

Mechanisms for similarity based cooperation

A. Traulsen^a

Program for Evolutionary Dynamics, Harvard University, Cambridge MA 02138, USA
and
Max-Planck-Institute for Evolutionary Biology, 24306 Plön, Germany

Received 14 August 2007

Published online 23 January 2008 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract. Cooperation based on similarity has been discussed since Richard Dawkins introduced the term “green beard” effect. In these models, individuals cooperate based on an arbitrary signal (or tag) such as the famous green beard. Here, two different models for such tag based cooperation are analysed. As neutral drift is important in both models, a finite population framework is applied. The first model, which we term “cooperative tags” considers a situation in which groups of cooperators are formed by some joint signal. Defectors adopting the signal and exploiting the group can lead to a breakdown of cooperation. In this case, conditions are derived under which the average abundance of the more cooperative strategy exceeds 50%. The second model considers a situation in which individuals start defecting towards others that are not similar to them. This situation is termed “defective tags”. It is shown that in this case, individuals using tags to cooperate exclusively with their own kind dominate over unconditional cooperators.

PACS. 02.50.Le Decision theory and game theory – 87.23.-n Ecology and evolution

Dawkins introduced the term “green beard” effect to describe an earlier thought experiment in sociobiology by Hamilton: a gene that leads to a tag visible to other individuals (such as the famous green beard) and also to the tendency to help others with the same tag can quickly spread and lead to cooperation based on these tags [1–4]. For a long time, it seemed to be difficult to accept the idea that a single gene leads to a tag and a change in behavior towards others with this tag. However, several experiments have reported green beard effects in biology [5–8]. If the tag and the change in behavior cannot be untangled, there is no room for cheaters: no individual can fake the signal which serves cooperators to identify each other.

Usually, green beards serve as an example how things cannot work. If tag and behavior evolve independently, this allows cheaters to undermine the system. Individuals with a certain tag that do not help others will enjoy the support of others, but not return this support. Hence, they will spread in the system and cooperation will diminish. Although this rules out the idea of long-term cooperation based on this mechanism, it has been shown that such mechanisms can lead to cooperation for a limited amount of time, depending on the costs and benefits of cooperation and the number of available tags. In this case cooperation evolves only under rather restrictive conditions [9,10].

In social systems, the gene centered view does not necessarily hold and thus, the results from these models can be quite different. In this case, a strategy can be defined as a combination of tag and behavior towards others. Such tags signal certain properties to others. For example, accents make it easy to recognize others with the same native language, which can lead to support due to communication. Fads are another example in which there is a cost associated with the admission into a group by buying a certain product. Every member of this group then supports the group identity.

In the presence of cheaters, tag-based cooperation can evolve, e.g. when the formation of groups is made possible by spatial structure which helps to establish cooperation [11]. Another possibility is the formation of cooperative tag groups in a mixed population for a limited amount of time [10,12]. It has also been proposed that such tag-based cooperation is an important mechanism in the formation and maintenance of peer-to-peer computer networks [13,14].

Here, different mechanisms that have been introduced to explain cooperation based on similarity are compared. In particular, the mechanisms introduced by Riolo et al. [15] and Jansen and van Baalen [16] are the basis for the comparison. To this end, we develop minimal models which allow an analytical comparison in a finite population framework. Finite populations are essential here, as

^a e-mail: traulsen@fas.harvard.edu

they allow neutral drift, which often has an important influence on the dynamics.

We consider only a finite number of tags, K , which can be viewed as the number of beard colours available. In the model motivated by the mechanism of Jansen and van Baalen [16], we have the strategies “always defect” D and “intragroup cooperation” I , which corresponds to the “ethnocentric” strategy of Axelrod et al. [11,17]. D individuals have a fixed tag, but do not cooperate. Individuals of type I cooperate only with others that have the same tag as themselves. We have hence $2K$ different types of individuals in total. Riolo et al. [15] introduced their model with tags and strategies given by real numbers. In this sense, the number of possible strategies is infinite. However, it has been shown that the general dynamics of this model can be captured by a simple toy model based on the replicator dynamics [18], which considers only two tags and two strategies [12]. Here, this model is generalised to K tags. For this model, the strategies are “always cooperate” C and “intragroup cooperation” I . Unconditional cooperators C cooperate with everyone, regardless of the tags. Again, the number of different individual types is $2K$. In the following, we use a lower index to denote the tag of an individual, e.g. D_i is a defector with tag i .

Population dynamics is modelled in finite populations, including the effects of mutation and random drift. We consider a well mixed population with no spatial structure, as it is well known that spatial structure itself can lead to cooperation [1,19–23]. We apply the pairwise comparison process discussed in [21,24–26], which for weak selection is equivalent to the frequency dependent Moran process [27–29]. The fixation probabilities of this process are identical to a Moran process with exponential mapping from payoffs to fitness [30]. Following [31,32], we assume that there is a small probability μ that an error occurs and an individual switches to a random strategy, which might not be present in the population yet. In general, the state space of the system is the discretized simplex $S_{2K} = \{(l_1, \dots, l_{2K}) \in N_0^{2K} : \sum_{k=1}^{2K} l_k = N\}$, where l_k is the number of individuals of type k and N is the population size. However, because μ is small, the system spends most time at the corners of the simplex. Occasionally, fixation of mutants leads from one corner to another one. The interior of the simplex is only reached when a second error occurs while the system moves along the edge of the simplex. Hence, the full Markov process can be approximated by the embedded Markov chain on the corners of the simplex if the typical time $\tau_\mu = 1/\mu$ between two mutations is much greater than the fixation time in which the system moves along the edge. For neutral selection, this time is $N(N-1)$. The frequency dependent selection in our system decreases this time [26,29]. Since the systems analysed here show no stable coexistence of two or more strategies, the maximum fixation time is found in the neutral case. Hence, we only have to ensure $\mu \ll N^{-2}$. The probability of switching from one strategy to another under frequency dependent selection is derived in the Appendix. The system has a symmetry between the different tags - exchanging the color of two beards cannot change

