

UC Irvine

Structure and Dynamics

Title

Evolution of Cultural Groups and Persistent Parochialism

Permalink

<https://escholarship.org/uc/item/2cr2681m>

Journal

Structure and Dynamics, 4(2)

Author

Kim, Jae-Woo

Publication Date

2010-11-02

DOI

10.5070/SD942003311

Peer reviewed

Introduction

The evolution of cooperation in human societies has attracted scholarly attention across disciplines. Several mechanisms are proposed such as kin selection (Hamilton 1964), direct reciprocity (Trivers 1971; Axelrod 1984), and indirect reciprocity (Alexander 1987; Boyd and Richerson 1989). However, it is a theoretical and empirical puzzle to explain society-wide and stable cooperation in sizeable groups of non-kin, given the setting of one-shot interaction or exchange without reciprocity. Network reciprocity (Nowak and May 1992) has been regarded as another route to cooperation in a one-shot Prisoner's Dilemma.

All those mechanisms commonly highlight the importance of preferential interactions among cooperators. However, the first three mechanisms above consider the role of 'cues' in the evolution of cooperation (McElreath et al. 2003). The kinship hypothesis is that the propensity to cooperate between two individuals is determined by the genetic relatedness between them. Kinship is a cue in this kin-based reciprocity. Direct reciprocity is predicated on repeated interactions between the same two individuals. Therefore, the behavior of an opponent on the previous round can serve as a cue in the Tit-for-Tat strategy. An individual's image score is a heuristic cue in an indirect form of reciprocity, that is, 'your cooperative behavior will be rewarded by third parties (not necessarily the recipient of your help) reading your increased reputation,' as in the model by Nowak and Sigmund (1998).

The last mechanism stresses the role of 'territoriality' (network embeddedness) in the evolution of cooperation. The argument for natural selection of defection is predicated on a well-mixed population, where everybody interacts equally likely and also randomly with everybody else. This approximation is used by all standard approaches to evolutionary game dynamics (Nowak 2006). But, real populations are not well mixed. Instead, some individuals interact more often than others, which means that people embedded in social networks interact with a subset of population. Natural selection favors defection over cooperation in 'unstructured' populations, but cooperation can be enhanced without either direct or indirect reciprocity among genetically unrelated agents with the help of 'network reciprocity.'

Table 1: Two Routes to Cooperation in Existing Research

	Cue-based Cooperation	Network Reciprocity
What actions to choose	Discriminators (Conditional strategies)	Indiscriminators (Unconditional strategies)
How partners are selected	Random Matching (Unstructured populations)	Spatiality (Structured populations)

Both 'cue-based cooperation' and 'network reciprocity' have complementary strengths and shortcomings (Table 1). Existing models of cue-based cooperation in the evolutionary Prisoner's Dilemma game assume unstructured populations in which agents randomly interact with one another (i.e. global mating or playing). However, they aptly underline the role of 'discriminators' (Nowak and Sigmund 2005) in the dynamics of cooperation. The

population in models of cue-based cooperation is composed of not only the ‘good’ guys (indiscriminating altruists) and the ‘bad’ guys (indiscriminating defectors) but also the discriminators who use conditional strategies depending on the trustworthiness of interaction partners. This approach is more realistic in the sense that cooperation in human societies more often hinges on our biological and cognitive capacities to use various symbols and signals than other animals. These heuristic cues, albeit inherently fallible, help human agents distinguish those who are more trustworthy (and more likely to cooperate) from those who are not.

In the existing classical models of network reciprocity, the population consists of the ‘good’ guys who unconditionally help all neighbors and the ‘bad’ guys who unconditionally refuse to help all neighbors, without the discriminators with intermediate levels of trust. In other words, it is unrealistically assumed that human agents have either the highest or lowest level of tolerance. However, research on evolutionary games in networks is aptly concerned with structural mechanisms of cooperation by emphasizing ‘assortative meeting’ (Eshel and Cavalli-Sforza 1982) through which cooperators benefit from more interactions with their own kind and less interactions with defectors. This approach implies that ‘viscosity’ (Hamilton 1964) – limited dispersal of offspring in the neighborhood instead of their random dispersal – renders network reciprocity workable (Grim et al. 2006) in animal societies, while clustering characterizes norm-generating and sustaining networks in human societies.

Tag-based Parochial Cooperation

One challenge to the theory of kin selection is that it assumes psychological mechanisms through which individuals can identify how much genes they share. “Of course, the trick is to distinguish between ‘us’ and ‘them.’ In the ancestral society, it was hardly a problem. Clan members were related by blood or ties of marriage, and lived their lives in close proximity” (Lopreato 2002: 423). In this aspect, proximity must be a critical heuristic cue for reciprocity on the basis of kin-recognition in earlier societies. However, how to explain the evolution of cooperation in increasingly differentiated and highly mobile societies today consisting of genetically unrelated people?

People strain to detect ‘ethnic markers’ such as genetically transmitted phenotypes (e.g. skin color) and behavioral characteristics (e.g. speech, manner), according to van der Berghe (1981: 28-29) who theorizes the dynamics of extended nepotism and conflict in ethnic groups as extended forms of the family. Stating that the theory of kin selection is silent on a route to altruism toward nonrelatives, Krebs (1987) regards ‘phenotype matching’ as vital for humans to identify who is who. In general, “people often related to each other in ways that are influenced by observable features such as sex, age, skin color, and style of dress. These cues allow a player to begin an interaction with a stranger with an expectation that the stranger will behave like others who share these same observable characteristics... This happens because the observed characteristics allow an individual to be labeled by others as a member of a group with similar characteristics” (Axelrod 1984: 146-7).

Holland (1993; 1995) suggests that ‘tags’ as observable markers are engaged in group processes of complex adaptive systems. Human agents may have tags as the phenotype of “memes” (Dawkins 1976) to form “memetic kin” (Heylighen and Campbell 1995).

Reading tags enables humans to make distinctions between in-group members and ‘others’ (cf. Tajfel 1974; Turner 1982). Recent research on tag-based systems reports that tag-based discriminating actions among randomly selected agents significantly enhance the level of global cooperation in a one-shot multi-agent Prisoner’s Dilemma game (Riolo et al. 2001).

Agents may display “in-group favoritism” (Hammond and Axelrod 2006; Efferson et al. 2008) or “parochialism” (Bowles and Gintis 2004) in choosing how to interact, based on their tolerance of differences with others. For instance, what action an ego chooses is contingent on whether or not her perceived social distance to an opponent is less than or equal to her tolerance level (Riolo et al. 2001). In Hammond and Axelrod’s model, one of four colors is assigned to agents with tags (ethnic markers) whereby their group membership is determined. An ethnocentric agent helps a partner of her own color, and otherwise does not help. In these examples, the predisposition of individuals to behave cooperatively depends on the identities of their interaction partners: ‘insiders’ are favored over ‘outsiders,’ but not necessarily with out-group hostility.

Issues in Tag-based Cooperative Societies

Riolo et al. (2001) demonstrate that if agents cooperate only with others with tolerably similar tags and they leave offspring in proportion to fitness measured by the payoffs at the previous generation, then societies reach high levels of cooperation. But, tolerance decreases drastically as the average level of cooperation (measured by the donate rate in a helping game) increases rapidly during the first few generations. After this transient period, “the agents in the resulting ‘dominant tag cluster’ have an advantage as there are more of them to help each other” (Riolo et al. 2001: 442). Riolo et al. (2001: 442) report that “about 75-80% of the agents have tags that are so similar that they are within each other’s tolerance range.”

Cooperators in the dominant tag cluster are, however, vulnerable to invasion by relatively intolerant mutants with tags within the range of tolerance of the typical members of the dominant cluster. Once these mutants have higher payoffs, a transition to a new tag cluster occurs. The average tolerance drops significantly once again, but the average donation rate returns to its previous level (Riolo et al. 2001: 441). A significant erosion of tolerance is a major trend, but if more tolerant agents benefit from each other to spread across the population, the average tolerance can increase temporarily. They conclude: “in our model, the cycle of increasing and decreasing tolerance could reflect, for example, a loss of sensory discrimination in a population when there is little selection pressure to retain it, followed by a recovery when a more discriminating individual succeeds” (Riolo et al. 2001: 442).

Riolo et al.’s research shows issues of tag-based cooperative societies. The ethnocentric tendency in the population becomes increasingly reinforced over generations since parochial agents displaying stronger discrimination against out-group are more likely to survive. Nonetheless, they continue to cooperate with each other because the degree of tag diversity in the population decreases over generations. In other words, agents adopt intolerance from more successful others as societies become more homogenized, but high levels of cooperation are still possible in spite of a huge erosion of tolerance because it is more likely that cooperation occurs among more homogeneous agents.

