



Shade T. Shuttters and David Hales (2013)

Tag-Mediated Altruism is Contingent on How Cheaters Are Defined

Journal of Artificial Societies and Social Simulation 16 (1) 4<<http://jasss.soc.surrey.ac.uk/16/1/4.html>>

Received: 16-Feb-2012 Accepted: 28-Aug-2012 Published: 31-Jan-2013



Abstract

Cooperation is essential for complex biological and social systems and explaining its evolutionary origins remains a central question in several disciplines. Tag systems are a class of models demonstrating the evolution of cooperation between selfish replicators. A number of previous models have been presented but they have not been widely explored. Though previous researchers have concentrated on the effects of one or several parameters of tag models, exploring exactly what is meant by cheating in a tag system has received little attention. Here we re-implement three previous models of tag-mediated altruism and introduce four definitions of cheaters. Previous models have used what we consider weaker versions of cheaters that may not exploit cooperators to the degree possible, or to the degree observed in natural systems. We find that the level of altruism that evolves in a population is highly contingent on how cheaters are defined. In particular when cheaters are defined as agents that display an appropriate tag but have no mechanism for participating in altruistic acts themselves, a population is quickly invaded by cheaters and all altruism collapses. Even in the intermediate case where cheaters may revert back to a tag-tolerance mode of interaction, only minimal levels of altruism evolve. Our results suggest that models of tag-mediated altruism using stronger types of cheaters may require additional mechanisms, such as punishment strategies or multi-level selection, to evolve meaningful levels of altruism.

Keywords:

Cooperation, Evolution, Green Beard, Social Parasitism, Chromodynamics



Introduction and Background

- 1.1 Understanding the conditions under which cooperation emerges within evolving populations has been studied within biological, social, and computational sciences. While motivating questions and applications vary by discipline, a number of general mechanisms and models have been identified which potentially have broad applicability. For example, direct reciprocity, where a pair of agents help each other (Trivers 1971), can support high-levels of cooperation when future interactions are likely and agents can recognize and punish those who have not cooperated in the past (Axelrod 1984). Likewise, kin selection, or inclusive fitness theory, asserts that cooperation can thrive among interacting individuals that exhibit a large degree of genetic similarity (Hamilton 1964). Indirect reciprocity, where agents help those who have helped others, also promotes cooperation when information concerning past behavior (reputation) is available (Nowak & Sigmund 1998; Nowak & Sigmund 2005). In addition, placing agents within various spatial and network structures can support high, and often dynamic, levels of cooperation (Cohen, Riolo, & Axelrod 2001; Hui, Xu, & Zheng 2007; Konno 2010; Nowak & May 1992; Santos & Pacheco 2006).
- 1.2 There is a large and growing body of work exploring these general cooperation-promoting mechanisms in relation to the necessary and sufficient conditions required for them to be effective (Nowak 2006). However, less explored in this respect is the tag mechanism. The tag approach was presented by Holland(1993), discussed by Axelrod(1984), who termed them "labels"^[1], and implemented in a number of variants (Cohen, Riolo, & Axelrod 1999; Hales 2000; Riolo, Cohen, & Axelrod 2001).
- 1.3 Tags are features attached to agents that can be observed by other agents for the purpose of deciding whether to cooperate. During interactions agents take into account their own tag, their partner's tag, and a behavioral trait or strategy. The combination of these three determines if an agent will cooperate. During evolution both the tags and behavioral traits evolve through replication and mutation. What is striking about the tag approach is that it sustains high levels of cooperation between very simple agents with no requirement that agents have knowledge of past interactions.

Related work

- 1.4 Holland (1993) was among the first to discuss the notion of tags in the context of classifier systems, a way of representing agent decision-making via an evolving set of interacting rules. He argued that externally visible tags provide a powerful mechanism for dynamically structuring populations, giving examples from both biology and human society. The important point is that an initially arbitrary tag, through evolution, can come to have a meaning by being a predictor of agent behavior. This is achieved by agents biasing their interactions towards others with the same tag. He discussed their application in the iterated prisoners dilemma (IPD) game indicating how agents can increase cooperation by supporting the spread of the tit-for-tat strategy^[2].
- 1.5 Riolo (1997) operationalized the important points from Holland's work by taking agents playing the IPD and giving each a tag represented as a real number on [0, 1]. These evolving agents sampled the population randomly for game partners and played a game with a probability proportional to tag similarity, meaning agents with different tags still had a chance to interact. Riolo included a cost associated with searching the population and found that reciprocal strategies and cooperation could emerge if the search cost was sufficiently low and the number of IPD rounds was sufficiently high.
- 1.6 Though Riolo's model showed the effects of tags on an IPD, it did not consider their effect when games are one-shot, a phenomenon growing more prevalent in human societies. To address this concern Hales (2000) presented a model applied to the single round prisoners dilemma with no cost for searching the population. Agent strategies were represented as a binary trait indicating a strategy of cooperation or defection while tags were represented as fixed length bitstrings. Games only took place when tags matched exactly. If no exact match could be found in the population, a game was played with a randomly chosen agent regardless of its tag. Results showed that even when interactions were not iterated, high levels of cooperation were still possible if tag length was sufficiently large.
- 1.7 In all of these previous cases, specific model parameters were tested and analyzed, primarily using agent-based models. Recently, a general analytical model of tag donations was developed by Traulsen and Nowak(2007). The authors present an evolutionary model applied to general cost and benefit conditions of cooperation in a donation game, demonstrating analytically the conditions under which cooperation emerges, especially the conditions related to the cost-benefit ratio and the number of tags. In their model, tags are discrete and interactions can only take place between agents sharing the same identical tag. They prove mathematically that, given a sufficiently large number of tags, cooperation will be selected when benefit is greater than cost.

Paper motivation and structure

- 1.8 The primary motivation for the current study begins with one of the most widely cited tag models, that of Riolo, Cohen, and Axelrod(2001). Their elegant model (hereafter RCA) consisted of a simple unilateral donation game in which a donation incurs a cost to the donor and a benefit to the recipient. Tags are represented as real numbers on [0, 1]. Strategies are represented as a tolerance value on [0, 1] and dictate how similar tags must be for a donation to occur. The model is defined such that a donation *must* occur between agents with identical tags. There is no cost for searching the population, which is implemented through repeated random sampling (called pairings). The surprising result from the RCA model was that high levels of donations evolved despite evolutionary pressure favoring agents that did not donate.
- 1.9 Roberts and Sherratt (2002) modified the RCA model by removing the requirement that agents must donate to another with an identical tag. Instead, their model (hereafter RS) allowed agents to ignore such partners, a change that almost completely eliminated altruism in the original RCA model.

