image scoring framework, two individuals are chosen at random, one as a potential donor and one as a recipient. The donor has the possibility to pay the cost c to support the recipient, who then obtains the benefit $b>c$. A pair of individuals only meets once. Therefore, the donor does not expect a return from the recipient himself, but hopes that through establishing a good reputation, the cost for his cooperation will be compensated by the benefit from an altruistic act of someone else in future encounters. We assume that the donor does not support the recipient directly, but invests in a public good and allows the recipient to take advantage of this good. Discriminators aim to improve their payoff by investing the cost c in their reputation and thus attract help from others, because the reputation, which is initially good for all individuals, only stays good if they help others and keep public confidence in this way. If they deny help, their reputation becomes bad. Defectors avoid the risk of investments in their reputation and do not support others. Thus, their good reputation from the beginning of each round remains good only until they are chosen as potential donors and deny help. Whereas in the short run nondonors yield the higher payoff by saving the costs for cooperation, in the long run cooperators increase the chance of obtaining a benefit based on their good reputation, and cooperation might thus yield the higher payoff.

This reputation-based cooperation can be exploited by "scammers" who manipulate their own reputation [23], especially if these do not need to fear imminent punishment. Therefore, we introduce an additional strategy, namely, fraud. Scammers only pretend to invest in the public good and encourage others to use the common good. While this allows them to maintain a good reputation, it undermines the system in the long run, since the common resource is overused. In reality, this can happen if transfers are made via anonymous public funds, if checks bounce, or if credit cards are misused. If the society is unable to detect and prevent such fraud, it will quickly spread and destroy the system. However, if there is a certain probability α $(0\leq\alpha\leq 1)$ that fraud is discovered, the situation becomes more interesting. For $\alpha=0$, fraud is never discovered and will spread in the population. For $\alpha=1$, reputation can never be faked successfully. Whenever a scammer is discovered, he has to pay a penalty proportional to the number of potential donors, $\rho(x+z)$, but does not change his behavior. The constant ρ can take any real value, but it seems to be reasonable that it is of the same order of magnitude as the benefit from cooperation, b. This leads to a new kind of dynamics between discriminators, defectors, and scammers. We follow Nowak and Sigmund [20] and model the dynamics with two different stages. On a fast time scale of interactions, reputations change and payoffs are accumulated. It is assumed that the information on the new reputation is available to all individuals before the next round occurs. On a slower time scale, the differences between the accumulated payoffs leads to a change of strategies. In this process, the numbers of discriminators, defectors, and scammers change.

A. Change of reputation and payoff accumulation

We consider three different types of individuals: discriminators, defectors, and scammers. All three types can have good or bad reputation (or image score). At the beginning of each generation, the image score is set to good for all individuals. Individuals interact for several rounds. The frequency of players with bad and good reputations, i.e., image score $i=0,1$, is denoted as x_i (discriminators), y_i (unconditional defectors), and z_i (scammers).

The frequency of individuals with image scores 0 or 1 changes from round to round, since a donation can change the reputation. An upper index denotes the round; e.g., x_1^2 is the frequency of discriminators with a good reputation in round 2. In the first round, all individuals have image score 1. Thus, the initial condition of each generation is given by $x_0^1=y_0^1=z_0^1=0$. The frequencies x_1^1, y_1^1, and z_1^1 (which sum up to 1) reflect the composition of the population. This composition changes based on the success of the strategies on a slower time scale (see below). In round 1, individuals obtain payoffs based on the initial reputation, which is always positive. Consequently, the reputation in round j depends on the actions in round $j-1$. Thus, actions in round j determine the future payoff obtained in round $j+1$.

For example, consider a discriminator with a good reputation. If he is chosen as a potential donor (which happens with probability 1/2), his image score changes. If he is paired with an individual in good reputation, he cooperates and his reputation remains good. If he is paired with an individual in bad reputation, he does not cooperate. Then, his reputation becomes bad. The frequency of discriminators with a good reputation decreases due to this process from round j to round $j+1$. However, it increases when discriminators with a bad reputation cooperate again with others. Thus, x_1 changes from round j to round $j+1$ as

$$x_1^{j+1} = x_1^j + [x_0^j\varphi^j - x_1^j(1-\varphi^j)]/2. \quad (1)$$

Here $\varphi^j=x_1^j+y_1^j+z_1^j+z_0^j$ is the fraction of all players who obtain help in round j. In an equivalent similar way, we obtain

$$x_0^{j+1} = x_0^j + [x_1^j(1-\varphi^j) - x_0^j\varphi^j]/2,$$

$$y_0^{j+1} = y_0^j + y_1^j/2, \quad y_1^{j+1} = y_1^j - y_1^j/2,$$

$$z_0^{j+1} = z_0^j + [z_1^j(1-\varphi^j) - z_0^j\varphi^j]/2,$$

$$z_1^{j+1} = z_1^j + [z_0^j\varphi^j - z_1^j(1-\varphi^j)]/2. \quad (2)$$

The total numbers of discriminators $x=x_0^j+x_1^j$ remains constant during a generation. Equivalently, also $y=y_0^j+y_1^j$ and $z=z_0^j+z_1^j$ remain constant. Note that the real reputation of scammers changes despite the fact that their fake reputation is always good. We note that the reputation dynamics depends only on the fraction of players with different reputation and strategies. It is independent of the parameters of the underlying game, i.e., the cost c, the benefit b, the penalty ρ, and the detection probability α.