However, Riolo et al. do not explore theoretical possibilities of whether culture is globally divergent in spite of its local convergence. According to the replication by Edmonds and Hales (2003), emergent societies in Riolo et al.'s model almost always reach universal cooperation in completely homogenized populations consisting of agents who have the minimum tolerance and the identical markers. However, "tags can also present major obstacles in overcoming segregation... Although the simulations by Riolo et al. do not produce dominant clusters that split into rival tribes, any territorial distribution would favor such 'speciation.' Tags would then act as self-reinforcing stereotypes, making it hard for tolerance to cross the divide" (Sigmund and Nowak 2001: 405). Recall that agents in Riolo et al.'s model play a one-shot Prisoner's Dilemma game with randomly selected partners in unstructured populations. This is another reason why it is difficult for Riolo et al. to examine the spatial configuration of parochialism and tag clusters. The present study intends to investigate the dynamics of tag-based cooperation in structured populations where agents play the game with immediate neighbors, as in Kim (2010).

There is another fundamental limitation of Riolo et al.'s model. They discover that if a mutant with a lower level of tolerance has similar tags tolerable to existing cooperators, such an immigrant earns a higher payoff (fitness) so that she can easily invade the population. The number of offspring with her tolerance and tags increases over generations. Since less parochial agents perceive a broader range of interaction partners as in-group, they are more likely to be vulnerable to more parochial mutants displaying similar tags. However, there are no unconditional defectors in their model since agents with the minimum tolerance are those who still cooperate with partners displaying the identical tags. We seek to examine the evolutionary stability of parochial cooperative strategies in the presence of mutant defectors displaying similar tags.

Existing studies on cue-based cooperation (Nettle and Dunbar 1997; Miller et al. 2002) commonly report its instability in the presence of defectors who learn 'secret handshaking' (Robson 1990). For example, cooperation can evolve through signaling among cooperators, but it fails to continue immediately after defectors are introduced who use the same communication token (Miller et al. 2002). For similar reasons, we expect that cooperation on the basis of similarity in tag-based societies tends to disintegrate in the face of defectors with tolerably similar markers because existing discriminating cooperators perceive those defectors as in-group to trust them.

In Search for Solutions

"Not only that the adoption of linguistic variables correlates with membership of a social group but that when a group feels itself threatened by outsiders, it will increase its usage of the linguistic markers that make it distinctive, thus producing divergence over time" (Nettle and Dunbar 1997: 94). In line with this sociolinguistic study by Labov (1972), Nettle and Dunbar (1997) simulate the dynamics of cooperation in artificial societies where agents have dialects in the setting of Prisoner's Dilemma. The main finding is that those who (POLYGLOT) cooperate only if their partner has a dialect similar to their own can beat unconditional defectors, but they are not robust against defectors (CHEAT) who can mimic dialects shared among POLYGLOT. However, as POLYGLOT changes their dialects more frequently, the chance of invasion by CHEAT significantly declines. Similarly, we first

consider the relationship between the speed of tag change and the speed of tolerance change. Hales (2004) and Edmonds and Hales (2005) suggest that “tags should mutate faster than strategies” for efficient and stable cooperation. In other words, cooperative tag groups need to spread by mutation of tags before free-riders (by mutation on strategies) invade the group.

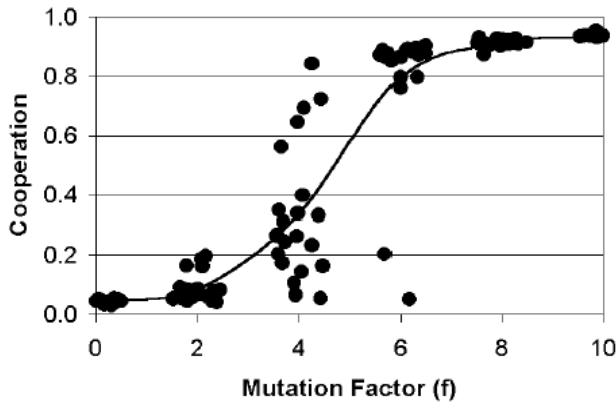


Figure 1: Level of Cooperation against Mutation Factor (Hales 2004: 8)

Applying different rates of mutation μf and μ to the tag change and the strategy change respectively, Hales (2004) demonstrates that high levels of cooperation continue when the mutation factor $f > 5$ (Figure 1). From the dispersion of the level of cooperation across multiple replications in each experimental condition, we notice that cooperation also becomes more stable as tags mutate faster than strategies. It is how to select partners, not what actions to choose, that is influenced by ‘tagging’ in their models. That is, strategies are not necessarily based on tag-based perception and tolerance. Informed by Labov (1972) and Nettle and Dunbar (1997), we still want to apply the idea of “tags should mutate faster than strategies” to our models in which tolerance is a proxy strategy. We explore the dynamics of cooperation across different levels of the mutation factor.

The current study points out existing tag-based models are basically concerned with genetic evolution of cooperation, and they assume completely mutable tags. Consequently, indistinguishable mutant defectors can be born. In other words, defectors have no difficulties in mimicking cues shared among discriminating cooperators. From the viewpoint of cultural evolution of cooperation, tags are no longer completely mutable – some markers are more mimicable (e.g. linguistic codes), whereas some others are not (e.g. skin colors). We thus suggest conceptualizing tags as cultural markers instead of genetically inheritable ones. Agents may have core markers which are not subject to both imitation and its error. We construct two new societies, one where agents have non-negotiable binary tags at the same dimension (‘caste society’) and the other where core markers spread across different dimensions of the tag space (‘modern society’).

Models and Experimental Designs

Agents are selected in a random order without any particular schedules, and they are updated synchronously. All models consist of the same steps (Table 2). Our NetLogo model applets (Wilensky 1999) are available at OpenABM (<http://www.openabm.org>).

Table 2: Stages of Simulation

```
Initialization
For each generation (round),
  For each agent,
    Similarity perception
  End
  For each agent,
    Interactions
  End
  For each agent,
    Fitness calculation (payoff calculation)
  End
  For each agent,
    Reproduction (payoff-based imitation)
  End
End
```

Initialization. Each agent i has a list of binary tags with its length L . We assume that agents perceive similarity in a dichotomized way: $t_i \in \{0, 1\}^L$. $t_{i(l)}$ denotes agent i 's trait value on the l th position of her tag list. Each agent also has tolerance $T_i \in \{0, 1, \dots, L+1\}$. 100 agents with randomly assigned tags and tolerance are located on the torus. The number of adjacent neighbors is 8 (i.e. Moore neighborhood). We note here that tags are arbitrary markers in the sense that they are not innately related to the behavioral propensity to cooperate or not. Tolerance is an exact predictor of strategy (i.e. the higher tolerance, the more likely to cooperate with partners), but it is not observable in tag-based societies.

Similarity perception. If $L \geq 1$, agent i has perception of similarity to neighbor j based on Hamming Distance defined by $HD_{ij} = \sum_{l=1}^L |t_{i(l)} - t_{j(l)}|$. Consequently, $HD_{ij} = HD_{ji}$. If $HD_{ij} < T_i$, agent i accepts agent j as in-group; otherwise, out-group.

Interactions. Agents simultaneously play a one-shot Prisoner's Dilemma game with all their neighbors at each generation (or round) without expectations of (in-group) reciprocity. Agent i cooperates with her neighbor j (strategy $s_{ij} = 1$) if j is perceived as in-group; otherwise, agent i declines to cooperate (strategy $s_{ij} = 0$). If $L = 0$, there is no similarity perception since all agents are identical ($HD = 0$). Since $\min(T) = 0$ and $\max(T) = 1$ at $L = 0$, one group of agents always cooperate ($T = 1$), but the other group always defect ($T = 0$). In this way, our model at $L = 0$ represents a population composed only of unconditional cooperators and defectors (i.e. no parochial agents). Generally at $L \geq 1$, agents with the

minimum tolerance ($T = 0$) are unconditional defectors, while agents with the maximum tolerance ($T = L + 1$) are unconditional cooperators, but agents who have in-between values ($T \in \{1, 2, \dots, L\}$) are discriminating egoists.

Fitness calculation. If an ego cooperates, she pays cost c and produces benefit b . If she defects but her partner cooperates, she gets benefit b produced by the partner, without paying cost c . If $b > c > 0$, the payoff matrix in Table 3 satisfies the Prisoner’s Dilemma game in which mutual cooperation is Pareto optimal. The total payoff of agent i at each round can be defined by $\pi_i = \sum_{j \in N_i} -s_{ij}c + s_{ji}b$. Here, the neighborhood N_i is the set of immediate partners who are directly connected with agent i .

Table 3: Payoffs in Prisoner’s Dilemma Game

	Cooperate	Defect
Cooperate	$b-c$	$-c$
Defect	b	0

Reproduction. One player i is randomly chosen, and then another agent j ($j \neq i$) is randomly selected from N_i . Only if $\pi_j > \pi_i$, agent i dies without producing her offspring, and agent j can leave her offspring at her own site and also at the empty site once occupied by agent i . This reproduction rule is the same as the rule of local payoff-based imitation: each individual’s payoff is compared to the payoff of a role model drawn from the agent’s neighbors; if the role model, say agent j , turns out to be more successful than the ego, say agent i , agent i adopts tags and tolerance from agent j . Equivalent to the concept of mutation in genetic inheritance, imitating both tags and tolerance from successful neighbors may be disrupted by error with probability of mutation rate μ . In this case, agents i ’s tags and tolerance are replaced by a new set of traits – tags and tolerance – that are randomly generated.