the dynamics of the system. Since we are interested only in the probability that the system is in one of the C , I , or D states (regardless of the tag), we only have to consider the transitions between strategies and can disregard transitions between tags without a change in the strategy. This reduces the number of types from $2K$ to 2 and allows computation of the stationary fraction of both strategies analytically for weak selection.

First, we analyze a model with cooperative tags. This yields a condition under which circumstances intragroup cooperators dominate over defectors in the population. Then, we turn to defective tags. We show that independent of the number of tags, the strategy using tags always dominates. Finally, the combination of all strategies is discussed.

1 Cooperative tags

Let us first consider a minimal model motivated by the mechanism discussed by Jansen and van Baalen [16]. In this model, individuals defect towards others with different tags. Cooperative individuals cooperate only with others with an identical tag. Since tags are used to establish cooperation in a population of defectors, this model is termed “cooperative tags”. There are only two strategies,

D Always defect.

I Cooperate with others that are from your own group.

Besides a strategy, each individual has one out of K different tags, denoted by a lower index. In the presence of I_i individuals, D_i individuals are advantageous over I_i . But if there are no D_i individuals, I_i individuals are advantageous compared to other types of defectors (D_j with $i \neq j$) as soon as there are at least two I_i individuals. This leads to a coupled dynamics of tags and strategies, see Figure 1a. Typically, a defecting population is replaced by a group with a different tag that cooperates. This group is later exploited by defectors of the same tag, leading to a defecting population with a different tag. These cycles can also be observed in simulations of unstructured populations, see Figure 2a.

The payoff matrix of a system with two different tags i and j is given by

$$\begin{array}{c} I_i \quad D_i \quad I_j \quad D_j \\ \begin{array}{l} I_i \\ D_i \\ I_j \\ D_j \end{array} \left(\begin{array}{cccc} b-c & -c & 0 & 0 \\ b & 0 & 0 & 0 \\ 0 & 0 & b-c & -c \\ 0 & 0 & b & 0 \end{array} \right). \end{array} \quad (1)$$

Here, b is the benefit from cooperation and $c < b$ is the cost for cooperation. The payoff matrix describes a Prisoner’s Dilemma in a population of one tag only. If several tags are available, the dilemma can be resolved, as new tags provide niches where cooperation can thrive until cooperators are exploited by defectors with the same tag again. In the following, we derive a condition which relates the benefits and costs to the number of available tags, stating under which conditions cooperation will dominate.

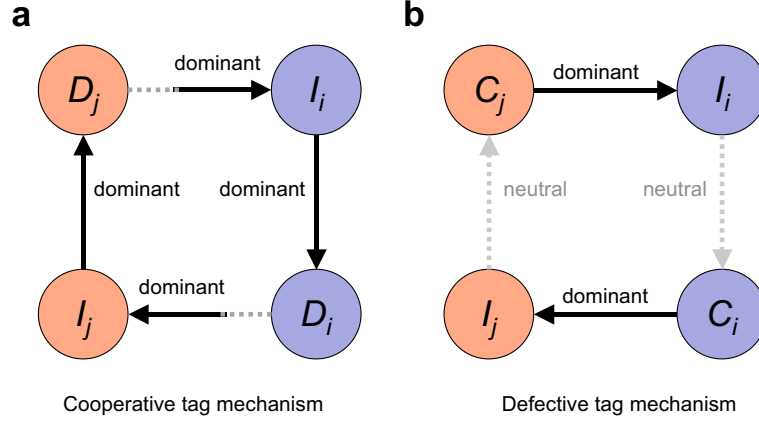


Fig. 1. Mechanisms for similarity based cooperation in finite populations. (a) In the cooperative tag mechanism, a defecting population with tag j can be invaded by intragroup cooperators of tag i . The first of these mutants is neutral. Then intragroup cooperation become advantageous. As soon as defecting individuals of the same tag appear, they can take over the population and the cycle continues. (b) Defective tag mechanism. A population of unconditional cooperators is invaded and taken over by intragroup cooperators of a different tag. These intragroup cooperators are supported by the whole population, but do not return this support. Neutral drift can transform these into unconditional cooperators again.

Let us first consider the transition from defectors to intragroup cooperators. A D_i population can be invaded by disadvantageous I_i individuals. Mutants of this type are produced at rate $\mu/(2K)$, as only one out of $2K$ possible types has a specific tag and a specific strategy. The payoff of l such I_i individuals is $\pi^I(l) = (b-c)(l-1)/(N-1) - c(N-l)/(N-1)$. The remaining $N-l$ D_i individuals have a higher payoff, $\pi^D(l) = bl/(N-1)$. For weak selection, where strategy update depends only weakly on the payoffs and is close to a random process, $\beta \ll 1$, the probability that a single I_i individual takes over a D_i population is (see Appendix)

$$\phi_{D_i \rightarrow I_i} = \frac{1}{N} - \frac{\beta}{2} \left(c + \frac{b-c}{N} \right). \quad (2)$$