Based on the fractions of players in good and in bad reputation, we now calculate the average payoffs P_C, P_D, and P_S of cooperative discriminators, defectors, and scammers, respectively. We denote the payoff of a discriminator with good reputation as $P_{C,1}$ and the payoff of a discriminator with bad reputation as $P_{C,0}$. Equivalent notation is used for defectors and scammers. When an individual is chosen as a donor (which happens with probability 1/2) and cooperates, then the cost of cooperation c is subtracted from its payoff. For example, a discriminator with a bad reputation cooperates with probability φ^j in round j. Thus, its payoff changes on average by $-c\varphi^j/2$ in that round. Cooperation leads to the benefit b. For example, an unconditional defector in good reputation increases his payoff by b whenever he interacts with a cooperator or a scammer and is chosen as a recipient. Thus, its payoff changes by $+b(x+z)/2$. For scammers, we have to distinguish two cases. With probability $1-\alpha$, they are viewed as individuals with a good reputation, but never pay the cost for cooperation. With probability α, they are punished for their fraud. The average payoffs for all players obtained in round j are

$$P_{C,0}^j = -\varphi^j c/2,$$

$$P_{C,1}^j = [-\varphi^j c + b(x+z)]/2,$$

$$P_{D,0}^j = 0, \quad P_{D,1}^j = b(x+z)/2,$$

$$P_{S,0}^j = P_{S,1}^j = (1-\alpha)b(x+z)/2 - \alpha\rho(x+z)/2, \tag{3}$$

where c is the cost, b the benefit, α the detection probability, and ρ the penalty for fraud. Figure 1 shows examples of the dynamics of reputations and payoffs during a generation.

Equations (1) and (2) indicate that there are always some players of a certain strategy that are in good reputation and others that are in bad reputation. The success of a strategy depends on the accumulated payoffs averaged over these two reputations. After n rounds, discriminators have accumulated the average payoff

$$\mathcal{P}_C = \sum_{j=1}^{n} \frac{P_{C,0}^j x_0^j + P_{C,1}^j x_1^j}{x_0^j + x_1^j}. \tag{4}$$

The sums for defectors and scammers simplify to closed analytical expressions; defectors accumulate the payoff

$$\mathcal{P}_D = \sum_{j=1}^{n} \frac{P_{D,0}^j y_0^j + P_{D,1}^j y_1^j}{y_0^j + y_1^j} = b(x+z)(1-2^{-n}) \tag{5}$$

and scammers obtain

$$\mathcal{P}_S = \sum_{j=1}^{n} \frac{P_{S,0}^j z_0^j + P_{S,1}^j z_1^j}{z_0^j + z_1^j} = \frac{(1-\alpha)b - \alpha\rho}{2}(x+z)n. \tag{6}$$

For $n=0$, no interactions take place and all payoffs are zero. For a single interaction, $n=1$, defectors have always the highest payoff among the three strategies. Discriminators pay the cost of cooperation, but cannot take advantage of their reputation. Scammers avoid paying the cost, but with probability α they are discovered and have to pay the punishment ρ. Thus defectors are more successful than scammers and fraud does not pay for a single interaction (cf. Fig. 1). Whether scammers or discriminators obtain a higher payoff depends on the parameters and on the initial condition. From $\mathcal{P}_C > \mathcal{P}_S$ we obtain with $n=1$ the condition $\alpha(x+z) > c/(b+\rho)$. In this case, discriminators are more successful than scammers. For $n \geq 2$, the dynamics becomes more complex.

Based on these accumulated payoffs for the short-term dynamics, we can now address the long-term dynamics that changes the fraction of discriminators, defectors, and scammers.