We are concerned with four models. In Model 1, tags as genetically inheritable traits are completely mutable, as is described in the reproduction step above. On the contrary, tags as cultural markers are not necessarily mimicable in Model 2 and Model 3. Each agent has a core marker at the same dimension in ‘caste’ societies (Model 2), whereas in different dimensions in ‘modern’ societies (Model 3). Core markers are not subject to both payoff-based imitation and its error. Suffice to say here that the society as a whole has one unchangeable master culture in Model 2, whereas not in Model 3.

While exploring three models, we intend to test the idea that “tag should mutate faster than tolerance” for efficient and stable cooperation in tag-based societies. There are two major parameters in their studies: the rate of tag mutation (mf); and the rate of strategy mutation (m). f denotes the mutation factor, and m is equivalent to μ . For example, tags mutate 10 times faster than strategy (equivalent to tolerance in our study) if $f = 10$. Yet another model in the present study (Model 4) is exactly the same as Model 1 (i.e. markers are completely mutable), but we apply differential mutation rates to tags and tolerance given a new parameter f .

We first undertake a set of experiments across the first three models (Table 4). Under varying conditions of the benefit-to-cost ratio (b/c) and the tag length (L) in the absence of mutation or imitation error ($\mu = .00$), we examine the possibility of the survival of cooperation, the possibility of universal cooperation, and the average tolerance and the averaged social distances each agent feels toward her neighbors at cooperative equilibrium. Notice that Model 4 behaves exactly the same way as Model 1 does under no mutation.

Table 4: Parameter Setting of Experiments under No Mutation (Model 1, 2, and 3)

Parameters	Values
Benefit-to-cost ratio	2, 4, 6, 8, 10
Tag length	0, 2, 4, 6, 8, 10
Mutation rate	0

We explore the dynamics of cooperation in the presence of mutation ($\mu = .01$) under a single experimental condition ($b/c = 4$ and $L = 5$) when markers as genetic traits are completely mutable (Model 1). We present a typical run in Model 1 in order to show that cooperation is not robust against mutant defectors with tolerably similar markers. Next, we are interested in whether or not its long-run dynamics significantly differ in two societies where each agent has one core marker (Model 2 and 3). Under varying conditions of the benefit-to-cost ratio (b/c), we select an intermediate tag length ($L = 5$). The same rate of imitation error ($\mu = .01$) is used (Table 5). Each replication has 5,000 steps, and 100 replications are executed for each experimental condition.

Table 5: Parameter Setting of Experiments under Mutation (Model 2 and 3)

Parameters	Values
Benefit-to-cost ratio	2, 4, 6
Tag length	5
Mutation rate	.01

In order to test the idea, “tag should mutate faster than tolerance” for efficient and stable cooperation, we implement exactly the same design used in the previous experiment. The range of the mutation factor is displayed in Table 6. We run 100 independent replications for each experimental condition. Each replication has 5,000 steps.

Table 6: Parameter Setting of Experiments under Mutation (Model 4)

Parameters	Values
Benefit-to-cost ratio	2, 4, 6
Tag length	5
Mutation rate	.01
Mutation factor	5, 10, 15, 20, 25, 30

For both experiments in Table 5 and 6 in the presence of mutation (imitation error), we measure the average tolerance and the average social distance at the global level, as in the

first experiment. In order to examine the spatial configuration of parochialism and groups as tag clusters in structured populations, we introduce two new measures at the local level: the degree of link homophily in tags and the degree of link homophily in tolerance. The minimum is 0 and the maximum is 1 for both measures. For agent i , a function $f_j = 1$ only if $\overline{HD}_{ij} = \overline{HD}_{jr}$, and otherwise $f_j = 0$, given that $j \in N_i$ and $r \in N_j$. \overline{HD}_{ij} denotes $\frac{1}{k} \sum_r^k \overline{HD}_{ij}$. In the same way, \overline{HD}_{jr} indicates $\frac{1}{k} \sum_r^k \overline{HD}_{jr}$. k is the number of neighbors each agent i and j has. We can define the degree of link homophily in tags as $\frac{1}{N} (\sum_i^N (\frac{1}{k} \sum_j^k f_j))$. Here, N is the population size. Given another function $g_j = 1$ only when $T_i = T_j$, and otherwise $g_j = 0$, the degree of link homophily in tolerance can be measured by $\frac{1}{N} (\sum_i^N (\frac{1}{k} \sum_j^k g_j))$. Since unconditional defectors ($T = 0$) never cooperate toward neighbors, we exclude them from the population when calculating the degree of link homophily in tolerance.

Results

Evolution of Tag-based Cooperation under No Mutation in Model 1, 2, and 3

Our model at $L = 0$ is equivalent to the classical model of the evolutionary Prisoner's Dilemma game with network reciprocity. Since agents do not have markers, they cannot distinguish 'us' from 'them.' As a result, the population is purely composed of those who always help neighbors and those who always decline to do. Under this condition, there exist only three possible states in the long run in the absence of mutation. The equilibrium population consists of defectors (the 'bad' guys), altruists (the 'good' guys), or both. Either altruists alone or both altruists and defectors constitute emergent cooperative societies. Also, emergent societies at $L = 0$ are always culturally homogeneous, either cooperative or betrayal.

Table 7 shows that network reciprocity alone without 'labeling' is not sufficient for the survival of cooperation when $b/c < 8$. Recall that the interaction space is the Moore neighborhood. The results at $L = 0$ indicate that unconditional cooperators can survive facing defectors in structured populations only when the benefit-to-cost ratio in the Prisoner's Dilemma game is higher than the number of interaction partners. However, tag-based local interaction and the local dispersion of offspring (i.e. 'viscosity') significantly increase the likelihood of the survival of cooperation in all three models. When agents use only a few dimensions ($L = 2$) to distinguish 'us' from 'them,' cooperation is more likely to evolve even at the benefit-to-cost ratios lower than the number of neighbors. The likelihood that cooperators ($T \geq 1$) survive also increases as the relative benefit of cooperation increases. In all three models, if agents make more fine-grained distinctions with four or more dimensions of tags, the likelihood increases up to almost 100% within the whole range of the b/c ratios under experiment.

Table 7: Likelihood of Survival of Cooperation and Universal Cooperation in Model 1, 2, 3

L	b/c				
	2	4	6	8	10
0	0	0	0	0	43 (0)
2	23/0/0 (22/0/0)	56/18/39 (50/0/0)	73/62/92 (63/0/0)	62/83/98 (50/0/0)	88/96/98 (46/2/0)
4	100/94/99 (89/3/0)	100/89/100 (81/16/0)	98/93/100 (85/26/0)	96/97/100 (81/27/0)	98/100/100 (70/30/0)
6	100/95/100 (81/54/0)	100/100/100 (80/70/0)	100/100/100 (91/66/0)	98/100/100 (82/80/1)	100/100/100 (87/73/0)
8	100/98/100 (87/82/2)	100/100/100 (84/82/2)	100/100/100 (87/83/0)	100/100/100 (90/80/2)	100/100/100 (79/82/5)
10	100/100/100 (87/85/6)	100/100/100 (95/89/8)	100/100/100 (89/89/16)	100/100/100 (84/87/8)	100/100/100 (86/85/13)

Note: Results from Experiments in Table 4. The numbers at each cell represent how many times out of 100 trials societies reach the cooperation equilibrium in the absence of mutation. For example, when $b/c = 4$ and $L = 4$, it is 100% in Model 1, 89% in Model 2, and 100% in Model 3. The numbers in parentheses at each cell indicate the likelihood of universal cooperation. At $b/c = 4$ and $L = 4$, it is 81% in Model 1, 16% in Model 2, and 0% in Model 3.

We observe quite different patterns in the possibility of ‘universal’ cooperation across three societies. Universal cooperation is less likely to occur in both caste and modern societies in which each agent has one immutable marker resistant to cultural homogenization. This tendency is more remarkable when cooperation is relatively costly and the dimension of the tag space is comparatively small. Nonetheless, as the number of available cultural markers increases, there are no significant differences in the likelihood of universal cooperation between societies with completely mutable markers (Model 1) and caste societies where each agent has one non-negotiable marker at the same dimension. Universal cooperation is, however, much less likely to happen in modern societies where core markers are distributed across different dimensions than that in caste societies, given the same number of cultural markers. We note that the disparities between both societies do not decrease even if the tag space size increases.