The generalisation to larger β is straightforward, but does not lead to equations that are easy to interpret (see Appendix). A D_i population can also be invaded by an I_j individual ($i \neq j$). These individuals are advantageous as soon as there are two of them. They are produced at rate $\mu(K-1)/(2K)$, as there are $K-1$ available I_j types out of total number of $2K$. The payoff of l I_j invaders is given by $\pi^I(l) = (b-c)(l-1)/(N-1)$. Defecting residents have a vanishing payoff, $\pi^D(l) = 0$. Note that for $l=1$, both payoffs are identical. Thus, a single I_j individual in a population of D_i is neutral. A second individual of this kind can only be produced by random drift. But as soon as there are two individuals of type I_j , they are advantageous. A single I_j individual reaches fixation with probability

$$\phi_{D_i \rightarrow I_j} = \frac{1}{N} + \frac{\beta}{6}(b-c) \left(1 - \frac{2}{N} \right). \quad (3)$$

The combined rate of these processes, $\rho_{D \rightarrow I} = \mu \frac{1}{2K} \phi_{D_i \rightarrow I_i} + \mu \frac{K-1}{2K} \phi_{D_i \rightarrow I_j}$, is given by

$$\rho_{D \rightarrow I} = \frac{u}{2K} \left[\frac{K}{N} - \frac{\beta}{6} \left(3c + (b-c) \left(\frac{N+1}{N} - K \frac{N-2}{N} \right) \right) \right]. \quad (4)$$

Let us now calculate the rate of the opposite transitions from I to D . An I_i population can be exploited and invaded by advantageous D_i individuals. Mutants of this type are produced at rate $\mu/(2K)$. The payoffs are in the case of l D_i invaders $\pi^D(l) = b(N-l)/(N-1)$ and $\pi^I(l) = -cl/(N-1) + (b-c)(N-l-1)/(N-1)$. Hence, the fixation probability for $\beta \ll 1$ is

$$\phi_{I_i \rightarrow D_i} = \frac{1}{N} + \frac{\beta}{2} \left(c + \frac{b-c}{N} \right). \quad (5)$$

The I_i population can also be invaded by D_j individuals, which are disadvantageous as they cannot exploit the cooperating I_i individuals. Such individuals are produced at rate $\mu(K-1)/(2K)$. The payoffs are $\pi^I(l) = (b-c)(N-l-1)/(N-1)$ and $\pi^D(l) = 0$. The fixation probability is

$$\phi_{I_i \rightarrow D_j} = \frac{1}{N} - \frac{\beta}{3}(b-c) \left(1 - \frac{2}{N} \right). \quad (6)$$

The combined rate of the processes transforming an I population into a D population, $\rho_{I \rightarrow D} = \mu \frac{1}{2K} \phi_{I_i \rightarrow D_i} + \mu \frac{K-1}{2K} \phi_{I_i \rightarrow D_j}$ is given by

$$\rho_{I \rightarrow D} = \frac{\mu}{2K} \left[\frac{K}{N} + \frac{\beta}{6} \left(3c + (b-c) \left(\frac{2N-1}{N} - 2K \frac{N-2}{N} \right) \right) \right]. \quad (7)$$

In equilibrium, the probability of being in the I state, P_I , has to fulfill $P_I \rho_{I \rightarrow D} = (1-P_I) \rho_{D \rightarrow I}$. For weak selection,