B. Change of strategies

The average payoffs that are accumulated over n rounds determine how successful a strategy is. As usual in evolutionary game theory, we equate payoff and fitness and players produce offspring proportional to their payoff. Note that each strategy consists of players with good reputation and of players with bad reputation. The accumulated payoffs considered here are the averages over the two reputations. We use the replicator dynamics to identify strategies that are successful in the long run based on their payoffs [45]. There are different microscopic processes that lead to slightly different differential equations [46–49]. As they do not change the stability of fixed points, we do not have to consider these alternative descriptions here. In the replicator equation, the fraction of the three strategies changes as

$$\dot{x} = x(\mathcal{P}_C - \langle \mathcal{P} \rangle),$$

$$\dot{y} = y(\mathcal{P}_D - \langle \mathcal{P} \rangle),$$

$$\dot{z} = z(\mathcal{P}_S - \langle \mathcal{P} \rangle), \tag{7}$$

where $\langle \mathcal{P} \rangle = x\mathcal{P}_C + y\mathcal{P}_D + z\mathcal{P}_S$ is the average payoff in the population. The replicator dynamics does not change the normalization, $x+y+z=1$. Due to the way that payoffs are calculated in Eqs. (3)–(5), the system is highly nonlinear. Note that the payoffs and hence the dynamics depends for a given initial condition on the number of rounds n, the cost c, the benefit b, the punishment ρ, and the detection probability α.

III. RESULTS

The global dynamics of the system is qualitatively determined by the dynamics between two strategies. As a starting point for our analysis, we assume that the detection probability α is independent of the number of scammers.

A. Constant detection probability

First, we consider the case of discriminators and scammers only. In this case, there are no individuals that do not receive help, $\varphi = x_1 + z = 1$. Hence, we have $\mathcal{P}_C = \mathcal{P}_S$ for detection probability