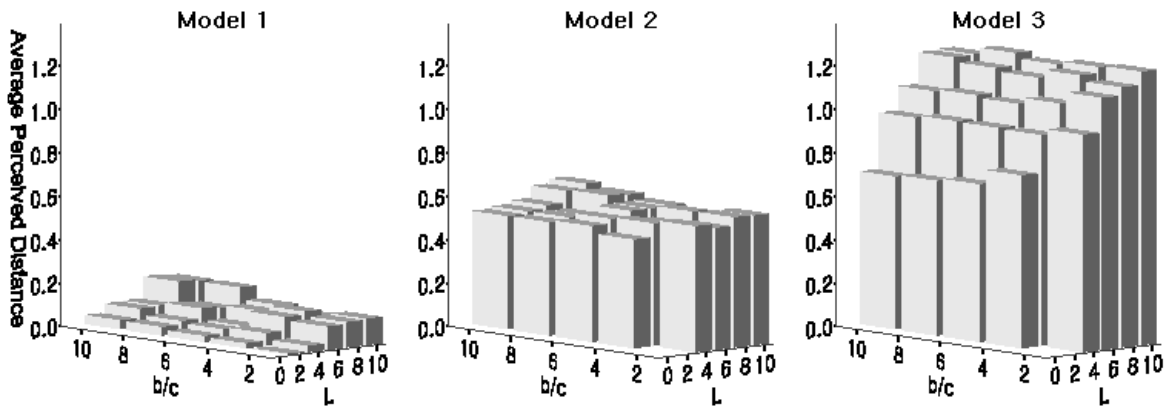


Figure 2: Average Perceived Distance at Cooperative Equilibrium in Model 1, 2, 3
Note: Results from Experiments in Table 4. Model 1 (societies with completely mutable tags); Model 2 (‘caste’ societies); Model 3 (‘modern’ societies)

Table 8: ANOVA Test on Average Perceived Distance in Model 1, 2, 3

L	Model 1	Model 2	Model 3	Post-hoc Test (Scheffé)
4	.06 (.134)	.55 (.122)	.98 (.118)	All pairs **
6	.10 (.192)	.56 (.145)	1.10 (.177)	All pairs **
8	.13 (.249)	.60 (.232)	1.21 (.256)	All pairs **
10	.14 (.299)	.60 (.235)	1.25 (.314)	All pairs **

Note: The number in each cell denotes the average. The number in each parenthesis indicates the standard deviation. * $p < .05$ ** $p < .01$ (two-tailed tests).

High levels of cooperation necessitate huge losses of tag diversity in parochial societies where markers as genetic inheritable traits are completely mutable (Model 1). The degree of tag diversity, in spite of society-wide cooperation, is much higher in both caste and modern societies because one immutable marker per agent increases heterogeneity (Figure 2). Holding the tag length constant, Model 3 ranks highest, Model 2 second highest, and Model 1 lowest in the average perceived dissimilarity. The differences between any two of them are statistically significant (Table 8).

Additionally, we notice that the average perceived social distance at the equilibrium of cooperation remains almost constant across varying lengths of tag strings in Model 2 (and Model 1), whereas it increases as the tag space dimension becomes larger in Model 3. In other words, although payoff-based imitation results in assimilative influence, agents in modern societies continue to look at each other from some distances, without losing the average degree of cultural heterogeneity at the beginning.

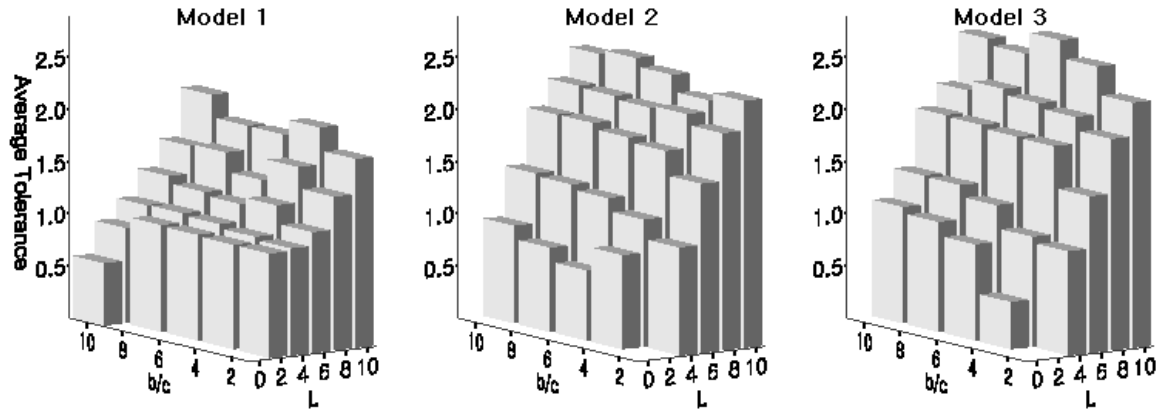


Figure 3: Average Tolerance at Cooperative Equilibrium in Model 1, 2, 3
Note: Results from Experiments in Table 4. Model 1 (societies with completely mutable tags); Model 2 ('caste' societies); Model 3 ('modern' societies)

Table 9: ANOVA Test on Average Tolerance in Model 1, 2, 3

L	Model 1	Model 2	Model 3	Post-hoc Test (Scheffé)
4	1.08 (.226)	1.32 (.453)	1.22 (.381)	All pairs **
6	1.28 (.457)	1.87 (.411)	1.84 (.367)	Model 1 and Model 2 ** Model 1 and Model 3 **
8	1.58 (.694)	2.16 (.405)	2.16 (.419)	Model 1 and Model 2 ** Model 1 and Model 3 **
10	1.94 (.904)	2.42 (.604)	2.59 (.549)	All pairs **

Note: The number in each cell denotes the average. The number in each parenthesis indicates the standard deviation. * $p < .05$ ** $p < .01$ (two-tailed tests).

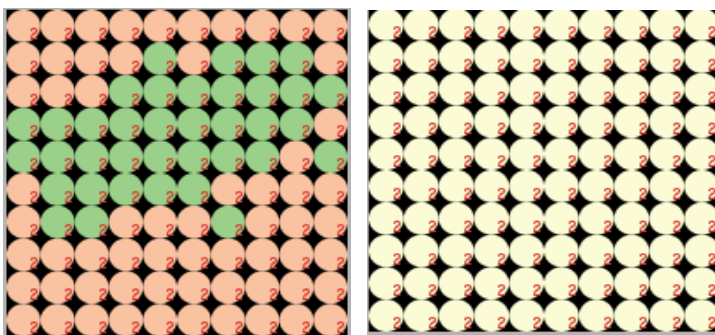
Figure 3 shows no differences in the average tolerance at the equilibrium of cooperation across three societies when the tag space is not much differentiated ($L = 2$). This is because the most discriminating cooperative strategy ($T = 1$) is dominant in emergent societies. In other words, as long as the tag space size is very small, both caste and modern societies also rapidly lose tolerance although they retain higher degrees of cultural heterogeneity than societies with completely mutable markers, as in Figure 2. At $L = 2$, the average tolerance (and its standard deviation) is .54 (.078) in Model 2, .71 (.051) in Model 3, and .03 (.082) in Model 1.

However, societies do not necessarily grow into a fully blown parochialism when there remain sufficient amounts of heterogeneity for partner identification. Figure 3 also indicates that both caste and modern societies with one immutable marker tend to preserve more tolerance, as they become less homogenized. At $L \geq 4$, the average tolerance in those societies at the cooperative equilibrium is significantly higher than the average in societies with genetically inheritable tags, according to the ANOVA test (Table 9). We note that as the tag space becomes larger, agents in modern societies tend to have higher levels of tolerance than those in caste societies.

Spatial Distribution of Tags and Tolerance in Model 1

Next, we examine the spatial distribution of markers and tolerance in emergent cooperative societies at equilibrium. We are particularly concerned with how it sometimes happens that emergent societies still reach high levels of global cooperation without the complete loss of tolerance or tag diversity, or without both. In other words, what kind of spatial topology makes it possible that ‘tagging’ serves as self-enforcing labeling to sustain the co-existence of within-group parochial cooperation and between-group non-cooperation at the global level in spite of assimilative influence at the local level?

The archetypal case of the cooperative equilibrium in Model 1 is that the most discriminating cooperators ($T = 1$) dominate the population. Recall that $T = 1$ is the minimum level for cooperation because those agents cooperate only toward neighbors with identical phenotypes, and otherwise refuse to do. Most of time societies based on the strongest parochialism reach universal cooperation in completely homogeneous populations. But, it occasionally happens even in Model 1 without core markers that less parochial agents (either with or without the most discriminating cooperators) govern emergent societies attaining universal cooperation.



(a) Distribution of Markers (b) Distribution of Strategies

Figure 4: Nonaligned Groups under a Single Level of Tolerance in Model 1

Note: The number colored red in each agent indicates her tolerance level. Agents of the same color have the same tag list in (a). The brightness of each node in (b) is proportional to how many neighbors with whom she cooperates out of 8. The brighter, the more cooperation. $T = 2$ in (b). $b/c = 4$, $L = 4$, and $\mu = .00$.

Panel (a) of Figure 4 shows the emergence of two groups with different sets of markers salient (green and pink). As in (b), all members in the population cooperate with their neighbors. We observe that tag diversity remains in spite of universal cooperation in an emergent cooperative society purely consisting of agents with $T = 2$. Recall that more tolerant agents perceive a broader range of neighbors as in-group. In other words, it is more likely that emergent societies composed of less parochial guys maintain tag diversity. Generally speaking, when tolerance becomes homogenized at $T \geq 2$, cultural groups with different markers can evolve.