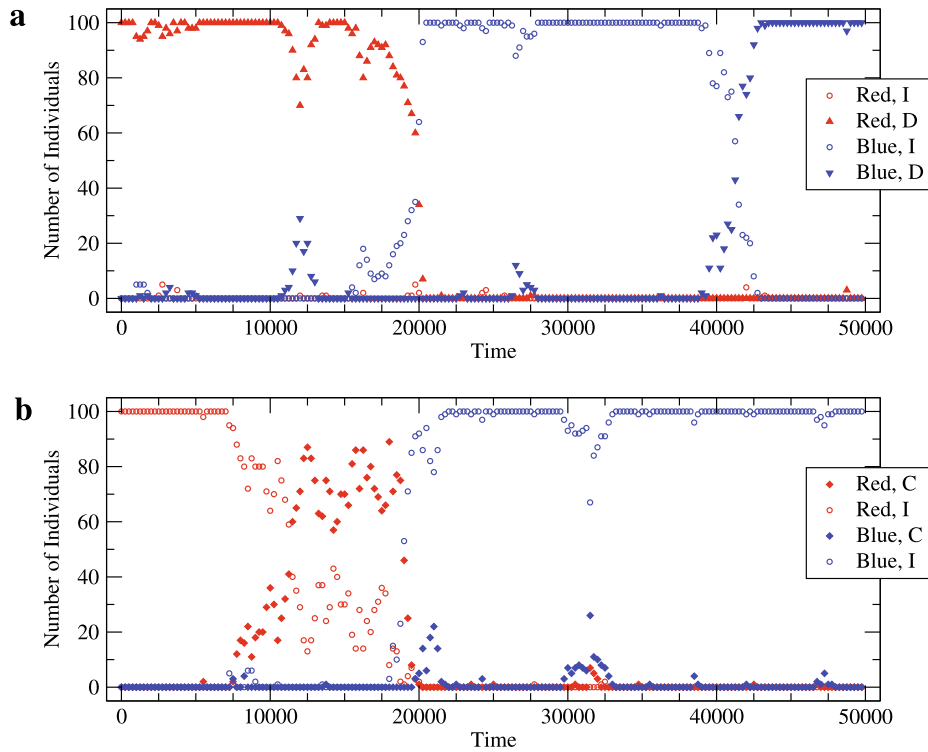


Fig. 2. Population dynamics for similarity based cooperation in finite populations. (a) Cooperative tag mechanism with two tags ($K = 2$), red and blue. A red defector population can be invaded by a blue intratag cooperator population, which builds a self-serving clique (between $T = 15\,000$ and $T = 20\,000$). This kind of cooperation is vulnerable to defectors of the same tag, which invade at $T \approx 40\,000$, destroying this kind of cooperation. Typically, tags change when the population is defecting. On average, the less cooperative strategy, always defect, dominates. (b) Defective tag mechanism. A group of red intratag cooperators is transformed to unconditional red cooperators by random drift ($T \approx 8\,000$). These are vulnerable to blue intratag cooperators, which invade at $T \approx 20\,000$. Typically, tags change when most individuals are unconditional cooperators. Although most individuals usually cooperate, the less cooperative strategy (intragroup cooperation) dominates (population size $N = 100$, intensity of selection $\beta = 0.5$, cost of cooperation $c = 0.5$, benefit from cooperation $b = 1.0$, mutation rate $\mu = 2 \times 10^{-3}$).

this equation results in

$$P_I \approx \frac{1}{2} + \frac{\beta}{8} \left((N-2)(b-c) - N(b+c) \frac{1}{K} \right). \quad (8)$$

This is in good agreement with simulations of the process. In Figure 3a, the dependence on the intensity of selection β is shown in comparison with simulations of the process, providing a generalization for strong selection. From equation (8), we can derive a simple condition under which intragroup cooperators dominate the population, i.e. have an average abundance above 50%. This condition is given by

$$K > \frac{N}{N-2} \frac{b+c}{b-c}. \quad (9)$$

Since N is usually large, the population size has a small influence only. If the cooperation game is very efficient, $b \gg c$, only a few tags are required for I individuals to dominate. However, if costs and benefit are very similar, $b \approx c$, many different tags are needed in order to reach dominance of I individuals. Alternatively, condition (9)

can be written as

$$\frac{b}{c} > \frac{K + \frac{N}{N-2}}{K - \frac{N}{N-2}} \approx \frac{K+1}{K-1}. \quad (10)$$

The approximation is valid for $N \gg 1$. For $K = 1$, this mechanism does not work, as no niches for cooperators are available. In the worst case of two tags, $K = 2$, (which only works for $N > 4$), the benefit to cost ratio for which intragroup cooperators still dominate is 3 for large populations. For large populations, b/c converges to 1 for $K \rightarrow \infty$, i.e. when enough tags are available, cooperation is always observed. This results from the fact that a cooperating population can only be exploited by a single type of mutants, while a defecting population can be invaded by $K - 1$ types of intragroup cooperators. In [10], this has been termed “altruistic freedom”. If mutants with the same tag are produced more frequently, this has to be reflected in equations (4) and (7) and the critical cost to benefit ratio will be different.

With increasing intensity of selection β , a complex dependence of the abundance of the two strategies can be