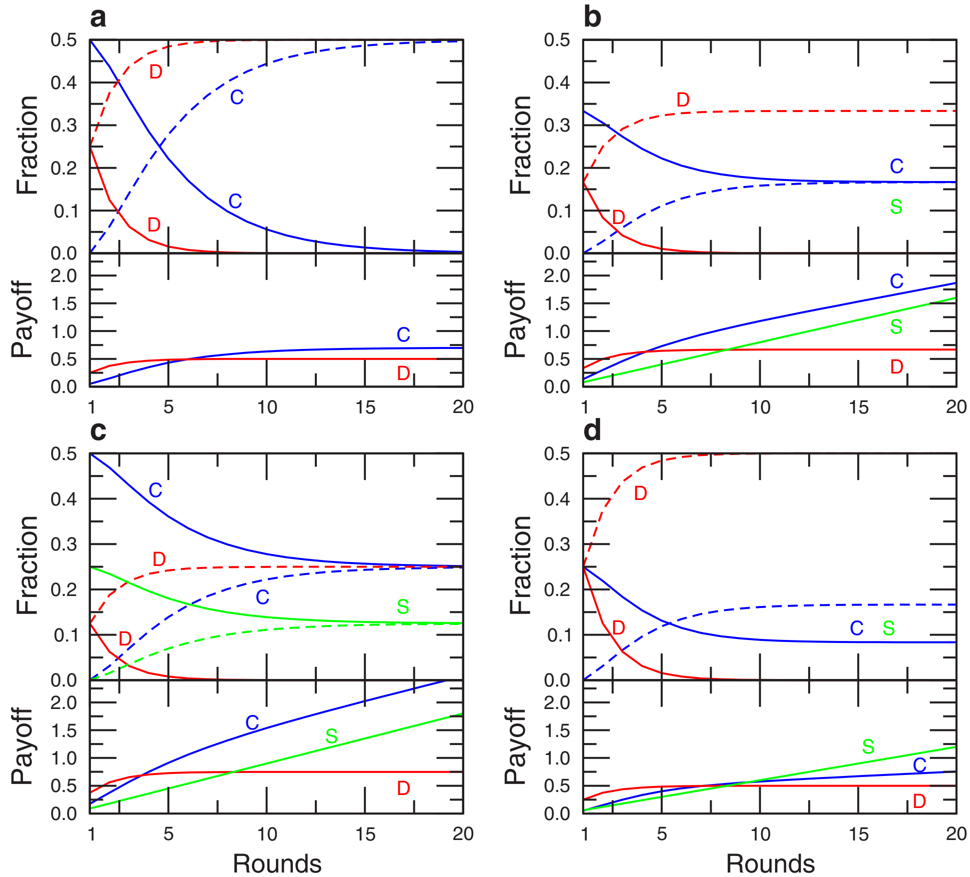

FIG. 1. (Color online) Time evolution of reputation and payoffs for the three different types (discriminating cooperators C with fraction x, defectors D with fraction y, scammers S with fraction z) for four different initial conditions. Initially, all individuals are in good reputation. Full lines show the dynamics of the fraction of individuals with good reputation, dashed lines show the fraction of individuals with bad reputation. (a) Without scammers, the reputation of both discriminating cooperators and defectors becomes bad after a few rounds. If the number of rounds is high enough, the accumulated payoff of discriminators becomes higher, which would in the long-term dynamics lead to an increase of their fraction (initial condition $x=0.5$, $y=0.5$). (b) Dynamics starting from a symmetric mixture of all three strategies, $x=y=z=1/3$. Because of the same initial values for discriminators and scammers, both have the same distribution, but different payoffs [cf. Eq. (2)]. Here, the payoff of discriminators is highest for more than four rounds, which leads to an initial increase of the fraction of discriminators in the replicator dynamics. (c) Starting from $x=0.5$, $y=0.25$, and $z=0.25$, the number of discriminators and defectors with bad reputation reaches $x_0^\infty=0.25$ in the long run. Cooperators again obtain the highest payoff if the number of rounds is sufficiently high. (d) If the initial condition is $x=0.25$, $y=0.5$, and $z=0.25$, the number of rounds determines the winning strategy: For only one round, defectors are most successful. If the number of rounds n satisfies $1<n<8$, then cooperators are most successful. Finally, scammers have the highest payoff for $n>8$ (in all panels, the parameter values are $\alpha=0.38$, $b=1.0$, $c=0.4$, $\rho=1.0$).

$$\alpha_1 = \frac{c}{b+\rho}, \tag{8}$$

independent of n. An α larger than α_1 leads to a drift toward discriminators, a smaller α to a drift toward scammers. However, a society dominated by scammers is not possible, as this would produce benefits at no costs. Hence, reputation loses its meaning before fraud takes over. In reality, there are additional constraints, e.g., a maximum fraction of scammers. In order to tackle fraud, one should aim for a low cost to benefit ratio c/b and a high penalty to benefit ratio ρ/b if it is not possible to increase the detection probability α.

Without discriminators, $x=0$, the critical detection probability can be computed from $\mathcal{P}_D=\mathcal{P}_S$, which yields

$$\alpha_2 = \frac{b}{b+\rho}\left(1-\frac{2}{n}+\frac{2^{1-n}}{n}\right). \tag{9}$$

For $\alpha>\alpha_2$, defectors are better off than scammers in the absence of discriminators and fraud vanishes ultimately as illustrated in Fig. 2(b). Note that, for large n, the ratio α_1/α_2 reduces to the cost to benefit ratio of cooperation c/b, which appears to be a crucial parameter for all mechanisms of cooperation [29].

The dynamics between discriminators and defectors depends only on the cost to benefit ratio and on the number of rounds in which the reputation of individuals changes [20,50]. If the number of rounds is sufficiently high, the dynamics is bistable. The position of the unstable equilibrium is given by the numerical solution of $\mathcal{P}_C=\mathcal{P}_D$.

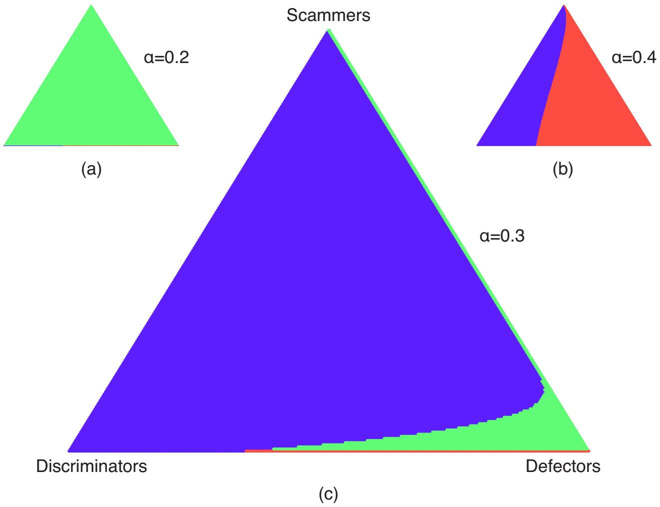

FIG. 2. (Color online) Simplices representing the evolutionary dynamics of the discriminator, defector, and scammer strategies for different detection probabilities α. Each point in the given area is colored depending on the fixed point of the dynamic. The corners represent the pure strategies. The sides represent competition between two strategies. For all diagrams the same parameter values are taken except for the probability α that a scammer is detected. (a) If the probability that a scammer is detected is below $\alpha_1=0.25$, scammers can take over the population ($\alpha=0.2$). Discriminators and defectors survive only if no scammers are initially present. (b) For $\alpha>\alpha_2\approx0.336$, scammers go extinct since they are discovered too frequently. Depending on the initial condition, cooperators or defectors prevail ($\alpha=0.4$). (c) For intermediate detection probabilities, $\alpha_1<\alpha<\alpha_2$, the dynamics leads from defectors to scammers, but from scammers to discriminators ($\alpha=0.3$). The basin of attraction of the discriminators is significantly larger than in situations with high α [$b=1.0$, $\rho=1.0$, $c=0.5$, $n=6$, blue (dark gray) are discriminators, red (medium gray) are defectors, green (light gray) are scammers].