- 1.10 Continuing to explore the nature of the RCA model, Spector and Klein(2006, hereafter SK) modified RS by restricting the potential pool of partners, or interaction neighborhood, to a subset of the total population. This change, which effectively added social structure to the original model, showed that high donation rates could be recovered if an agent's neighborhood was very small compared to the total population.
- 1.11 In this paper we revisit the RCA model and the two subsequent variants, RS and SK, to explore the conditions that lead to high levels of altruism. We examine a number of existing parameters with a particular interest in how cheaters are defined. We introduce the idea of weak, medium, and strong cheaters (i.e. agents that derive benefits without incurring a cost) and show that the way a cheater is defined can have a dramatic effect on levels of altruism. This is particularly relevant when relating the results of tag models to actual biological and social phenomena or when attempting to apply such mechanisms to engineer distributed computer systems.
- 1.12 In the following section we describe the RCA, RS and SK models in detail along with their principle findings (see Table 1 for primary differences between each). We then introduce a classification scheme for cheater types followed by a description of a simulation model that explores the effects of those types. Next we present and discuss the results obtained from the model, with comparisons to results from previous models, and finally conclude with potential future directions for research in this area.



The RCA tag model

- 2.1 The RCA model consists of a population of 100 agents. Each agent comprises two real valued numbers: a tag $t \in [0, 1]$ and a behavioral trait called a "tolerance threshold" $T \in [0, \infty)$. In each generation each agent is paired with $P = 3$ agents from the population at random (with replacement). During each pairing, or interaction, the initiating agent A determines whether to make a donation to a recipient agent B. A donates to B if their tags are sufficiently similar. Agent A determines this by comparing its own tag (t_A) with the tag of agent B (t_B). If the difference between the tags is less than or equal to A's tolerance threshold of T_A then a donation takes place from A to B. Hence a donation occurs when $|t_A - t_B| \leq T_A$. A donation causes B's payoff to increase by the benefit of $b = 1$ and A's payoff to decrease by the cost of $c = 0.1$. Donation therefore represents altruistic behavior because A incurs a cost to help B. An agent with a high T will donate to others holding a wide range of tag values but an agent with a low T will only donate to those with very similar tag values. In the extreme cases an agent with $T \geq 1$ will donate to all other agents whereas an agent with $T = 0$ will only donate to those with tag values identical to its own.
- 2.2 After each generation of 300 pairings (100 agents $\times$ 3 pairings) the payoff of each agent, its accumulated costs and benefits, is considered as the fitness of the agent. Evolution is then implemented in the form of a tournament selection process in which each agent A compares its own fitness f to a randomly chosen other B. If $f_B > f_A$, then A copies B's traits - that is, both tag and tolerance values. But if $f_B < f_A$, then A's strategy remains unchanged. If $f_B = f_A$ a coin toss is used to determine whether the strategy of A or B prevails³¹. This selection process may be interpreted as not only a biological evolutionary process but also a mechanism of learning in fixed populations (Riolo et al. 2001).
- 2.3 Mutation is applied to newly copied values with probability 0.1. Mutation of the tag involves its replacement with a new random tag uniformly chosen from $[0, 1]$ while mutation of the tolerance T is achieved by adding Gaussian noise to the old value with mean 0, standard deviation 0.01. If mutation results in $T < 0$, T is reset to 0.
- 2.4 The model was explored via computer simulations. At the beginning of a simulation run all tags and tolerances were initialized to random values selected uniformly from $[0, 1]$. Each simulation run was executed for 30,000 generations. In the original published results a number of runs were performed exploring different values of both the pairing P and the cost c parameters. Average donation rates and tolerances were reported over a range of P and c values.
- 2.5 Results showed high average donation rates were obtained in conjunction with low tolerance values. This was significant because it indicated high levels of altruism without reciprocity in randomly mixing populations.



Reaction to the RCA model - the RS and SK models

- 3.1 In addition to problems with the RCA model pointed out by Edmonds and Hales(2003), Roberts and Sherratt(2002) took issue with the way cheaters were defined in RCA. In particular, RCA did not allow cheating in the case that two agents had identical tags. Even though agent A may have a tolerance of 0, if it interacted with agent B who had an identical tag, A would still altruistically donate to B. Roberts and Sherratt implemented their own version (RS) of the RCA model, in which the tolerance of an agent could evolve to a value less than 0. They implemented this by changing the tolerance range from $T \in [0, \infty)$ to $T \in [-10^{-6}, \infty)$. Though the RCA and RS models were otherwise identical, this small change to the tolerance range created striking differences in simulation outcomes. While RCA evolved altruistic donation rates of 73.6%, RS evolved rates of 1.48%.
- 3.2 Riolo et al (2002) disagreed with the Roberts and Sherratt's conclusion, but admitted that the ability of the RCA model to evolve cooperative behavior was limited to "some conditions." Spector and Klein (2006) further explored what these conditions might be. In their own implementation (SK) of the RCA model, the sensitivity of several model parameters was examined. Results demonstrated that when an agent's interactions were restricted to a subset of the total population, in contrast to the population-wide interactions in RCA, high levels of altruism evolved even using the RS requirement that $T \in [-10^{-6}, \infty)$.
- 3.3 In particular, SK restricted the interaction pool of agents by assuming the population was structured as a ring. They then parameterized how far along this ring an agent could interact when seeking donation partners as well as reproduction partners. With a population of 100 agents, this neighbor radius was implicitly fixed at 50 in the RCA and RS models, but was varied by SK. In the case of radius = 1, agents could only interact with two agents, one on either side. With radius = 1 and all other parameters matching RS, altruistic donation rates in the SK model rose from 1.48% (observed in RS) to 55.8%. Spector and Klein further showed that lowering the mutation rate from 0.1, used by RCA and RS, to 0.01 increased rates of altruism to 84.1%.