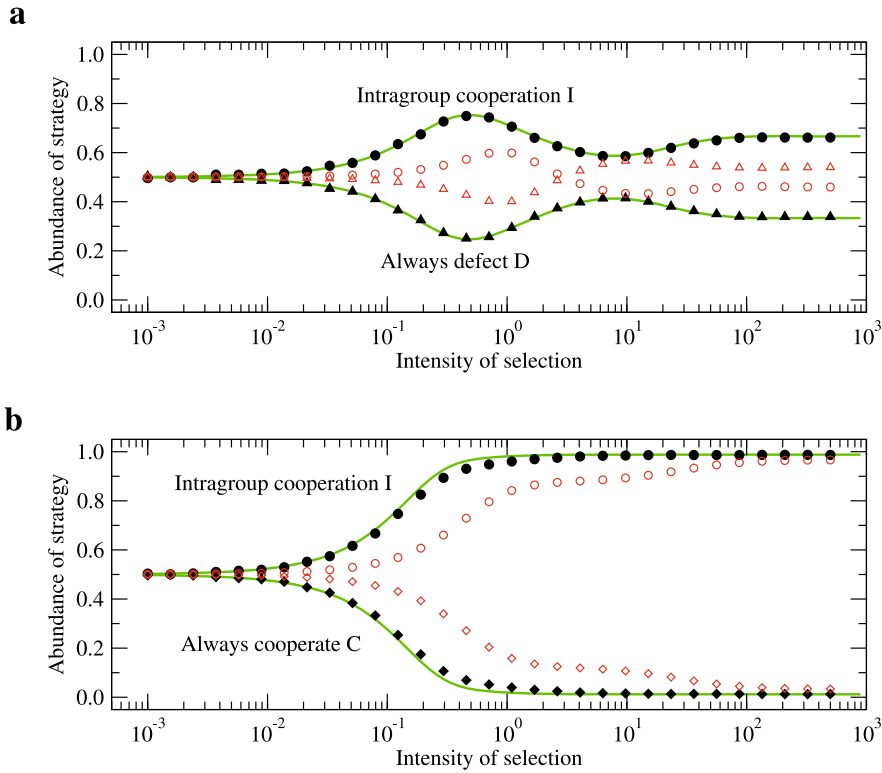


Fig. 3. The abundance of the different strategies in the two mechanisms can be computed from the transition probabilities between the two states alone. Here, the dependence on the intensity of selection β is shown. Symbols are computer simulations and lines give the theoretical result derived from equation (30). For small error probability ($\mu = 10^{-3}$, filled symbols), simulations and the theoretical result agree very well. The weak selection approximation is valid if the intensity of selection is below 0.1, where it basically coincides with the lines. (a) Cooperative tag mechanism with five tags. Since the chosen parameters fulfil the condition (9), intragroup cooperators I dominate for small intensity of selection, $\beta \ll 1$. The maximum abundance of strategy I is observed at $\beta \approx 0.5$. With increasing intensity of selection, a complex change in the stationary distribution is observed. When errors are more frequent ($\mu = 0.1$, open symbols), the theoretical prediction becomes misleading for high intensity of selection, as always defect dominates here. (b) Defective tag mechanism with $K = 5$ tags. For all intensities of selection β , the less cooperative intragroup cooperators I dominate the population. For $\mu = 0.1$ (open symbols), the theoretical result deviates from the simulations, but still gives a qualitative estimate (Parameters: population size $N = 20$, $K = 5$, cost of cooperation $c = 0.2$, benefit from cooperation $b = 1.0$, averages over 10^9 time steps).

observed. The result for weak selection can be misleading, as the picture can be different for strong selection. For example, in a system with the parameters $N = 20$, $K = 5$, $c = 0.2$, and $b = 1.0$, condition (9) predicts that strategy I dominates the system. With increasing β , the abundance of the I strategies, x_I , increases, as expected. However, x_I reaches a maximum for $\beta \approx 0.5$ at $x_I \approx 0.75$, see Figure 3a. Then, it decreases until $\beta \approx 8$, where the abundance of strategy (I) reaches $x_I \approx 0.59$. Finally, it increases again until it reaches a constant value $x_I = 0.67$ for $\beta \rightarrow \infty$. This complex behavior arises from the interplay of the four frequency dependent transitions that enter into the stationary distribution.

We can also introduce a parameter s which measures the ratio of two mutation rates. At rate μ , only the strategy changes. At rate μs , tags and strategies change simultaneously. For $s \ll 1$, tag and strategy mutations are

independent. For $s \gg 1$, they are strongly coupled. Incorporating this parameter, equation (10) changes to

$$\frac{b}{c} > \frac{(K-1)s + \frac{2N-2}{N-2}}{(K-1)s - \frac{2}{N-2}}. \quad (11)$$

given that $(K-1)s - 2/(N-2) > 0$. Otherwise, the opposite relation is valid. For $s = 1$, we recover equation (10). For $s \gg 1$, this condition reduces to $b > c$ and is always fulfilled. The situation corresponds to a single gene greenbeard effect. For $s \rightarrow 0$, the relevant condition is $b/c < 1 - N < 0$, which is never fulfilled. The situation corresponds to a greenbeard effect caused by two independent genes, which would not lead to cooperation.

2 Defective tags

In the model of Riolo et al., individuals always cooperate with others that are sufficiently similar to them [15]. Since they have no possibility of denying help if others similar to themselves are present, real defectors that never cooperate with anyone are absent [33]. The driving force behind the population dynamics can be easily understood if a minimalist model is considered [12], see. Figure 1b. In contrast to [12], here a finite population with more than two tags is analyzed. We term the model “defective tags”, as a tag is not used to establish cooperation in a population of defectors, but rather to found a self-serving group of cooperators that do not help others. As in the analysis above, we consider K different tags and two strategies. The two strategies are

C Always cooperate

I Cooperate with others that are from your own group

As for the cooperative tag model, it is instructive to write down a payoff matrix for the interaction of the two strategies with two different tags,

$$\begin{array}{c} \begin{array}{cccc} & C_i & I_i & C_j & I_j \\ C_i & \begin{pmatrix} b-c & b-c & b-c & -c \\ b-c & b-c & b & 0 \\ b-c & -c & b-c & b-c \\ b & 0 & b-c & b-c \end{pmatrix} \\ I_i & \\ C_j & \\ I_j & \end{array} \end{array} \quad (12)$$

For K tags, the payoff matrix has dimension $2K \times 2K$. Again, we consider population dynamics in a finite population as described in the Appendix. Typically, a population of unconditional cooperators is invaded by intragroup cooperators of a different tag. These are then again transformed to unconditional cooperators, see Figure 2b. In an infinite population, this process has been modeled by an external drift from intolerant to tolerant strategies [12]. In finite populations, such an external extension of the selection dynamics is no longer necessary, as random drift is naturally incorporated in this process, see Figures 1b and 2b. The symmetry between tags allows us to compute the stationary fraction of both strategies by taking into account only transitions that change the strategy, but leave the tag unchanged. First, we compute the rate in which the population is transformed from unconditional cooperation to intragroup cooperation. A C_i population can be replaced by an I_i population. Such a transition is neutral and a single mutant fixates with probability $1/N$. Mutants of this type are produced at rate $\mu/(2K)$, leading to a transition rate $\phi_{C_i \rightarrow I_i} = \mu/(2KN)$. A C_i population can also be invaded by an I_j individual, which is advantageous in this case. Individuals of such type appear at rate $\mu(K-1)/(2K)$. If there are l individuals of this type, their payoff is $\pi^I(l) = (b-c)(l-1)/(N-1) + b(N-l)/(N-1)$. The remaining $N-l$ C_i individuals have a smaller payoff, $\pi^C(l) = -cl/(N-1) + (b-c)(N-l-1)/(N-1)$. Hence, the probability that a single I_j individual takes over a C_i population of a different tag is

$$\phi_{C_i \rightarrow I_j} = \frac{1}{N} + \frac{\beta}{6} \left(b + 2c + \frac{b-c}{N} \right). \quad (13)$$