Let us now return to the full dynamics of all three strategies (see Fig. 2). As described above, for $\alpha<\alpha_1$, scammers dominate the system. For $\alpha>\alpha_2$, scammers have no influence on the dynamics. The replicator dynamics of the system has two stable fixed points, unconditional defection and discriminating cooperation. A very different situation is observed for intermediate detection probability $\alpha_1<\alpha<\alpha_2$ [Fig. 2(c)]. The size of this region increases with the number of rounds n, since α_2 is an increasing function of n. For large n, it is given by $\alpha_2-\alpha_1\approx(b-c)/(b+\rho)$ and increases with larger $b-c$, but decreases with larger $b+\rho$.

Qualitatively, the situation is the same for all parameters within this region: While discriminators prevail compared to scammers, the scammers can still outperform defectors. The fixed point of unconditional defection is unstable and the dynamics leads to scammers. Errors on the trajectory toward scammer dominance will inevitably lead to the revival of discriminating cooperators. Similar, mutations in the strategies or stochasticity arising from a finite population size can also lead from the very narrow path from defectors toward scammers (cf. Fig. 2) into the basin of attraction of discriminators. These will finally take over the system and cannot be outperformed by the other two strategies. Hence, fraud has a counterintuitive positive effect on the evolution of cooperation, as it considerably enlarges the basin of attraction of discriminators and destabilizes the situation in which only defectors are present. Interestingly, when few scammers are present, discriminators benefit from this when competing with defectors, as their basin of attraction is increased.

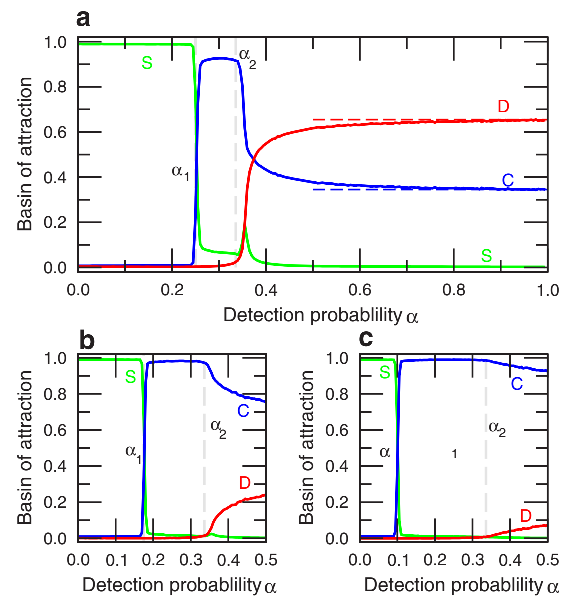

FIG. 3. (Color online) Size of the basin of attraction (cf. Fig. 2) of the different stable fixed points in dependence on the detection probability α. As in Fig. 1, C denotes cooperating discriminators, D unconditional defectors, and S scammers. (a) In a system with high cost ($c=0.5$), scammers dominate the system for $\alpha<\alpha_1=0.25$. For $\alpha_1<\alpha<\alpha_2=0.336$, the basin of attraction of discriminators is greatly enlarged, but scammers still dominate over defectors. If $\alpha>\alpha_2$, then defectors and discriminators dominate. Compared to the analytical result for the situation without any scammers (horizontal dashed lines), discriminators have a larger basin of attraction. (b) For smaller costs of cooperation, $c=0.35$, the value of α_1 is reduced to 0.175, whereas α_2 remains constant. Without scammers, discriminators have a larger basin of attraction than defectors, but in our case it is further enlarged due to the initial presence of scammers. (c) If we decrease the cost to $c=0.2$, the value of α_1 is reduced to 0.1. In all three panels, each data point is an average over 10^5 initial conditions. For each initial condition, we solved the replicator equations (6) numerically using a Euler discretization with $\Delta t=0.01$. After $T=10^5$ time steps, we determined the strategy with the highest abundance. The small peaks of the scammer curve in (a) and (b) for values of $\alpha>\alpha_2$ indicate a very slow dynamics, in which the trajectory is after T time steps still close to the unstable scammer fixed point. For $\Delta t\to0$ and $T\to\infty$, no scammers are present for $\alpha>\alpha_2$ (in all panels, the parameter values are $b=1.0$, $\rho=1.0$, $n=6$ rounds).

To demonstrate the positive effect of scammers and the increase of the $\alpha_1<\alpha<\alpha_2$ region with decreasing cost of cooperation c, we have numerically calculated the basin of attraction (see Fig. 3) for three different costs of cooperation. The numerical results are in very good accordance with the analytical results. For high α the proportion of discriminators and defectors are determinined by the ratio c/b. For $c=0.5$ and $\alpha<\alpha_2$, the discriminators can even have a larger basin than defectors, in contrast to the situation without scammers. For small c, discriminators are advantageous compared to defectors in the absence of scammers. Nonetheless, in the presence of scammers their basin of attraction increases.