Ultraweak, weak, medium and strong variants of cheating

- 4.1 In this paper we assert that the above (RCA, RS and SK) models suffer from a narrow definition of cheating with regard to tag-mediated altruism. We term the original RCA model definition of cheaters as *ultraweak cheaters*, which RS argue are actually not true cheaters because they always act altruistically toward an agent displaying an identical tag to its own.
- 4.2 RS and SK use what we term *weak cheaters*. These are agents that may evolve a tolerance so low that they do not donate to any other agent but will likely receive donations. However, these agents still have a trait for tolerance and the mechanism for altruistically donating. If mutation drives the tolerance of a weak cheater back to a positive value, it may resume altruistic behavior.
- 4.3 In this paper we go further and explore the effects of what we define as *medium* and *strong cheaters*. These are agents that have no ability to donate and thus no tolerance trait. They simply display a tag and reap the rewards of displaying that tag to altruistic donors. There is no possibility of a mutation in *tolerance* causing a return to altruistic behavior because they have no capacity for such behavior and thus - they effectively have no tolerance trait. These stronger cheaters are created through addition of a binary trait that turns off or on the ability to act altruistically. In some treatments we allow this trait to mutate in either direction, resulting in agents we label as *medium cheaters*, while in others it may only turn off, resulting in *strong cheaters*. This latter treatment is an important case because it is characteristic of many real-world biological instances, particularly when interactions are interspecific (Buschinger 2009).
- 4.4 Numerous biological examples exist in which individuals of one species exhibit a tag allowing it to exploit the cooperative tendencies of a host population. This phenomenon is known variously as social parasitism or aggressive mimicry. These tags can include biochemical signatures, functional structures, or other detectable features exhibited by cheaters (Hölldobler & Wilson 1994; Stebbins & Galan 2001) that induce an altruistic response from others. Some more commonly known examples of interspecific tag cheating include several species of birds that lay their eggs in the nests of other bird species (Rothstein 1990; Zink 2000). The exploited mother provides food and shelter to the cheater, reducing resources allocated to her own offspring, while the cheater typically contributes nothing material to the host family. Similar examples of nest exploitation exist across many species of insects and spiders (Crowe, Fitzgerald, Remington, Ruxton, & Rychtář 2009; Zink 2000) - especially ants (Buschinger 2009; Hölldobler & Wilson 1994; Kronauer & Pierce 2011) - and often involve the cheaters actually eating offspring of the host population without detection.
- 4.5 It should be noted that in the biological examples above, the invasion of a parasitic species is analogous in our model to a strong cheater arising via mutation. However, even though in one case cheaters are introduced by invasion and in the other by mutation, the effect is identical. Members of the host population, reacting to tags exhibited by the cheater, consider the cheater to be one of its own species and, once established, the cheater's species or strategy can propagate through the population.

Table 1: A comparison of the current simulation models and the three previous models on which the current model is built.

Model	Tolerance can evolve below 0?	Neighborhood size can be less than full population?	Agents can turn off tag-tolerance mechanism?	Difference from previous model
-------	-------------------------------	---	--	--------------------------------

RCA	No	No	No	Allowed tolerance to evolve below 0
SK	Yes	Yes	No	Allowed neighborhood size to be less than the full population
current	Yes	Yes	Yes	Allowed agents to deactivate the tag-tolerance mechanism



The simulation model

- 5.1 We re-implemented the RCA, RS and SK models and explored, through simulation, different treatments (parameter value sets) with populations allowing the four distinct cheater types: *ultraweak, weak, medium and strong*.
- 5.2 When populations are composed of medium cheaters, the binary trait that turns off the altruistic pathway can be reversed through mutation, allowing a resumption of altruistic behavior (assuming the agent's tolerance is sufficiently high). In the treatment with strong cheaters, the binary tag determining whether an agent donates or not cannot be reversed - once the tag-tolerance mechanism for behavior is deactivated it remains so permanently.
- 5.3 It is important to note that simulations allowing medium or strong cheaters are initialized so that all agents have their binary trait turned on at the start, meaning that all agents begin interacting exactly as in the RCA model. Hence medium or strong cheaters can only arise through subsequent mutation. Consequently, we are exploring whether there is evolutionary pressure to turn off the tag-tolerance mechanism and, in simulations with medium cheaters, to turn it back on.
- 5.4 Using each of the four cheater types defined above, we tested the sensitivity of several model parameters. Treatments included high, standard, and low values for each of the model parameters: population size N, pairings per generation P, cost to donate c, mutation rate m, and neighbor radius R.
- 5.5 In all simulations the benefit conferred during an altruistic act = 1.0, the number of generations per simulation run = 30,000, and the number of replications (independent simulation runs with a different initial pseudo-random number seed) using each parameter set = 100. During the final 1,000 generations of each simulation, data were collected tabulating the percentage of interactions that resulted in an altruistic act (donation rate) and the mean tolerance value among all agents in the population. These values were then averaged across all 100 replications of a parameter set. Standard deviations were also calculated for both primary output statistics but were relatively small and are not presented here.
- 5.6 Collection of data over the final 1,000 generations differs from the RCA, SK, and RS studies, in which mean donation rates and tolerances were collected over the full 30,000 generations. This step was taken to avoid distortions due to stochastic noise and aberrations associated with the first several generations. We feel it more accurately reflects the long-run evolutionary outcome of these simulations. This caused no discernible differences in mean donation rates, but did result in lower mean tolerance values in all cases, sometimes by as much as an order of magnitude. For instance, in our reimplementation of the RCA model, average tolerance over 30,000 generations was 0.018, in exact agreement with the original RCA paper. However, over the final 1,000 generations the average tolerance was only 0.006. All statistics reported hereafter, except where noted, are based on data collected over the final 1,000 generations of each simulation run.



Results and discussion

- 6.1 Results of the array of treatments are presented in Table 2. The treatment column on the left indicates the varied parameter against the standard treatment (given in the first row of the column). On the right are shown the average donation rates for runs using each of the four cheater types. The table indicates with footnotes where the results reproduce reported values from the RCA, RS and SK models for specific parameter sets.

Effects of medium and strong cheaters

- 6.2 In our implementation of the RCA model, the introduction of strong cheaters caused donation rates to drop from 73.8% to 0.00%. With medium cheaters, in which the binary trait of non-donor was reversible, donation rates were only 1.6%. This is puzzling since Riolo et al (2002) claim that in their RCA model,

if unconditional defection is introduced by adding a binary trait that controls whether agents never donate, or donate using tags and tolerance, we find that cooperation also emerges, but again the extent of cooperation depends on many factors.
- 6.3 Though the paper suggests that the authors experimented with either medium or strong cheaters by adding a binary trait as we have done, it does not discuss how it was implemented, the levels of cooperation that emerged, or the many factors on which cooperation depended. In contrast, we find very low levels of altruistic behavior in the presence of medium cheaters and none using strong cheaters. Under every parameter combination we tested with strong cheaters, simulations evolved populations consisting entirely of non-donating strong cheaters. This complete displacement of the initial population by strong cheaters was usually accomplished within 10-20 generations (see Figure 1 for a comparison of temporal dynamics).
- 6.4 In our implementation of the SK model, with neighbor radius = 1, mutation rate = 0.01, and tolerance allowed to drop to -10^6 (as in RS), populations evolved donation rates of 86.6% (compared to SK's result of 84.1% using the same parameters). However, introducing medium cheaters to the SK model caused donation rates to drop to 15.7%. With strong cheaters altruistic behavior disappeared completely. These results clearly demonstrate that cooperation based on tags and tolerance is contingent on how one constructs and defines a cheater.