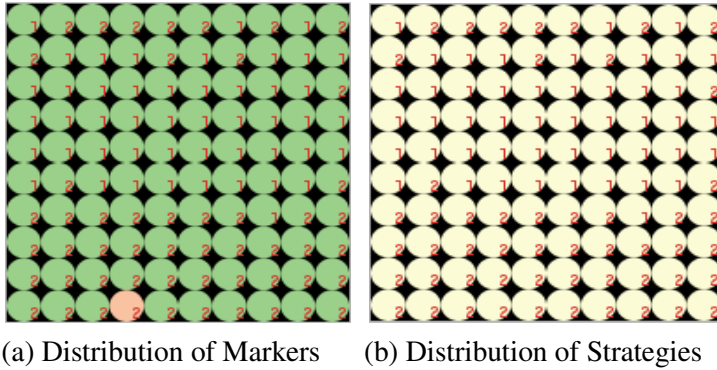


Figure 5: Nonaligned Groups under Two Levels of Tolerance in Model 1

Note: The number colored red in each agent indicates her tolerance level. Agents of the same color have the same tag list in (a). The brightness of each node in (b) is proportional to how many neighbors with whom she cooperates out of 8. The brighter, the more cooperation. $T = 1$ and $T = 2$ in (b). $b/c = 4$, $L = 4$, and $\mu = .00$.

The emergent society in Figure 4 indicates that universal cooperation does not necessitate the complete loss of tag diversity if agents in populations remain less parochial. A single cooperative phenotype constitutes it, however. Another pattern is the emergence of cooperative societies in which agents with different levels of tolerance co-exist (cf. ‘polymorphism’ in evolutionary biology). The society in Figure 5 represents this case where only one agent has a different list of tags (pink) in (a). Tag diversity does not completely disappear even in the population reaching universal cooperation because agents with $T = 2$ accept neighbors with the same markers except in one dimension as their own as in-group members. However, this pattern of group formation is very rarely observed.

Dynamics of Tag-based Cooperation under Mutation in Model 1

We present the dynamics of tag-based cooperation and its evolutionary stability in the presence of 1% mutation with snapshots from (a) to (d) in Figure 6. At the beginning (a), an almost equal number of agents (100 / 7) have each level of tolerance, from 0 through 6, given the tag length $L = 5$. For the first few generations, there is a transient state in which within-group cooperation co-exists with between-group discrimination with a significant amount of tag diversity. However, agents become less tolerant in increasingly homogeneous environments. In other words, dominant groups displaying stronger parochialism emerge as the number of clusters with different markers decreases over time. The level of global cooperation goes up in spite of a continuous loss of the average tolerance because cooperation occurs among more homogeneous agents. In (b), agents with $T = 1$ are dominant who cooperate only toward identical neighbors, but the emerging society can reach universal cooperation in completely homogeneous populations. In (c), even the most discriminating cooperative agents fail to maintain their cooperative society in the face of mutant defectors who display exactly the same tags as their own. This invasion by indistinguishable mutant defectors is followed by a downward spiral of cooperation. Because defectors do not benefit from each other, global cooperation can evolve once parochial cooperators happen to cluster together (‘network reciprocity’) in (d).

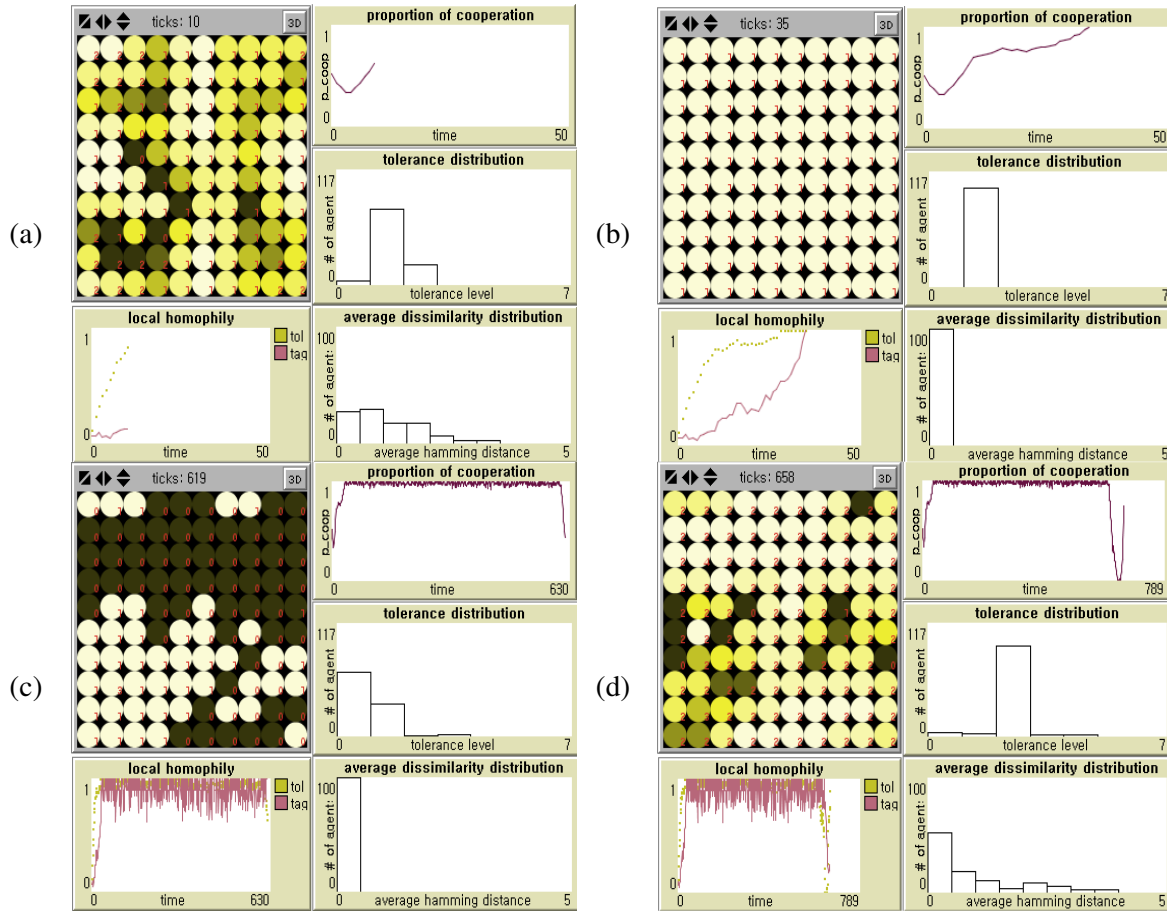


Figure 6: Dynamics of Tag-based Cooperation in Model 1

Note: A typical run of the simulation. The number in each agent indicates her tolerance level. The brightness of each node is proportional to how many neighbors with whom she cooperates out of 8. The brighter, the more cooperation. $b/c = 4$, $L = 6$, and $\mu = .01$.

Generally, parochial cooperators recognize each other by sending and receiving signals of trustworthiness. In the present study, tags as observable markers play the same role in partner identification for discriminators. Once emerging defectors mimic signaling among those discriminators (equivalent to mutant defectors with tolerably similar markers in our model), “secret handshaking” (Robson 1990) loses its advantages: a self-enforcing belief is broken that helping similar partners produces a differential share of the benefits of cooperation. Parochial cooperators must establish new signals (See Skyrms 2004; Traulsen and Nowak 2007). In the presence of mutation, this competition leads to endless spirals of global cooperation and global betrayal without reaching absorbing states.

Spatial Distribution of Tags and Tolerance in Model 2 and 3

Our findings so far identify that emergent cooperative societies with completely mutable tags (Model 1) are extremely parochial and highly homogeneous. Under the strong force of homogenization, tagging rarely acts as self-enforcing labeling enough to stabilize the

segregation of emerging enclaves. Given increased cultural heterogeneity in caste and modern societies with one immutable marker per agent, how differently is parochialism spatially distributed? And, how are cultural groups linked to each other in emergent cooperative societies? As in Model 1 (Figure 5), polymorphism (i.e. the co-existence of more than two cooperative phenotypes) does not frequently occur in caste and modern societies, but we observe that parochial cooperation stops at boundaries of heterogeneous cultural groups.