For $n=3$ rounds, one can analytically show that fraud enhances cooperation. In a system without fraud, $z=0$, and $n=3$ rounds, there is an unstable fixed point x^* for the dynamics between discriminators and defectors if $0<5c<12b$ given by

$$x^* = -2 + \frac{\sqrt{4b^2 - bc - 3c^2}}{b-c}. \tag{10}$$

If we add a small fraction of scammers near this fixed point, we can ask whether the difference between discriminator and defector payoff becomes positive or negative. For $z \ll 1$, we find

$$\mathcal{P}_C - \mathcal{P}_D \approx \left(\frac{2b-c}{8}x^* + \frac{4b-5c}{8}\right)z > 0. \tag{11}$$

Since this payoff difference is always positive for $c<b$, discriminators are always better off when few scammers are present compared to situations without scammers. For $n=2$. the reasoning is similar: For $0<4c<b$, the unstable fixed point is given by $x^*=3c/(b-c)$. For small z we find near the fixed point $\mathcal{P}_C-\mathcal{P}_D \approx (b-c)z/4$, which is positive if $c<b$. Thus, for $n=2$ and $n=3$ the basin of attraction of discriminators grows due to the presence of scammers, regardless of which strategy ultimately prevails.

B. Adaptive detection probability

The analysis in the previous paragraph is a necessary prerequisite to tackle the more realistic case in which the probability to detect scammers increases with their presence. If only a small fraction of the population steals from public funds, this is unlikely to have a detrimental effect on the system and will not be detected. However, if this fraction grows to larger values, such fraud endangers the common enterprise and the detection probability grows. The simplest approach to such a dynamic detection rate is to make it proportional to the number of scammers, i.e., $\alpha=\beta z$. In this way, it is hard to detect a small number of scammers, while it becomes significantly easier to detect them if their number increases. With this extension, the fixed points at the pure strategies are destabilized (see Fig. 4). The ultimate outcome of the dynamics is given by stable fixed points where a certain fraction of scammers is present. If discriminators and scammers coexist this fraction is given by α_1/β [see Eq. (8)]. If defectors and scammers coexist, it is given by α_2/β [see Eq. (9)].

However, we do not have to restrict ourselves to a linear function for the detection probability: Our results hold for any strictly increasing function $f(z)$. A steeper increase leads to a smaller fraction of scammers, which is given by the solution of $f(z)=\alpha_1$ (absence of defectors) or $f(z)=\alpha_2$ (absence of discriminators). Thus, the stable fixed points of the system can be calculated analytically for any strictly increasing function $f(z)$. Besides the three trivial unstable fixed points at $x=1$, $y=1$, and $z=1$, the system also has two nontrivial unstable equilibria. The fixed point between discriminators and defectors has been calculated from $\mathcal{P}_C=\mathcal{P}_D$ in Eq. (9) for the special case of $n=3$. In general, it can only be determined numerically. The fixed point in the interior that appears only for adaptive detection probability has to be determined numerically from $\mathcal{P}_C=\mathcal{P}_D=\mathcal{P}_S$ for a given set of parameters.

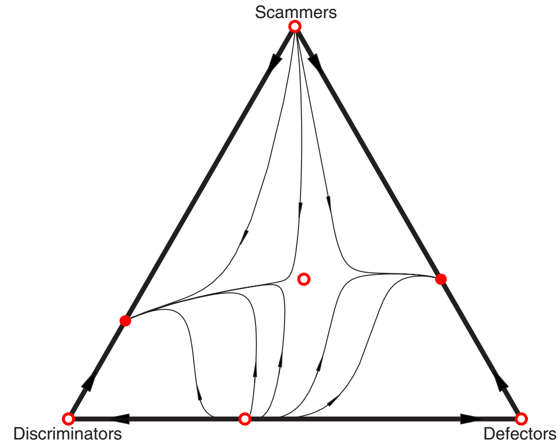

FIG. 4. (Color online) Dynamics between discriminators, defectors, and scammers with a scammer detection probability α proportional to their abundance z in the system $\alpha(z)=\beta z$. Since scammers are very successful when their abundance is low, situations without scammers become unstable. However, since high abundance of scammers implicates a high detection probability, the fixed point with scammers only is destabilized. Depending on the initial condition, the system ends up in one of two fixed points. (i) For a coexistence of discriminators and scammers, the fraction of scammers is given by $\alpha_1/\beta=0.3$. (ii) For coexistence of defectors and scammers, the fraction of the latter is given by $\alpha_2/\beta=0.358$ ($b=1.0$, $c=0.5$, $\rho=1.0$, $n=7$, $\beta=1$).