Table 2: Donation rates evolved under several parameter sets. Red indicates which value in a treatment deviated from the standard (control) parameter set.

Treatment	Model parameters (other than cheater type)					Mean donation rate over final 1,000 generations			
	Pop. Size	Pairings	Cost	Mutation Rate	Neighbor Radius	Cheater type			
						Ultraweak	Weak	Medium	Strong
Standard	100	3	0.1	0.01	10	97.3%	16.9%	5.2%	0.0%
Pairings - low	100	1	0.1	0.01	10	88.0%	8.7%	1.7%	0.0%
Pairings - high	100	10	0.1	0.01	10	97.9%	16.8%	7.9%	0.0%
Cost - low	100	3	0.05	0.01	10	97.5%	16.9%	4.2%	0.0%
Cost - high	100	3	0.5	0.01	10	95.5%	8.3%	2.2%	0.0%
Mutation - low	100	3	0.1	0.001	10	99.8%	48.5%	0.3%	0.0%
Mutation - high	100	3	0.1	0.1	10	70.3%	20.3%	4.2%	0.0%
Radius - low	100	3	0.1	0.01	1	91.8%	86.6% ***	15.7%	0.0%
Radius - high	100	3	0.1	0.01	50	97.7%	8.4%	1.5%	0.0%
Population - low	40	3	0.1	0.01	10	98.1%	31.7%	2.6%	0.0%
Population - high	900	3	0.1	0.01	10	94.4%	33.5%	9.0%	0.0%
RCA model	100	3	0.1	0.1	50	73.8% *	3.8% **	1.6%	0.0%

* replication with parameters used in RCA (Riolo et al 2001), which reported 73.6%

** replication with parameters used in RS (Roberts and Sherratt 2002), which reported 1.5%

*** replication with parameters used in SK (Spector and Klein 2006), which reported 84.6%

Multiple cheaters

6.5 A possible criticism of our approach is that in treatments with medium and strong cheaters, we leave the lower tolerance bound at $T \geq -10^{-6}$, effectively allowing weak cheaters to evolve in addition to medium or strong types. One may argue that the combined effect of both types in the same population is the reason for the large drop in donation rates. To test for this, we ran additional simulations using medium and strong cheaters but with the lower tolerance bound at $T \geq 0$, as in RCA. In the case of strong cheaters, this change had no effect and all simulations still evolved to a population consisting of all non-donors. Using medium cheaters, populations evolved slightly higher donation rates. However, the effect was minimal and mean donation rates over all 12 parameter sets listed in Table 2 increased by an average of 1.3%.

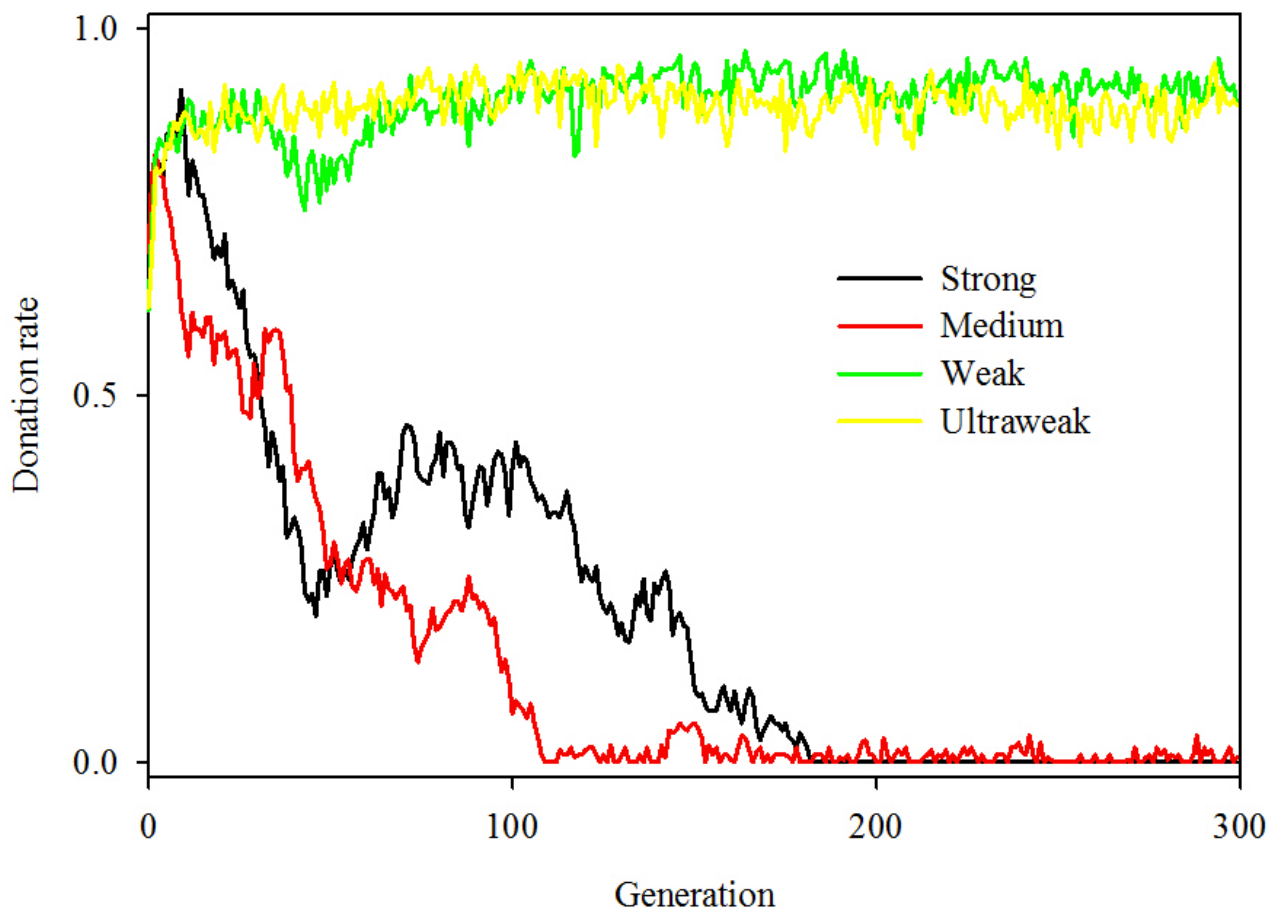


Figure 1a.