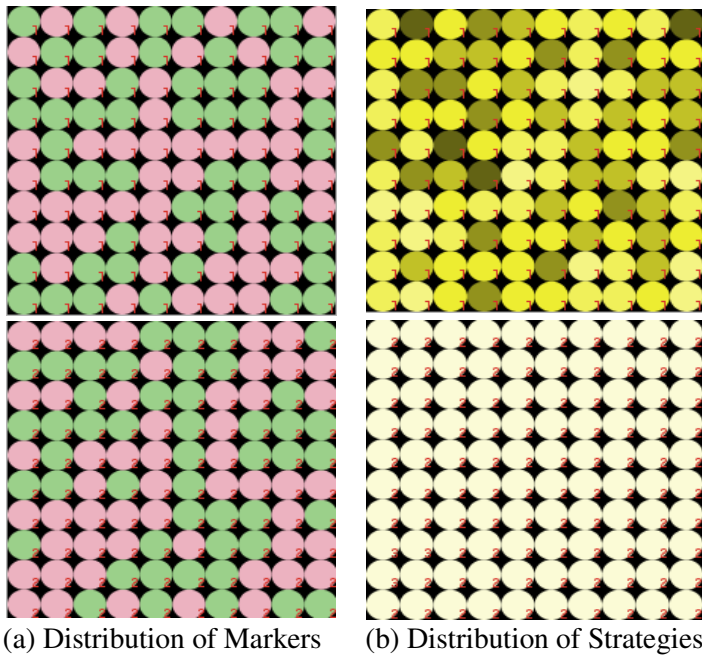


Figure 7: Nonaligned Groups under a Single Level of Tolerance in Model 2

Note: The number colored red in each agent indicates her tolerance level. Agents of the same color have the same tag list in (a). The brightness of each node in (b) is proportional to how many neighbors with whom she cooperates out of 8. The brighter, the more cooperation. $T = 1$ for all agents (Upper) and $T = 2$ for all (Lower). $b/c = 4$, $L = 4$, and $\mu = .00$.

The upper row of Figure 7 represents a typical example of emergent cultural groups in caste societies. The society purely consists of agents displaying the strongest parochialism, but it reaches an intermediate equilibrium level of global cooperation (.485) due to increased cultural heterogeneity. The average perceived social distance is .515. All members have the tag list [1000] in one cultural group (green), while [0000] in the other (pink). Given this persistent heterogeneity, the emergent society does not reach universal cooperation because all agents are conditional cooperators displaying the strongest degree of parochialism. The emergence of multiple cultural groups much more frequently occurs in Model 2 than in Model 1. The lower row of Figure 7 presents another case of cooperative caste societies divided into multiple groups. The population is composed of two groups: one with [0011] (green) and the other with [1011] (pink). As in the upper case, cultural

markers in all other dimensions are converged except those in the first dimension. However, the level of global cooperation at equilibrium is 1 because agents are more tolerant of diversity ($T = 2$ for all agents). This pattern of the emergence of multiple groups is similar to the case in Figure 4, but it is much more likely to happen in caste societies. More importantly, cultural groups in Model 2 are not necessarily highly segregated in contrast to groups in Figure 4.

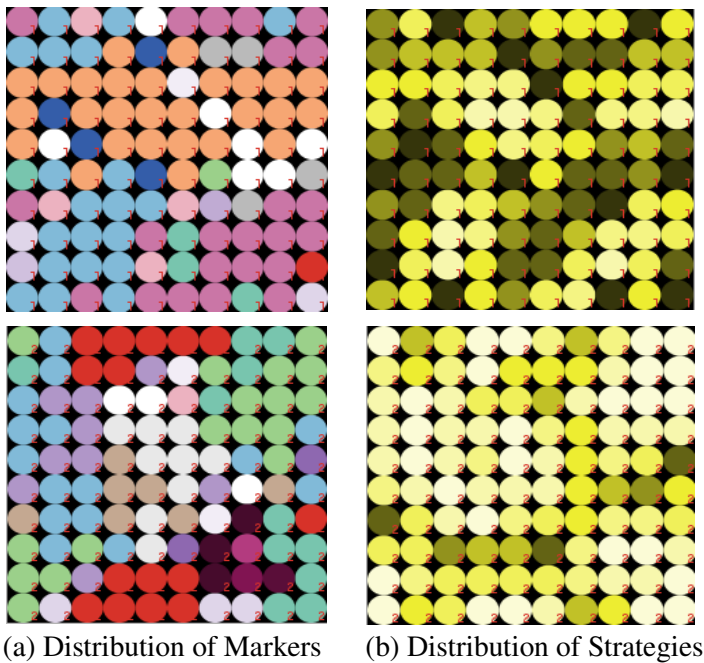


Figure 8: Nonaligned Groups under a Single Level of Tolerance in Model 3

Note: The number colored red in each agent indicates her tolerance level. Agents of the same color have the same tag list in (a). The brightness of each node in (b) is proportional to how many neighbors with whom she cooperates out of 8. The brighter, the more cooperation. The level of global cooperation is .398 (Upper) and .745 (Lower). $T = 1$ for all agents (Upper) and $T = 2$ for all (Lower). $b/c = 4$, $L = 4$, and $\mu = .00$.

Most of time, a single group of cooperators evolve to dominate the population in both caste and modern societies. In modern societies (Figure 8), all agents have the same level of tolerance (i.e. $T = 1$ in the upper row and $T = 2$ in the lower row), but we find different patterns of the spatial configuration of markers and tolerance. Recall that non-negotiable markers in modern societies are distributed across different dimensions of the tag space. Consequently, cultural heterogeneity is not only high at the global level (.98 in the upper and 1.00 in the lower of Figure 8) but also more persistent at the local level. This is why it is more difficult for emergent cultural groups to expand their boundaries by assimilating other group members. While small-sized homogeneous social circles evolve, agents with more heterogeneous markers are not firmly affiliated with any of adjacent groups. Members at group boundaries, although they hold the same level of tolerance, continue to modify

their strategies toward those heterogeneous agents. It is not shown here, the cooperative equilibrium in modern societies is always dynamic, not static, in spite of the absence of imitation errors.

Dynamics of Tag-based Cooperation under Mutation in Model 2, 3, and 4

Recall that a broader range of mutant defectors can more easily demolish cooperative societies consisting of less parochial agents. Also, even the most discriminating cooperators are inevitably gullible to mutant defectors with exactly the same markers in tag-based societies given that tolerance is not observable. Are there mechanisms that may enable tag-based societies to stably maintain parochial cooperation in spite of persistently emerging defectors in the presence of mutation (imitation errors)? Before presenting the dynamics of cooperation in caste and modern societies, we first test the idea in Hales (2004) and Edmonds and Hales (2005): “tags must mutate faster than strategies” for tag-based societies to achieve high and stable cooperation.

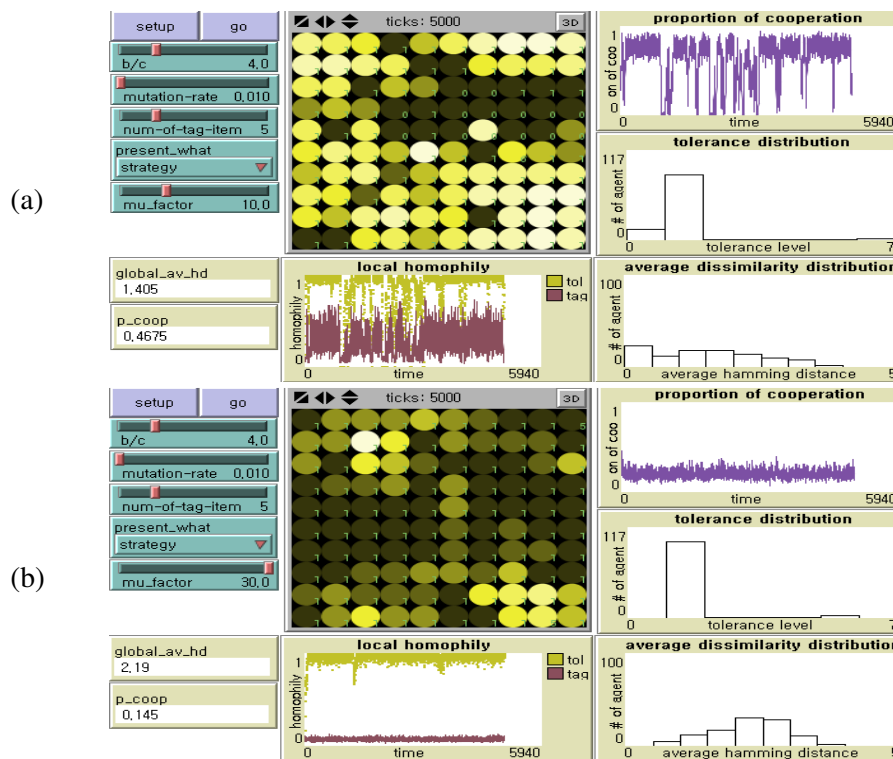


Figure 9: Dynamics of Tag-based Cooperation across Mutation Factors in Model 4

Note: The brightness of each node is proportional to how many neighbors with whom she cooperates out of the total number of neighbors. The brighter, the more cooperation. $b/c = 4$, $L = 5$, and $\mu = .01$. $f = 10$ (a); $f = 30$ (b).

Figure 9 shows the dynamics of cooperation across different levels of the mutation factor f when the fixed mutation rate $\mu = 0.01$. At $f = 10$, an emergent society is not that robust in the presence of indistinguishable mutant defectors in (a). It is not shown here, but

we observe cooperation evolve more stably, but sometimes with cascades of non-cooperation at $f = 20$. If tags mutate 30 times faster than tolerance, cooperation becomes highly robust in (b). Notice that the level of global cooperation declines as tags mutate faster than tolerance in Model 4. In other words, the trade-off relationship between the efficiency of tag-based cooperation and its stability depends on the ratio of how fast tags mutate to how fast tolerance does.