IV. DISCUSSION

Here, we have introduced a framework for reputation mechanisms that takes into account fake reputations. Many social, political, and economical systems show characteristics that emerge from differences in real and fake reputations. For example, think of political scandals where individuals act at the expense of the community, not paying any costs until they are detected or, in the worst case, the system breaks down. This living at the expense of others is typically feasible during a limited amount of time only. The simplicity of our model allows a future comparison with behavioral experiments, which have successfully corroborated several theoretical results in game theory [41–43,51]. By concentrating on the mean-field dynamics of the system, we are able to obtain an analytical classification of the dynamics of the system.

In a finite population, stochastic effects would change this prediction. Without errors or mutations that lead to strategies that are not present, the system will ultimately reach one of the corners of the simplex. For small error rates, the new strategy is lost again or adapted by the whole population before a second error arises [52–54]. Thus, the stationary distribution is determined by the transition rates between the corners of the simplex. In this case, one would expect that discriminators dominate. However, this analytical approach involves at least a temporary dominance of scammers, which might not be feasible in real systems. For higher error rates, errors in the scammer corner of the simplex or on the trajectories from defectors to scammers in the parameter region $\alpha_1 < \alpha < \alpha_2$ will lead into the basin of attraction of discriminators. Even with high error rates, the system will spend most of the time in the adjacency of the discriminator corner. Thus, the possibility of faking image scores together with a small probability for errors can lead to cooperation based on reputation in this system. A possible extension of this model is to consider interactions on social networks. However, only the simplest cases of fixed [6–14,55–62] or evolving networks [63–68] allow tackling these problems analytically. In our case, additional complications occur due to the nonlinearity in the calculation of payoffs, which makes most analytical approaches unfeasible.

In conclusion, our results show that in indirect reciprocity discriminators might benefit from coexistent scammers that fake their reputation, depending on their detection probability and on the cost to benefit ratio. If scammers dominate over defectors, they can help discriminators to initiate cooperation, as the presence of scammers allows discriminators to obtain a good reputation. Once defectors are rare, scammers are displaced by discriminators. When the probability to detect scammers vanishes with the fraction of scammers, a certain amount of fraud is always found in the system. Thus, a limited presence of scammers in the population can increase cooperative behavior. Complex cooperative systems become vulnerable to self-interested scammers when a critical number is exceeded or if they cannot be detected at low abundance. This could explain why evolution did not eliminate fraud from social systems.

ACKNOWLEDGMENTS

We thank J. C. Claussen for many insightful discussions and comments on this manuscript. A.T. acknowledges support by the “Deutsche Akademie der Naturforscher Leopoldina” (Grant No. BMBF-LPD 9901/8-134).

[1] W. D. Hamilton, Am. Nat. **97**, 354 (1963).
[2] S. A. Frank, *Foundations of Social Evolution* (Princeton University Press, Princeton, 1998).
[3] J. Maynard Smith, Q. Rev. Biol. **51**, 277 (1976).
[4] D. S. Wilson, Proc. Natl. Acad. Sci. U.S.A. **72**, 143 (1975).
[5] A. Traulsen and M. A. Nowak, Proc. Natl. Acad. Sci. U.S.A. **103**, 10952 (2006).
[6] M. A. Nowak and R. M. May, Nature (London) **359**, 826 (1992).
[7] A. V. M. Herz, J. Theor. Biol. **169**, 65 (1994).
[8] K. Lindgren and M. G. Nordahl, Physica D **75**, 292 (1994).
[9] G. Szabó and C. Tőke, Phys. Rev. E **58**, 69 (1998).
[10] F. Schweitzer, L. Behera, and H. Mühlenbein, Adv. Complex Syst. **5**, 269 (2002).
[11] G. Szabó and C. Hauert, Phys. Rev. Lett. **89**, 118101 (2002).
[12] F. C. Santos, J. F. Rodrigues, and J. M. Pacheco, Phys. Rev. E **72**, 056128 (2005).
[13] F. C. Santos and J. M. Pacheco, Phys. Rev. Lett. **95**, 098104 (2005).
[14] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak, Nature (London) **441**, 502 (2006).
[15] M. Perc, New J. Phys. **8**, 22 (2006).
[16] G. Szabó and G. Fáth, Phys. Rep. **446**, 97 (2007).
[17] R. L. Trivers, Q. Rev. Biol. **46**, 35 (1971).
[18] R. Axelrod and W. D. Hamilton, Science **211**, 1390 (1981).
[19] C. Hauert and H. G. Schuster, Proc. R. Soc. London, Ser. B **264**, 513 (1997).
[20] M. A. Nowak and K. Sigmund, Nature (London) **393**, 573 (1998).
[21] H. Ohtsuki and Y. Iwasa, J. Theor. Biol. **231**, 107 (2004).
[22] H. Brandt and K. Sigmund, J. Theor. Biol. **231**, 475 (2004).
[23] H. Brandt and K. Sigmund, Proc. Natl. Acad. Sci. U.S.A. **102**, 2666 (2005).
[24] M. A. Nowak and K. Sigmund, Nature (London) **437**, 1291 (2005).
[25] F. A. C. C. Chalub, F. C. Santos, and J. M. Pacheco, J. Theor. Biol. **241**, 233 (2006).
[26] J. M. Pacheco, F. C. Santors, and F. A. C. C. Chalub, PLOS Comput. Biol. **2**, 1634 (2006).
[27] H. Brandt and K. Sigmund, J. Theor. Biol. **239**, 183 (2006).
[28] H. Ohtsuki and Y. Iwasa, J. Theor. Biol. **239**, 435 (2006).
[29] M. A. Nowak, Science **314**, 1560 (2006).
[30] R. Alexander, *The Biology of Moral Systems* (de Gruyter, New York, 1987).
[31] G. Bolton, E. Katok, and A. Ockenfels, Manage. Sci. **50**, 1587 (2004).
[32] C. Dellarocas, MIT Sloan School of Management working paper, 4297 (2003).
[33] T. Pfeiffer and M. A. Nowak, Curr. Biol. **16**, R946 (2006).
[34] M. Milinski, D. Semmann, H. J. Krambeck, and J. Marotzke, Proc. Natl. Acad. Sci. U.S.A. **103**, 3994 (2006).
[35] H. G. Schuster, *Complex Adaptive Systems* (Scator, Saarbrücken, 2002).
[36] A. Lotem, M. A. Fishman, and L. Stone, Nature (London) **400**, 226 (1999).
[37] E. Fehr and S. Gächter, Nature (London) **415**, 137 (2002).
[38] K. Panchanathan and R. Boyd, Nature (London) **432**, 499 (2004).
[39] O. Leimar and P. Hammerstein, Proc. R. Soc. London, Ser. B **268**, 745 (2001).
[40] K. Panchanathan and R. Boyd, J. Theor. Biol. **224**, 115 (2003).
[41] C. Wedekind and M. Milinski, Science **288**, 850 (2000).
[42] M. Milinski, D. Semmann, T. C. Bakker, and H. J. Krambeck, Proc. R. Soc. London, Ser. B **268**, 2495 (2001).
[43] M. Milinski, D. Semmann, and H. J. Krambeck, Nature (London) **415**, 424 (2002).