(a)

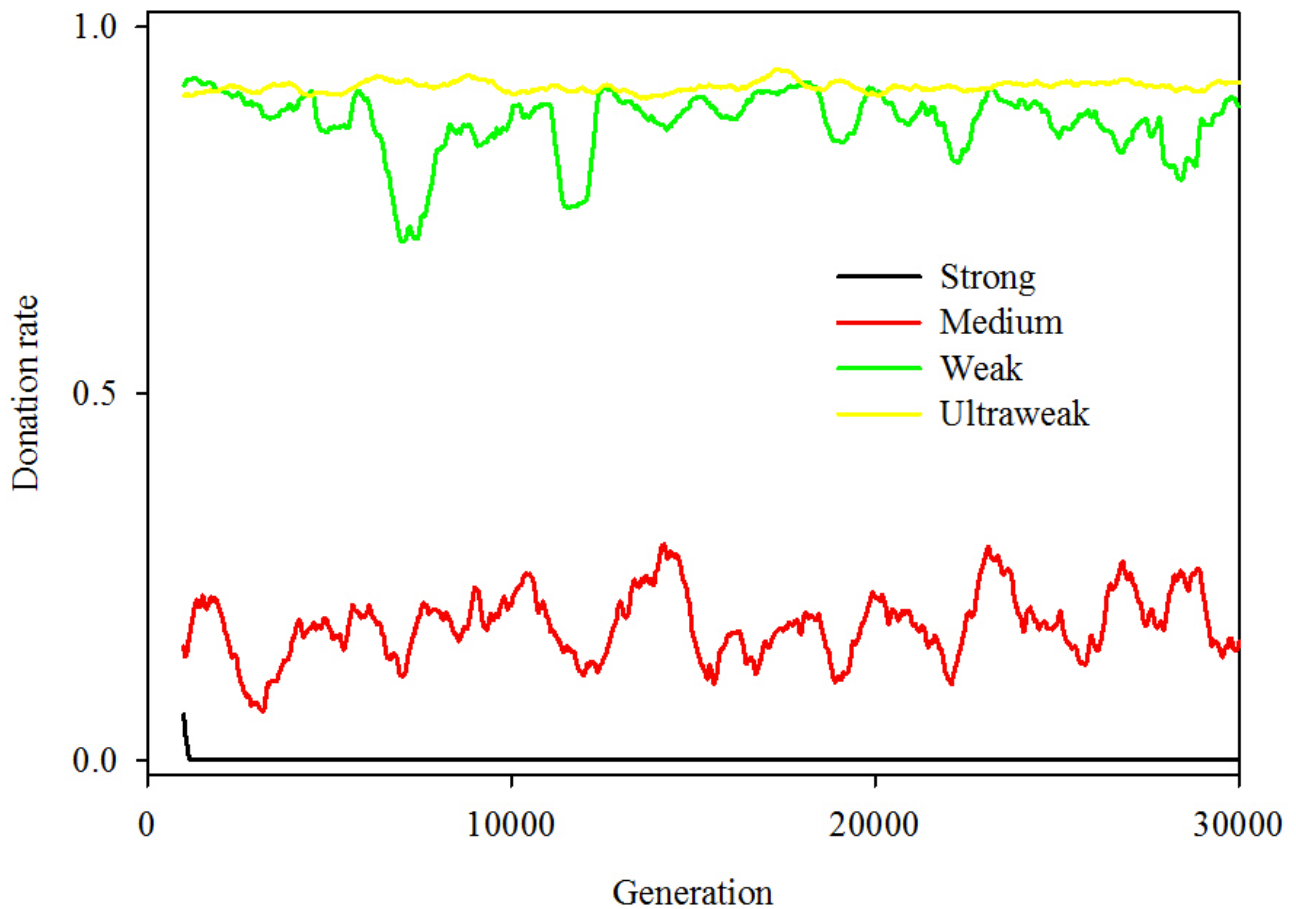


Figure 1b.

Figure 1. Temporal dynamics using four types of cheaters. Parameters were $N = 100$, $P = 3$, $R = 1$, $c = 0.1$, $b = 1$, $m = 0.01$, corresponding to treatment type "Radius - low" in Table 2. Results are presented for (a) the first 300 generations and (b) 1,000-generation moving average for the full 30,000 generations.

Effects of parameters governing interactions

- 6.6 Results show that parameters related to how agents interact with one another had significant effects on donation rates. These parameters include population size, neighbor radius, and pairings per generation. Since all treatments with strong cheaters result in full defection, and since treatments with weak cheaters are addressed by Spector and Klein(2006), we focus here on relevant results with medium cheaters. We found that decreasing neighbor radius or increasing population size led to higher donation rates. These parameters are related in that they reduce the percentage of agents within the total population with which a single agent is likely to interact. This finding concurs with conclusions by others that localization of interactions is a key to cooperation in structured societies (Flache 2002; Ohtsuki, Hauert, Lieberman, & Nowak 2006; Shutters 2011; Takács, Janky, & Flache 2008; Wilson & Wilson 2007).
- 6.7 However, this result is seemingly contradicted by the effect of pairings. Since decreasing pairings also decreases the number of potential interaction partners, one might expect, based on the previous paragraph, that its effect would be to increase donations. Instead, decreasing the number of pairings per generation actually decreased donation rates.
- 6.8 In attempting to explain this contradiction, we note that it is not completely accurate to say that increasing pairings increases the percentage of a population that an agent is able to interact with. Instead, it increases the percentage of the agent's *neighborhood* that it is likely to interact with. This is a subtle but important difference, as the neighborhood is a function of the neighbor radius. With respect to effects on cooperation, it appears that a tension exists between the number of interactions in which an agent engages and the number of potential interaction partners. This result suggests that restricting potential interaction partners allows a population to develop patches or pockets of cooperators, while increasing the number of interactions increases the likelihood that agents with similar tags *within such a patch* will find each other.
- 6.9 To disentangle the true effect of the three parameters that influence the number of interaction partners of an agent, we ran additional simulations testing the effects of pairings, neighbor radius, and population size in a 3x3x3 experimental arrangement, while keeping all other simulation parameters constant. Results confirm that, while increasing neighbor radius led to lower donation rates, for any given radius, increasing the number of pairings increased donation rates (Table 3).

Table 3: Combined effects of population size and neighbor radius for three different values of pairings per generation P. In all cases, increasing pairings increased donations while increasing radius decreased donations.