Table 10: Population Characteristics in Model 4

Variables	f	$b/c = 2$	$b/c = 4$	$b/c = 6$	Post-hoc Test (Scheffé)
(a) Average cooperation	5	.727 (.291)	.721 (.278)	.742 (.261)	
	10	.673 (.201)	.675 (.194)	.676 (.189)	
	15	.573 (.148)	.575 (.146)	.581 (.136)	
	20	.459 (.112)	.466 (.103)	.469 (.100)	2-4*2-6**
	25	.302 (.074)	.310 (.074)	.312 (.073)	2-4* 2-6**
	30	.197 (.043)	.201 (.043)	.202 (.045)	2-4** 2-6**
(b) Average tolerance	5	.878 (.313)	.902 (.329)	.929 (.318)	2-6**
	10	.943 (.228)	.952 (.239)	.960 (.244)	
	15	.978 (.168)	.982 (.186)	1.001 (.191)	2-6** 4-6*
	20	1.004 (.125)	1.017 (.126)	1.022 (.134)	2-4** 2-6**
	25	1.022 (.090)	1.031 (.101)	1.039 (.109)	2-4** 2-6**
	30	1.023 (.086)	1.034 (.096)	1.045 (.115)	All pairs **
(c) Average link homophily in tolerance	5	.841 (.268)	.834 (.252)	.852 (.230)	
	10	.907 (.171)	.901 (.172)	.900 (.167)	
	15	.938 (.111)	.928 (.123)	.929 (.116)	
	20	.956 (.080)	.954 (.079)	.951 (.081)	2-6*
	25	.967 (.060)	.964 (.066)	.961 (.069)	2-6*
	30	.968 (.060)	.964 (.065)	.960 (.070)	All pairs **
(d) Average perceived distance	5	.562 (.541)	.541 (.502)	.495 (.449)	2-6**
	10	.783 (.416)	.767 (.395)	.755 (.379)	
	15	1.074 (.337)	1.056 (.325)	1.046 (.306)	2-6**
	20	1.396 (.285)	1.379 (.267)	1.369 (.262)	2-4** 2-6**
	25	1.815 (.209)	1.797 (.211)	1.797 (.210)	2-4** 2-6**
	30	2.088 (.130)	2.082 (.130)	2.083 (.130)	2-4** 2-6**
(e) Average link homophily in tags	5	.471 (.223)	.464 (.221)	.477 (.214)	
	10	.269 (.129)	.270 (.127)	.271 (.127)	
	15	.145 (.069)	.146 (.069)	.147 (.069)	
	20	.085 (.033)	.086 (.033)	.086 (.033)	2-6*
	25	.063 (.014)	.063 (.015)	.063 (.015)	2-6*
	30	.066 (.014)	.066 (.014)	.066 (.014)	2-4*

Note: Results from Experiments in Table 6. $L = 5$ and $\mu = .01$. The number (in the parenthesis) at each cell indicates the average of each measure over time from 0 until 5,000 steps (and the averaged value of the standard deviation of each measure over time, not the standard deviation of the average of each measure). * $p < .05$; ** $p < .01$. In the last column, for example, 2-4** indicates the significant mean difference between $b/c = 2$ and $b/c = 4$ at $\alpha = .01$.

In Table 10, we present the characteristic levels of the average cooperation, the average tolerance, the average social distance (global), the average link homophily in tolerance, and the average link homophily in tags under varying conditions of the mutation factor and the b/c ratio. As tags mutate faster than tolerance when controlling for the b/c ratio, emerging societies are less likely to reach high levels of cooperation in terms of efficiency in (a). This is because societies are still based on the strongest parochialism ($T = 1$), as is displayed in (b), in spite of increased tag heterogeneity at the global and local level in both (d) and (e). The measure of local homophily in tolerance in (c) indicates that the degree of spatial homogeneity of parochialism increases as tags mutate faster than tolerance. In other words, parochial cooperators can develop new signals by changing tags rapidly against mutant defectors who consistently invade existing cooperative clusters.

There are a couple of problems with the application of the idea that tags must mutate faster than tolerance. First, it is reasonable to assume that tags are completely mutable in genetic inheritance. From the perspective of cultural evolution of cooperation, tags as cultural markers are not necessarily completely imitable. Second, the circumstances should be explained under which one genetic string of tags can mutate faster than another string of strategies. Relatedly, if tags are regarded as cultural markers, it should be justified why errors in imitating tags occur more frequently than errors in adopting tolerance.

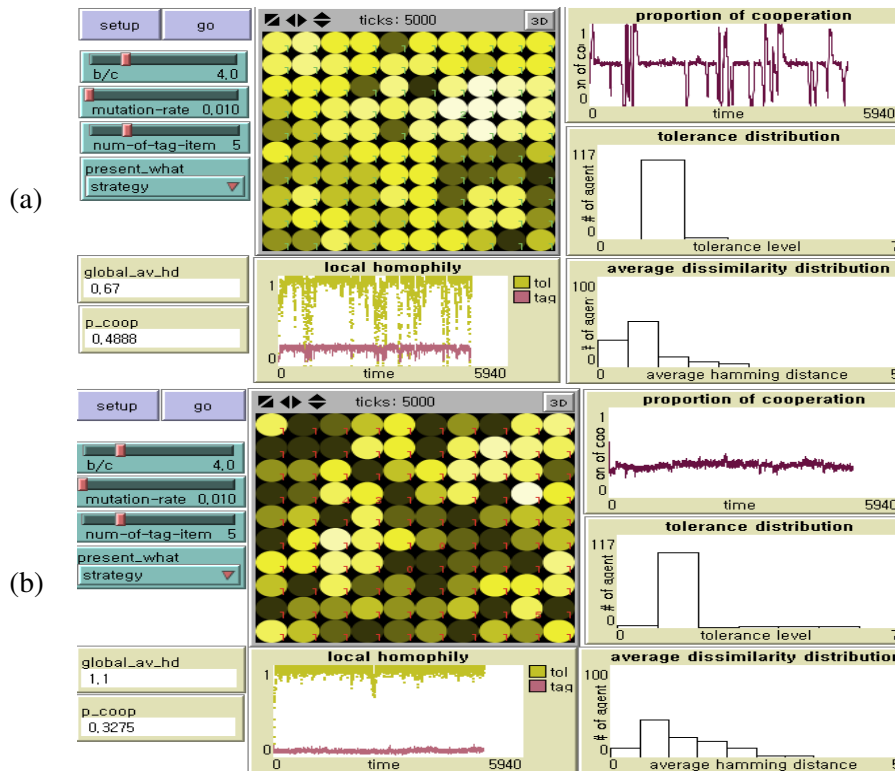


Figure 10: Dynamics of Tag-based Cooperation in Model 2, 3

Note: $b/c = 4$, $L = 5$, and $\mu = .01$. Model 2 (a); Model 3 (b). The brightness of each node is proportional to how many neighbors with whom she cooperates out of the total number of neighbors. The brighter, the more cooperation.

Next, we compare the dynamics of cooperation in caste societies to the dynamics in modern societies in search for alternative tag-based societies in which cooperation is not only more efficient but also stable. There is no significant difference in the possibility of cooperation survival between the two societies (See Table 7), but we find fundamental differences in evolutionary patterns of cooperation under 1% of imitation errors (Figure 10). At $b/c = 4$ and $L = 5$, caste societies in (a) very frequently undergo downward spirals of cooperation followed by its re-emergence, which is similar to the pattern in societies with completely mutable tags seen in Figure 6. On the contrary, given the same conditions, tag-based cooperation in modern societies is highly stable, as in (b). It is not shown here, but parochial cooperation in modern societies is robust against defectors with tolerably similar markers within a broader range of the b/c ratios (from 2 through 6) and the tag lengths (from 2 through 10).