[44] P. D. Taylor and L. Jonker, Math. Biosci. **40**, 145 (1978).
[45] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).
[46] D. Helbing, Physica A **193**, 241 (1993).
[47] D. Helbing, Physica A **196**, 546 (1993).
[48] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. Lett. **95**, 238701 (2005).
[49] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. E **74**, 011901 (2006).
[50] M. A. Nowak and K. Sigmund, J. Theor. Biol. **194**, 561 (1998).
[51] D. Semmann, H. J. Krambeck, and M. Milinski, Nature (London) **425**, 390 (2003).
[52] L. A. Imhof, D. Fudenberg, and M. A. Nowak, Proc. Natl. Acad. Sci. U.S.A. **102**, 10797 (2005).
[53] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund, Science **316**, 1905 (2007).
[54] A. Traulsen and M. A. Nowak, PLOS One **2**, e270 (2007).
[55] G. Abramson and M. Kuperman, Phys. Rev. E **63**, 030901(R) (2001).
[56] P. Holme, A. Trusina, B. J. Kim, and P. Minnhagen, Phys. Rev. E **68**, 030901(R) (2003).
[57] H. Ebel and S. Bornholdt, Phys. Rev. E **66**, 056118 (2002).
[58] F. C. Santos, J. F. Rodrigues, and J. M. Pacheco, Proc. R. Soc. London, Ser. B **273**, 51 (2006).
[59] F. C. Santos and J. M. Pacheco, J. Evol. Biol. **19**, 726 (2006).
[60] F. C. Santos, J. M. Pacheco, and T. Lenaerts, Proc. Natl. Acad. Sci. U.S.A. **103**, 3490 (2006).
[61] H. Ohtsuki, M. A. Nowak, and J. M. Pacheco, Phys. Rev. Lett. **98**, 108106 (2007).
[62] H. Ohtsuki, J. Pacheco, and M. A. Nowak, J. Theor. Biol. **246**, 681 (2007).
[63] B. Skyrms and R. Pemantle, Proc. Natl. Acad. Sci. U.S.A. **97**, 9340 (2000).
[64] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[65] M. G. Zimmermann and V. M. Eguíluz, Phys. Rev. E **72**, 056118 (2005).
[66] P. Holme and G. Ghoshal, Phys. Rev. Lett. **96**, 098701 (2006).
[67] J. M. Pacheco, A. Traulsen, and M. A. Nowak, Phys. Rev. Lett. **97**, 258103 (2006).
[68] F. C. Santos, J. M. Pacheco, and T. Lenaerts, PLOS Comput. Biol. **2**, 1284 (2006).