Population Size	Neighbor Radius	Mean donation rate over final 1,000 generations		
		P = 1	P = 3	P = 10
40	1	4.5%	9.0%	11.4%
40	10	1.6%	2.6%	3.1%
100	1	6.2%	15.7%	19.4%
100	10	1.7%	5.2%	7.9%
100	50	1.1%	1.5%	3.2%
900	1	6.2%	16.6%	18.3%
900	10	2.2%	9.0%	17.3%
900	50	1.0%	1.8%	9.2%

Note: in all cases, cost to donate = 0.1, mutation rate = 0.01, cheater type = medium

- 6.10 This finding relates to an ongoing debate over the effect that interaction density has on cooperative outcomes. A long-held assertion is that when a population is more densely connected the likelihood of cooperation increases (Marwell & Oliver 1993; Opp & Gern 1993). As Jun and Sethi (2007, p. 625) assert, "dense networks are more conducive to the evolution of cooperation." On the other hand, recent research suggests the opposite, demonstrating that denser networks inhibit cooperation in structured populations (Flache 2002; Flache & Macy 1996; Shutters 2011; Takács et al. 2008).
- 6.11 This study suggests the possibility that both sides of this debate are correct and that the debate is a product of semantic confusion. Each side's assertion depends on their answer to the question - what is a network? Is it a set of links between *potential* interaction partners? Or is it a set of links representing actual interactions? Confusion over this subtle distinction could lead to the debate described above. If a network is built of potential interactions, i.e. the neighbor radius, then Table 3 shows that increasing network density does diminish cooperative outcomes. Yet it also shows that if a network is built of actual interactions, i.e. the number of pairings, then increasing the density of those interactions, through pairings, increases cooperation.

Effects of mutation rate

- 6.12 As Spector and Klein (2006) revealed, mutation rate has a complex effect on altruistic outcomes. Under ultraweak cheaters increasing mutation rate led to lower donation rates. Because populations with ultraweak cheaters typically converge so that the population is nearly homogeneous with respect to tag value and agents have a very small tolerance value (Riolo et al. 2001), it is not surprising that such a population would benefit by minimizing disruptions in the form of random mutation.
- 6.13 In contrast to ultraweak cheaters, donation response to mutation rate was not strictly monotonic for either weak or medium cheaters. Under both types there was no statistically significant change in donation rate when decreasing mutation rate from $m = 0.1$ to $m = 0.01$. As mutation rate dropped further to $m = 0.001$ donation rates under both cheater types did change significantly. However, responses were opposite and while donation rate increased under weak cheaters from 16.9% to 48.5%, donations decreased from 5.2% to 0.3% using medium cheaters.
- 6.14 It is likely that the same beneficial effect of minimizing disruptions under ultraweak cheaters applies to weak cheaters. However, medium cheaters should benefit from the increased diversity that results from higher mutation because it allows for more frequent formation of new altruistic clusters, clusters which are potential targets of exploitation for the medium cheater.
- 6.15 A more detailed explanation of the effects of mutation on a non-tolerance based tag model are discussed by Hales and Edmonds(2005). They examine the effect of varying the mutation rates of the tag and strategy independently in the context of discrete tag models with binary strategy traits. The study found that if the mutation rate applied to the tag was sufficiently larger than that applied to the strategy high levels of cooperation emerged. This allowed for new cooperative tag groups to be created before existing groups were invaded by defection. It is unclear, however, if similar processes are operating in the models presented here.



Conclusion

- 7.1 We have explored the emergence of altruism in three existing evolutionary tag models using four different forms of cheating: ultraweak, weak, medium, and strong. Over a range of parameter values we found that these have dramatic effects on the rates of altruism observed, reducing or eliminating altruism in many instances.
- 7.2 We believe this indicates the importance of a careful examination of the assumptions underlying any simulation that models cheating behavior because the way cheating is defined may be as important as other model parameters. This becomes highly relevant if such models are to be used to understand or generalize about actual biological or social systems. This is particularly important for sociological interpretations such as "old school tie" and "secret handshakes" (Sigmund & Nowak 2001; Traulsen & Nowak 2007)^[4].
- 7.3 Also research in distributed computer systems applying tag approaches, such as in peer-to-peer systems, has demonstrated further classes of possible cheating behavior, such as lying about fitness and strategy, that could easily be enacted in social systems. Put simply, a cheater has an incentive to prevent the spread of its own cheating behavior and rely on others to cooperate so that the cheater can continue its exploitation (Arteconi, Hales, & Babaoglu 2007).

The future of tags

- 7.4 One might conclude from this study that the efficacy of tags as a mechanism for the evolution of cooperation, as presented in the RCA, RS, and SK models discussed in this paper, is limited by the existence of non-donors, both reversible and not. Yet this conclusion is not accurate. Indeed one of our hopes in producing this study is to offer a direction for future research related to tags. We close by suggesting at least two mechanisms by which tags may continue to induce cooperative outcomes even in the presence of medium or strong cheaters.
- 7.5 The first is through the development of mechanisms by which donating agents can first recognize non-donors and eventually punish them. Such policing behavior, as well as retaliation by those being policed, is widespread among social organisms (Frank 1995; Marlowe et al. 2008). Such developments typically result in long-term evolutionary arms races, the so called red queen effect (Dawkins & Krebs 1979), with complex population dynamics reminiscent of host-parasite systems.
- 7.6 The second is through multi-level selection, in which populations of agents compete against other populations for survival (Bowles 2006; Goodnight 2005; Traulsen & Nowak 2006; Wilson & Wilson 2007). Those populations that can internally regulate or eliminate non-donors typically displace populations that cannot (Güerke, Irlenbusch, & Rockenbach 2006), enabling tag-mediated altruism to persist dynamically with cheaters in a larger metapopulation.
- 7.7 Both of these mechanisms are not only fertile grounds for future investigation but are well-suited to broad interdisciplinary research programs.



Appendix

Algorithm pseudo code and parameter details

- 8.1 The simulation model used in this paper is described below. The model explores the conditions under which evolving agents altruistically donate to each other using a "tag" approach. The simulation replicates several existing models under certain parameter settings and provides novel results in other cases. As discussed in the main text of this article the previous models replicated are the RCA, RS, SK models. Results from different runs with various parameter settings are given in Table 2.

Parameters:

G = Number of generations = 30,000

The number of generations to run the simulation for.

N = Population size = {40, 90, 100}

The number of agents in the population.

b = Benefit of donation = 1

The amount of fitness benefit an agent receiving a donation obtains.

c = Cost of donation {0.05, 0.1, 0.5}

The amount of fitness cost applied to an agent making a donation.

P = Pairings {1, 3, 10}

The number of random pairings made for each agent in each generation.

m = Mutation rate {0.001, 0.01, 0.1}

The probability that agent traits will be mutated during reproduction to the next generation.