The combined rate of the processes leading from C to I , $\rho_{C \rightarrow I} = \mu \frac{1}{2K} \phi_{C_i \rightarrow I_i} + \mu \frac{K-1}{2K} \phi_{C_i \rightarrow I_j}$, is given by

$$\rho_{C \rightarrow I} = \frac{\mu}{2K} \left[\frac{K}{N} + \frac{\beta}{6}(K-1) \left(b + 2c + \frac{b-c}{N} \right) \right]. \quad (14)$$

Now we turn to the processes in which the population switches from I to C . First, an I_i population can be replaced by a C_i population. This transition is neutral and a single mutant fixates with probability $1/N$. Mutants of this type are produced at rate $\mu/(2K)$, leading to a transition rate $\phi_{I_i \rightarrow C_i} = \mu/(2KN)$. Second, the I_i population can also be invaded by a C_j individual, which is disadvantageous. Individuals of such type appear at rate $\mu(K-1)/(2K)$. If there are l C_j individuals, their payoff is $\pi^C(l) = (b-c)(l-1)/(N-1) - c(N-l)/(N-1)$. The remaining $N-l$ I_i individuals have payoff $\pi^I(l) = bl/(N-1) + (b-c)(N-l-1)/(N-1)$ each. Hence, the probability that a single C_j individual takes over an I_i population is

$$\phi_{I_i \rightarrow C_j} = \frac{1}{N} - \frac{\beta}{6} \left(2b + c + \frac{b-c}{N} \right). \quad (15)$$

The combined rate of the two processes is given by

$$\rho_{I \rightarrow C} = \frac{\mu}{2K} \left[\frac{K}{N} - \frac{\beta}{6}(K-1) \left(2b + c + \frac{b-c}{N} \right) \right]. \quad (16)$$

From the transition rates, we can again compute the equilibrium distribution of the system. For the probability that the system is in the C state, we have $P_C \rho_{C \rightarrow I} = (1 - P_C) \rho_{I \rightarrow C}$. Finally, we find for weak selection

$$P_C \approx \frac{1}{2} - \beta \frac{K-1}{K} \left(\frac{b-c}{12} + N \frac{b+c}{8} \right). \quad (17)$$

We have $P_C < 1/2$, i.e. unconditional cooperators never dominate the system and the system spends most time in I states. This holds also for stronger selection, when nonlinear terms in β are taken into account: see Figure 3b. The explanation for this is simple: The $C \rightarrow I$ transition is advantageous, whereas the $I \rightarrow C$ transition is disadvantageous. The more tags are present, the smaller the number of cooperators in equilibrium. However, this effect is small and does not lead to any qualitative differences, in contrast to the cooperative tag model (see above).

If tag and strategy are controlled by a single gene and cannot change independently, the up-down transitions in Figure 1b that drive the cyclic dominance are not possible and the dynamics of the system will be different. If the simultaneous change of tag and strategy occurs with a very small probability, the left-right transitions in Figure 1b occur with a very small probability, which changes the cyclic dominance. There is no qualitative change incorporating different mutation rates modeled by a parameter s here. The only possible way to increase the fraction of cooperators is to weaken the transition from C to I , i.e. reduce s . However, in this case neutral drift between the

two strategies dominates and the fraction of cooperators remains below 50%.

In the cooperative tag model, a higher number of tags increases the abundance of the more cooperative strategy for weak selection. In contrast, in the defective tag model, a higher number of tags decreases the abundance of unconditional cooperators.

3 Combination of all three strategies

We can also combine all three strategies, unconditional cooperators C , intragroup cooperators I and unconditional defectors D . In this case, the 6×6 payoff matrix for two different tags is given by

$$\begin{array}{c} C_i \\ I_i \\ D_i \\ C_j \\ I_j \\ D_j \end{array} \begin{array}{cccccc} C_i & I_i & D_i & C_j & I_j & D_j \\ \left(\begin{array}{cccccc} b-c & b-c & -c & b-c & -c & -c \\ b-c & b-c & -c & b & 0 & 0 \\ b & b & 0 & b & 0 & 0 \\ b-c & -c & -c & b-c & b-c & -c \\ b & 0 & 0 & b-c & b-c & -c \\ b & 0 & 0 & b & b & 0 \end{array} \right). \quad (18)$$

Now we have $3K$ different types in total. Again, we can calculate the transition probabilities between the different states. In addition, we have to calculate the transition rates from C to D and vice versa, which can be done using the calculation in the Appendix. Primed transition rates ρ' indicate that we are considering the $3K$ -strategy case. In analogy to the calculations above, we find

$$\rho'_{C \rightarrow D} = \frac{u}{3K} \left[\frac{1}{N} + \frac{\beta}{2} \left(c + \frac{b-c}{N} \right) \right] \quad (19)$$

$$\rho'_{D \rightarrow C} = \frac{u}{3K} \left[\frac{1}{N} - \frac{\beta}{2} \left(c + \frac{b-c}{N} \right) \right] \quad (20)$$

$$\rho'_{D \rightarrow I} = \frac{u}{3K} \left[\frac{K}{N} - \frac{\beta}{6} \left(3c + (b-c) \left(\frac{N+1}{N} - K \frac{N-2}{N} \right) \right) \right] \quad (21)$$

$$\rho'_{I \rightarrow D} = \frac{\mu}{3K} \left[\frac{K}{N} + \frac{\beta}{6} \left(3c + (b-c) \left(\frac{2N-1}{N} - 2K \frac{N-2}{N} \right) \right) \right] \quad (22)$$

$$\rho'_{C \rightarrow I} = \frac{\mu}{3K} \left[\frac{K}{N} + \frac{\beta}{6} (K-1) \left(b + 2c + \frac{b-c}{N} \right) \right] \quad (23)$$

$$\rho'_{I \rightarrow C} = \frac{\mu}{3K} \left[\frac{K}{N} - \frac{\beta}{6} (K-1) \left(2b + c + \frac{b-c}{N} \right) \right]. \quad (24)$$