R = Radius of interaction {1, 10, 50}

The radius around the agent from which random pairings with other agents are drawn. Constraint: $0 < R \leq n/2$. when $r = N/2 = 50$, radius is entire population (as in the RCA model). When $R = 1$, radius is only immediate two neighbours (as in SK model).

TLB = Tolerance Lower Bound. TLB $\{-10^{-6}, 0\}$

The tolerance lower bound determines the lowest value that any agent tolerance can be. This is determined by the CT parameter below.

CT = Cheater type. CT {ultraweak, weak, medium, strong}

The cheater type determines the kind of cheaters that should be implemented in the simulation. The following constraints hold:

If CT = ultraweak or weak then BT = 1 for all agents at all times.

If CT = ultraweak then TLB = 0, otherwise TLB = -10^6

Agent state variables:

t_i = tag value of agent i. $t \in [0..1]$.

The tag value is a real number that is attached to each agent and is observable by other agents.

T_i = tolerance value of agent i. $T \in [TLB, \infty]$.

The tolerance value is a real number stored by each agent used, in combination with tag values, to determine if to make a donation to another agent.

BT_i = binary trait of agent i. $BT \in \{0,1\}$.

The binary trait, stored by each agent, determines if to make donations or not: 0 = turn off all donation, 1 = donate using tag / tolerance mechanism.

f_i = fitness of agent i. $f \in [0, \infty]$.

Agents accumulate a fitness value in each generation through interactions involving costs and benefits.

Main Outline Algorithm:

```
Initialise population of agents ()    // see function below
Loop for G generations
  Loop for each agent A in population N
    Loop for P pairings
      Select a random agent B within agent A's radius
      Agents Interact (A, B)          // see function below
    End loop
  End loop
  Reproduce Population ()            // see function below
End loop
```

Functions used in main outline algorithm:

// Initialise population of agents

```
Function Initialise population of agents ()
  Create a ring of size N
  Loop for each agent A in population N
    Place agent A at next free position on ring
    Set tag ( $t_A$ ) and tolerance ( $T_A$ ) to uniform random values from  $[0..1]$ 
    Set binary trait to 1 ( $BT_A = 1$ )
    Set fitness of agent to 0
  End loop
End Function
```

// Agents A and B interact, A decides if to make a donation to B

```
Function Agents Interact (agent A, agent B)
  If binary trait of A is turned on ( $BT_A = 1$ ) then
    If tags of A and B are within tolerance of A (  $|t_A - t_B| \leq T_A$  ) then
      Fitness of B is increased by benefit ( $f_B = f_B + b$ )
      Fitness of A is decreased by cost ( $f_A = f_A - c$ )
    End if
  End if
End function
```

// Reproduce a new population based on fitness of agent in current population

```

Function Reproduce Population ( )
  Create a new ring of size N for the new population
  Loop for each agent A in population N
    Select a random agent B within agent A's radius R
    Reproduce agent (A or B) with higher fitness# (f) as agent C
    Agent C is placed in the same location on the ring as agent A
    With probability m mutate* tag (tC) and tolerance (TC) of agent C
    If CT = medium then
      with probability m mutate* binary trait (BTC)
    else if CT = strong and binary trait is turned on (BTC = 1) then
      with probability m mutate binary trait (BTC)
    End if
  End loop
  New population becomes the current population
  All agents in new population have fitness initialised to zero
End function

```

Notes:

[#]Note: if the fitness of A and B are the same a random selection is made.

*Note: mutation takes the following form:

A tag (t) is replaced with a uniformly random value from [0..1].

Tolerance (T) is changed by adding Gaussian noise with mean 0 and standard deviation 0.01. If this results in T < TLB then T is reset to TLB.

A binary trait (BT) is flipped (such that 1 becomes 0 and vice versa).



Acknowledgements

We thank Tate Holbrook and three anonymous reviewers for helpful comments related to this study.



Notes

¹ Axelrod mainly considered fixed labels that do not evolve over time.

² Holland does not present a detailed model or results but cites personal communication with Riolo in 1992.

³ The original RCA model did not use a fair coin toss to select between equal fitnesses but rather a "selected bias" approach (see Edmonds and Hales 2003). However for the purposes of the results presented here this difference is minor and we have implemented a fair selection.

⁴ Note that the work of Traulsen & Nowak does not suffer from the weak cheating issues because they effectively use medium cheaters and dispense with tolerance and searching mechanisms.



References

- ARTECONI, S., Hales, D., & Babaoglu, O. (2007). Greedy Cheating Liars and the Fools Who Believe Them. In S. A. Brueckner, S. Hassas, M. Jelasity & D. Yamins (Eds.), *Proceedings of the 4th international conference on engineering self-organising systems (ESOA'06)* (pp. 161-175). Heidelberg, Germany: Springer-Verlag. [doi:10.1007/978-3-540-69868-5_11]
- AXELROD, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- BOWLES, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science*, *314*(5805), 1569-1572. [doi:10.1126/science.1134829]
- BUSCHINGER, A. (2009). Social parasitism among ants: a review (*Hymenoptera: Formicidae*). *Myrmecological News*, *12*, 219-235.
- COHEN, M. D., Riolo, R. L., & Axelrod, R. (1999). *The emergence of social organization in the prisoner's dilemma: How context preservation and other factors promote cooperation. Working paper 99-01-002*. Santa Fe Institute. Santa Fe, New Mexico.
- COHEN, M. D., Riolo, R. L., & Axelrod, R. (2001). The role of social structure in the maintenance of cooperative regimes. *Rationality and Society*, *13*(1), 5-32. [doi:10.1177/104346301013001001]
- CROWE, M., Fitzgerald, M., Remington, D., Ruxton, G., & Rychtář, J. (2009). Game theoretic model of brood parasitism in a dung beetle *Onthophagus taurus*. *Evolutionary Ecology*, *23*(5), 765-776. [doi:10.1007/s10682-008-9271-3]
- DAWKINS, R., & Krebs, J. R. (1979). Arms Races between and within Species. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 489-511. [doi:10.1098/rspb.1979.0081]
- EDMONDS, B., & Hales, D. (2003). Replication, Replication and Replication: Some Hard Lessons from Model Alignment. *Journal of Artificial Societies and Social Simulation*, *6*(4), <http://jasss.soc.surrey.ac.uk/6/4/11.html>.
- FLACHE, A. (2002). The rational weakness of strong ties: Failure of group solidarity in a highly cohesive group of rational agents. *Journal of Mathematical Sociology*, *26*(3), 189-216. [doi:10.1080/00222500212988]
- FLACHE, A., & Macy, M. W. (1996). The weakness of strong ties: Collective action failure in a highly cohesive group. *Journal of Mathematical Sociology*, *21*(1-2), 3-28. [doi:10.1080/0022250X.1996.9990172]
- FRANK, S. A. (1995). Mutual Policing and Repression of Competition in the Evolution of Cooperative Groups. *Nature*, *377*(6549), 520-522. [doi:10.1038/377520a0]
- GOODNIGHT, C. J. (2005). Multilevel selection: the evolution of cooperation in non-kin groups. *Population Ecology*, *47*(1), 3-12. [doi:10.1007/s10144-005-0207-2]
- GÜRERK, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, *312*(5770), 108-111. [doi:10.1126/science.1123633]
- HALES, D. (2000). Cooperation without memory or space: Tags, groups and the prisoner's dilemma. In S. Moss & P. Davidsson (Eds.), *Multi-Agent-Based Simulation* (Vol. 1979, pp. 157-166). [doi:10.1007/3-540-44561-7_12]
- HALES, D., & Edmonds, B. (2005). Applying a socially inspired technique (tags) to improve cooperation in P2P networks. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, *35*(3), 385-395. [doi:10.1109/TSMCA.2005.846399]
- HAMILTON, W. D. (1964). Genetical Evolution of Social Behaviour 1 & 2. *Journal of Theoretical Biology*, *7*(1), 1-52. [doi:10.1016/0022-5193(64)90038-4]
- HOLLAND, J. (1993). *The Effect of Labels (Tags) on Social Interactions. Working Paper 93-10-064*. Santa Fe Institute. Santa Fe, New Mexico.