From these transition rates, we can construct a transition matrix between the three different strategies, see

e.g. [31,32]. The largest eigenvector (which has eigenvalue 1, as the transition matrix is stochastic) gives the stationary distribution. For weak selection and large populations ($\beta \ll 1$, $N \gg 1$ with $\beta N \ll 1$), we find

$$\begin{pmatrix} P'_C \\ P'_I \\ P'_D \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} + \frac{N\beta}{18} \begin{pmatrix} -b - 3c + \frac{b+c}{K} \\ +2b - 2\frac{b+c}{K} \\ -b + 3c + \frac{b+c}{K} \end{pmatrix}. \quad (25)$$

For a large number of tags, $K \gg 1$, these equations are easy to interpret. The abundance of unconditional cooperators C is always smaller than $1/3$, which is the neutral case. Intragroup cooperators, I , are present more than in the neutral case. Moreover, since $2b > -b+3c$ due to $b > c$, there are always more intragroup cooperators than defectors. For unconditional defectors D , the benefit to cost ratio is important: for $b/c < 3$, their abundance exceeds $1/3$, for $b/c > 3$ it is below $1/3$. Of course, for finite K (and smaller N) the conditions become more intricate.

Thus, the intragroup cooperators are successful under weak selection even if both unconditional defectors and unconditional cooperators are present. The combination of the three strategies has also been analyzed by Masuda and Ohtsuki in a setting where tags are not always observable [34].

4 Discussion

In structured populations, i.e. on spatial lattices or social networks, the dynamics of tag based system can be quite different. In all cases, strong segregation between different tags is observed [11,16,35,36].

The analytical results derived here are based on weak selection. The generalisation to higher intensities of selection is possible with the methods given in the Appendix, but does not lead to equations that are easy to interpret. Throughout this paper, it has been assumed that errors always lead to a random strategy. The picture changes if we assume that errors only change the tag or the strategy of an individual. In this case, for the cooperative tag mechanism the transition from D to I would require two independent errors and the system would spend more time in the D state. In the defective tag model, the system spends more time in the C state, as the transition to I requires two independent errors. However, alterations in the transition rates of the model can easily be done based on the framework introduced here.

Finally, which mechanism for similarity based cooperation is more likely to occur in reality? The defective tag model describes a mechanism in which tags indicate the action of an individual, i.e. I can predict the behavior of an individual based on its tag. Two individuals with the same tag will always cooperate with each other. This framework seems to be correct if the tag and the behaviour are tightly linked and cheating is not possible [6,7]. Another possibility for excluding cheaters that fake a certain tag are fads

in which the tag itself is costly, e.g. ownership of a certain car or MP3-player.

Whenever cheating is possible, the cooperative tag framework is more appropriate. In this case, unconditional cooperators can often be excluded from the beginning, as they are dominated by intragroup cooperators. Even in the presence of cheaters, cooperation based on tags can evolve for a limited amount of time, until the first cheater leads to a decay of cooperation again. Since cheaters do not enjoy the benefits of cooperation, new cooperative groups will form again and again [37].

When all three strategies are combined, then intragroup cooperators dominate under weak selection. For sufficiently high benefit to cost ratios, the presence of this strategy can even lead to a situation in which there are fewer defectors than in the neutral case.

The mechanisms for tag based cooperation discussed here show oscillations in the cooperative behavior, as cooperation based on similarity is unstable and can always be exploited. These “Tides of Tolerance” [38] are a characteristic feature of similarity based cooperation.

I thank Martin A. Nowak, Anna Dreber and Norman Siebrasse for helpful discussions. Support by the “Deutsche Akademie der Naturforscher Leopoldina” (Grant No. BMBF-LPD 9901/8-134) is gratefully acknowledged. The Program for Evolutionary Dynamics at Harvard is sponsored by J. Epstein.

Appendix A: Finite population framework

The payoffs are described above. Individuals adopt new strategies employing the mechanism described in [25]. A randomly chosen focal individual with payoff π^f compares its payoff to a second random individual with payoff π^c . The focal individual switches its strategy to the one of this role model with probability

$$p = \frac{1}{1 + e^{\beta(\pi^f - \pi^c)}}. \quad (26)$$

The intensity of selection is given by β . For $\beta \gg 1$, selection is strong and only more successful strategies are adopted. For $\beta \ll 1$, selection is weak. In this case, the process coincides with the frequency dependent Moran process under weak selection [27]. With a small probability $\mu \ll 1$, an individual tries an entirely new strategy which is possibly not present in the population yet. Since these errors are rare, there is usually only one strategy present in the population (the error rate has to be much smaller than the inverse fixation time of a newly introduced strategy). Only occasionally, an error occurs and the newly introduced strategy can take over the population. Consider a situation with l individuals of type A and $N - l$ individuals of type B . The payoff of the A individuals is

$$\pi^A(l) = a \frac{l-1}{N-1} + b \frac{N-l}{N-1} \quad (27)$$

and equivalently

$$\pi^B(l) = c \frac{l}{N-1} + d \frac{N-l-1}{N-1}. \quad (28)$$

The probability that the number of A individuals changes from i to $l \pm 1$ is

$$T^\pm(l) = \frac{l}{N} \frac{N-l}{N} \frac{1}{1 + e^{\pm\beta(\pi^A - \pi^B)}}. \quad (29)$$