- HÖLLDOBLER, B., & Wilson, E. O. (1994). *Journey to the Ants: A Story of Scientific Exploration*. Cambridge, Massachusetts: Belknap Press.
- HUI, P. M., Xu, C., & Zheng, D. F. (2007). Cooperation in evolutionary snowdrift game: Networking effects. *International Journal of Modern Physics B*, 21(23-24), 4035-4040. [doi:10.1142/S0217979207045165]
- JUN, T., & Sethi, R. (2007). Neighborhood structure and the evolution of cooperation. *Journal of Evolutionary Economics*, 17(5), 623-646. [doi:10.1007/s00191-007-0064-6]
- KONNO, T. (2010). A condition for cooperation in a game on complex networks. *Journal of Theoretical Biology*, 269(1), 224-233. [doi:10.1016/j.jtbi.2010.10.033]
- KRONAUER, D. J. C., & Pierce, N. E. (2011). Myrmecophiles. *Current Biology*, 21(6), R208. [doi:10.1016/j.cub.2011.01.050]
- MARLOWE, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., . . . Tracer, D. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 275(1634), 587-590. [doi:10.1098/rspb.2007.1517]
- MARWELL, G., & Oliver, P. E. (1993). *Critical Mass in Collective Action*. Cambridge: Cambridge Publishing. [doi:10.1017/CBO9780511663765]
- NOWAK, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560-1563. [doi:10.1126/science.1133755]
- NOWAK, M. A., & May, R. M. (1992). Evolutionary games and spatial chaos. *Nature*, 359(6398), 826-829. [doi:10.1038/359826a0]
- NOWAK, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685), 573-577. [doi:10.1038/31225]
- NOWAK, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298. [doi:10.1038/nature04131]
- OHTSUKI, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature*, 441(7092), 502-505. [doi:10.1038/nature04605]
- OPP, K. D., & Gern, C. (1993). Dissident Groups, Personal Networks, and Spontaneous Cooperation - the East-German Revolution of 1989. *American Sociological Review*, 58(5), 659-680. [doi:10.2307/2096280]
- RIOLO, R. L. (1997). *The Effects of Tag-Mediated Selection of Partners in Evolving Populations Playing the Iterated Prisoner's Dilemma*. Working Paper 97-02-016. Santa Fe Institute. Santa Fe, New Mexico.
- RIOLO, R. L., Cohen, M. D., & Axelrod, R. (2001). Evolution of cooperation without reciprocity. *Nature*, 414(6862), 441-443. [doi:10.1038/35106555]
- RIOLO, R. L., Cohen, M. D., & Axelrod, R. (2002). Does similarity breed cooperation? Reply. *Nature*, 418(6897), 500-500. [doi:10.1038/418500a]
- ROBERTS, G., & Sherratt, T. N. (2002). Does similarity breed cooperation? *Nature*, 418(6897), 499-500. [doi:10.1038/418499b]
- ROTHSTEIN, S. I. (1990). A Model System for Coevolution: Avian Brood Parasitism. *Annual Review of Ecology and Systematics*, 21(1), 481-508. [doi:10.1146/annurev.es.21.110190.002405]
- SANTOS, F. C., & Pacheco, J. M. (2006). A new route to the evolution of cooperation. *Journal of Evolutionary Biology*, 19(3), 726-733. [doi:10.1111/j.1420-9101.2005.01063.x]
- SHUTTERS, S. T. (2011). Punishment leads to cooperative behavior in structured societies. *Evolutionary Computation*, 20(2), 301-319. [doi:10.1162/EVCO_a_00062]
- SIGMUND, K., & Nowak, M. A. (2001). Evolution: Tides of tolerance. *Nature*, 414(6862), 403-405. [doi:10.1038/35106672]
- SPECTOR, L., & Klein, J. (2006). Genetic stability and territorial structure facilitate the evolution of tag-mediated altruism. *Artificial Life*, 12(4), 553-560. [doi:10.1162/artl.2006.12.4.553]
- STEBBINS, C. E., & Galan, J. E. (2001). Structural mimicry in bacterial virulence. [10.1038/35089000]. *Nature*, 412(6848), 701-705.
- TAKÁCS, K., Janky, B., & Flache, A. (2008). Collective action and network change. *Social Networks*, 30(3), 177-189. [doi:10.1016/j.socnet.2008.02.003]
- TRAULSEN, A., & Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, 103(29), 10952-10955. [doi:10.1073/pnas.0602530103]
- TRAULSEN, A., & Nowak, M. A. (2007). Chromodynamics of Cooperation in Finite Populations. *Plos One*, 2(3), e270. [doi:10.1371/journal.pone.0000270]
- TRIVERS, R. L. (1971). Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46(1), 35-57. [doi:10.1086/406755]
- WILSON, D. S., & Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *Quarterly Review of Biology*, 82(4), 327-348. [doi:10.1086/522809]
- ZINK, A. G. (2000). The Evolution of Intraspecific Brood Parasitism in Birds and Insects. *American Naturalist*, 155(3), 395-405. [doi:10.1086/303325]