The probability that a single individual of type A takes over a population of type B is

$$\phi_{B \rightarrow A} = \left[1 + \sum_{m=1}^{N-1} \prod_{l=1}^m \frac{T^-(l)}{T^+(l)} \right]^{-1}. \quad (30)$$

For large β , the fixation probabilities can be computed from this formula or from a closed expression that is an excellent approximation for this formula [25,26]. However, here we concentrate on weak selection, $\beta \ll 1$, where the fixation probabilities reduce to

$$\phi_{B \rightarrow A} = \frac{1}{N} + \frac{\beta}{6} \left(a + 2b - c - 2d - \frac{2a + b + c - 4d}{N} \right). \quad (31)$$

This result is identical to the corresponding result of the frequency dependent Moran process [27]. The results for stronger selection shown in Figure 3 and discussed in the text are derived from equation (30) without this simplification.

References

1. W.D. Hamilton, *J. Theor. Biol.* **7**, 1 (1964)
2. W.D. Hamilton, *J. Theor. Biol.* **7**, 17 (1964)
3. R. Dawkins, *The Selfish Gene* (Oxford University Press, New York, 1976)
4. R. Dawkins, *The extended phenotype* (Oxford University Press, Oxford, 1982)
5. D. Haig, *Proc. Natl. Acad. Sci. USA* **93**, 6547 (1996)
6. L. Keller, K.G. Ross, *Nature* **394**, 573 (1998)
7. D.C. Queller, E. Ponte, S. Bozzaro, J.E. Strassmann, *Science* **299**, 105 (2003)
8. K. Summers, B. Crespi, *Proc. Roy. Soc. Lond. B* **272**, 643 (2005)
9. G. Hartvigsen, L. Worden, S. Levin, *Complexity* **5**, 14 (2000)
10. A. Traulsen, M.A. Nowak, *PLoS One* **2**, e270 (2007)
11. R. Axelrod, R.A. Hammond, A. Grafen, *Evolution* **58**, 1833 (2004)
12. A. Traulsen, H.G. Schuster, *Phys. Rev. E* **68**, 046129 (2003)
13. D. Hales, B. Edmonds, *IEEE Transactions on systems, man, and cybernetics – Part A: systems and humans* **35**, 385 (2005)
14. T. Pfeiffer, M.A. Nowak, *Curr. Biol.* **16**, R946 (2006)
15. R.L. Riolo, M.D. Cohen, R. Axelrod, *Nature* **414**, 441 (2001)

16. V.A.A. Jansen, M. Van Baalen, *Nature* **440**, 663 (2006)
17. R. Axelrod, R.A. Hammond, Proc. Midwest Political Science Convention (2003)
18. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, 1998)
19. M.A. Nowak, R.M. May, *Nature* **359**, 826 (1992)
20. K. Lindgren, M.G. Nordahl, *Physica D* **75**, 292 (1994)
21. G. Szabó, C. Tóke, *Phys. Rev. E* **58**, 69 (1998)
22. F. Schweitzer, L. Behera, H. Mühlenbein, *Advances in Complex Systems* **5**, 269 (2002)
23. H. Ohtsuki, C. Hauert, E. Lieberman, M.A. Nowak, *Nature* **441**, 502 (2006)
24. L.E. Blume, *Games and Economic Behavior* **5**, 387 (1993)
25. A. Traulsen, M.A. Nowak, J.M. Pacheco, *Phys. Rev. E* **74**, 11909 (2006)
26. A. Traulsen, J.M. Pacheco, M.A. Nowak, *J. Theor. Biol.* **246**, 522 (2007)
27. M.A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, *Nature* **428**, 646 (2004)
28. C. Taylor, D. Fudenberg, A. Sasaki, M.A. Nowak, *Bull. Math. Biol.* **66**, 1621 (2004)
29. T. Antal, I. Scheuring, *Bull. Math. Biol.* **68**, 1923 (2006)
30. A. Traulsen, N. Shresh, M.A. Nowak, *Bull. Math. Biol.* (2007), submitted
31. L.A. Imhof, D. Fudenberg, M.A. Nowak, *Proc. Natl. Acad. Sci. USA* **102**, 10797 (2005)
32. C. Hauert, A. Traulsen, H. Brandt, M.A. Nowak, K. Sigmund, *Science* **316**, 1905 (2007)
33. G. Roberts, T.N. Sherratt, *Nature* **418**, 499 (2002)
34. N. Masuda, H. Ohtsuki, *Proc. Roy. Soc. Lond. B* **274**, 689 (2007)
35. A. Traulsen, J.C. Claussen, *Phys. Rev. E* **70**, 046128 (2004)
36. R.A. Hammond, R. Axelrod, *Theor. Pop. Biol.* **69**, 333 (2006)
37. D. Hales, *Change your tags fast! – A necessary condition for cooperation?*, in *Proceedings of the Workshop on Multi-Agents and Multi-Agent-Based Simulation* (2004), Vol. LNAI 3415, pp. 89–98
38. K. Sigmund, M.A. Nowak, *Nature* **414**, 403 (2001)