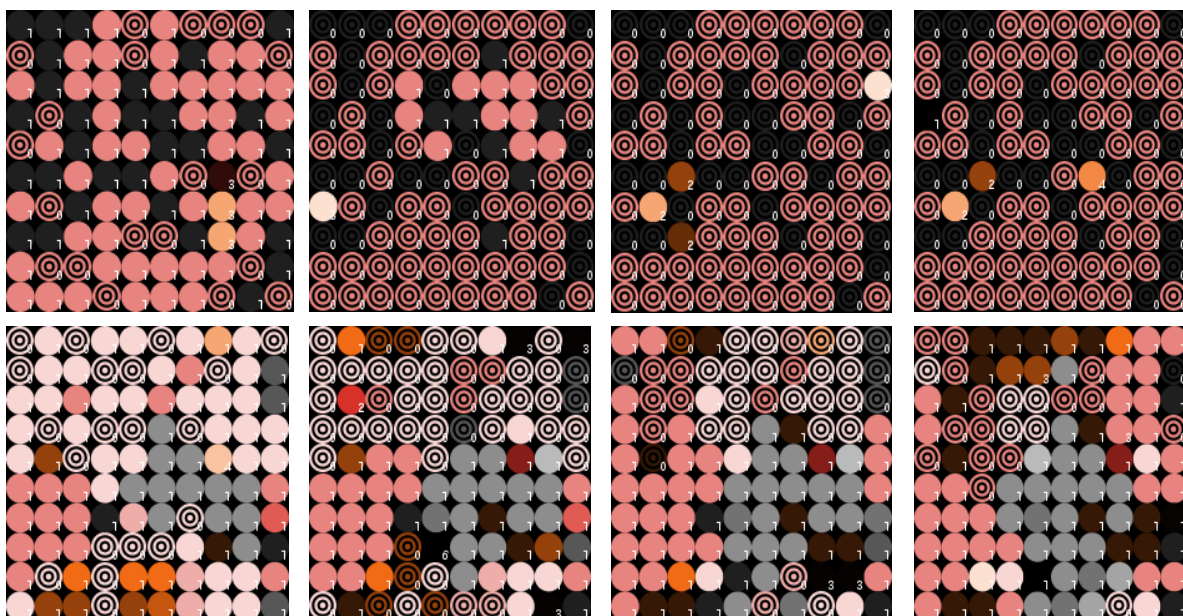


Figure 11: Robustness of Cooperation in Model 2, 3 under Random Attack by Defectors

Note: $b/c = 4$, $L = 5$, and $\mu = .01$. Model 2 (Top); Model 3 (Bottom). Agents of the same color have the same tag list. The number colored red in each agent indicates her tolerance level. Agents with $T = 0$ are shaped by ‘target.’ From the left to the right, time $t = 400$, $t = 405$, $t = 410$ and $t = 415$.

We develop additional experiment in order to further investigate why modern societies are relatively stable than caste societies in the presence of tolerably similar defectors. Instead of the invasion by a very small number of mutant defectors with arbitrary tags, we allow 20% of the total population to have the minimum tolerance and also the same tags as those shared by parochial cooperators in the dominant tag cluster. These defectors randomly attack the majority group in cooperative societies at time $t = 400$. Snapshots of two typical runs of simulation from both societies are presented in Figure 11. Two groups emerge in a caste society at the top where pink-colored cooperators are dominant. At $t = 400$, 20 out of

them become deviants. Cooperation in the caste society tends to disintegrate as defectors spread not only inside the dominant group but also over the other tag group colored dark grey. In a modern society at the bottom, 20 out of light pink-colored cooperators begin to play the role of invader at $t = 400$. On the contrary, learning from defectors can hardly outpace the reinforcement of locally scattered cooperation through network reciprocity, given a considerable number of groups in the modern society.

Table 11: Population Characteristics in Model 2, 3

Variables	$b/c = 2$		$b/c = 4$		$b/c = 6$	
	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
(a) Average cooperation	.427 (.196)	.361 (.048)	.492 (.238)	.429 (.070)	.525 (.234)	.526 (.096)
(b) Average tolerance	.938 (.409)	1.022 (.125)	1.081 (.506)	1.186 (.216)	1.147 (.508)	1.478 (.273)
(c) Average link homophily in tolerance	.816 (.287)	.951 (.075)	.823 (.238)	.926 (.089)	.839 (.195)	.900 (.097)
(d) Average perceived distance	.688 (.255)	1.216 (.150)	.652 (.222)	1.172 (.139)	.637 (.209)	1.162 (.141)
(e) Average link homophily in tags	.180 (.046)	.069 (.015)	.189 (.040)	.072 (.015)	.190 (.037)	.076 (.015)

Note: Results from Experiments in Table 5. $L = 5$ and $\mu = .01$. The number (in the parenthesis) at each cell indicates the average of each measure over time from 0 until 5,000 steps (and the averaged value of the standard deviation of each measure over time). At $b/c = 6$, no mean difference in the average cooperation between Model 2 and Model 3. In the rest of the case, the mean differences in all five measures between the two models are significant at $\alpha = .01$.

Table 11 presents the demographic characteristics of both caste and modern societies. First, modern societies are less efficient in terms of the level of global cooperation than caste societies particularly when cooperation is relatively costly ($b/c = 2$ and $b/c = 4$) in (a). Second, most of time, the strongest parochialism ($T = 1$) is dominant in the population, but agents who are more tolerant of culturally different neighbors more frequently remain in modern societies than in caste societies in (b). Third, polymorphism does not very often occur in both societies, but the average degree of link homophily in tolerance tends to decrease as the benefit-to-cost ratio goes up in modern societies, which is not the case in caste societies, according to (c). This pattern implies that when cooperation becomes less costly, it is more likely in modern societies that weak discriminators are adjacent to strong discriminators. Consistently high degrees of the average link homophily in tolerance in (c)

also indicate that modern societies are characterized by significantly stable cooperative clusters. Fourth, the average Hamming Distance in the population in (d) shows that agents in modern societies feel more dissimilar to each other than agents in caste societies, when holding the tag length L constant. Lastly, (e) indicates that the average degree of local link homophily in tags is significantly lower in modern societies than that in caste societies.

Similarities and Differences between Model 3 and Model 4

Our experiments so far discover that both modern societies and societies with tags mutating faster than tolerance render parochial cooperation on the basis of similarity sustainable in the presence of mutant defectors with tolerably similar markers. Finally, we are interested in scrutinizing similarities and differences in the dynamics of cooperation between these two societies (Table 12).

Table 12: Population Characteristics in Model 3, 4

Variables	Societies with tags mutating faster than tolerance (Model 4)	Modern societies (Model 3)
(a) Average cooperation	.310 (.074) ⁺ .201 (.043) ⁺⁺	.429 (.070)
(b) Average tolerance	1.031 (.101) ⁺ 1.034 (.096) ⁺⁺	1.186 (.216)
(c) Average link homophily in tolerance	.964 (.066) ⁺ .964 (.065) ⁺⁺	.926 (.089)
(d) Average perceived distance	1.797 (.211) ⁺ 2.082 (.130) ⁺⁺	1.172 (.139)
(e) Average link homophily in tags	.063 (.015) ⁺ .066 (.014) ⁺⁺	.072 (.015)

Note: The number (in the parenthesis) at each cell indicates the average of each measure over time from 0 until 5,000 steps (and the averaged value of the standard deviation of each measure over time). The results in the second column are excerpted from Table 10. The last two levels of the mutation factor ($f = 25$ and 30) are selected for the sake of comparison, where cooperation is significantly stable (See Figure 9). The outcomes in the last column are from Table 11. $b/c = 4$, $L = 5$ and $\mu = .01$. ⁺ at the mutation factor $f = 25$ and ⁺⁺ at $f = 30$.

Both societies are very similar in the following aspects. First, the dominant group in emerging societies consists of those who cooperate only with neighbors with exactly the same tags as theirs ($T = 1$), according to (b). Second, there is no significant difference ($\alpha = .01$) in the averaged values of the standard deviation of the average cooperation across time from 0 to 5,000 steps: .074 at $f = 25$ in comparison to .070 in (a). Third, the average degrees of link homophily in tolerance are very high: .964 at $f = 25$ and $f = 30$ in comparison to .926 in (c). These last two aspects refer to the noticeable robustness of cooperation in both societies. Lastly, persistent tag heterogeneity common to both cases

enables emergent cooperative societies to be more stable in the presence of mutant defectors. As in (e), the average degrees of link homophily in tags are sufficiently low: .063 at $f = 25$ and .065 at $f = 30$ in comparison to .072.

However, we underline different mechanisms contributing to the robustness of cooperation in those societies. In societies with tags mutating faster than tolerance, existing cooperative clusters are continuously reorganized into new ones before mutant defectors displaying tolerably similar signals attack them. Members in new groups have new signals whereby they recognize each other. In modern societies, immutable elements of cultural markers are distributed along different dimensions of the tag space. Parochial cooperators do not need to develop new signals in the presence of those defectors. Even if discriminating cooperators in some clusters accept mutant defectors displaying tolerably similar markers as in-group members, it is less likely to happen to their neighboring clusters.

These disparate routes to stable parochial cooperation result in different levels of global heterogeneity and thereby global cooperation in both societies. The average degree of tag heterogeneity at the global level in societies with tags mutating faster is much higher than that in modern societies: 1.797 at $f = 25$ and 2.082 at $f = 30$ in comparison to 1.172 from (d). Consequently, modern societies reach significantly higher levels of cooperation (.429) than societies with faster mutating tags (.310 at $f = 25$ and .201 at $f = 30$), as is presented in (a). We conclude that it is only in modern societies that relatively high and also stable cooperation is achievable.

Discussion and Conclusions

In the present study, we first confirm that ‘network reciprocity’ can promote global cooperation when the benefit-to-cost ratio in the Prisoner’s Dilemma game is higher than the average number of local interaction partners (Model 1, 2, and 3 under no mutation at $L = 0$. Recall that Model 4 = Model 1 under no mutation). It is otherwise impossible for altruists (indiscriminate cooperators) to survive in the face of defectors even in structured populations. The introduction of discriminating egoists is expected to promote the evolution of cooperation, but if they cooperate only with in-group members, parochial cooperation will stop at group boundaries. How is then society-wide cooperation possible in a mixed population of altruists and ethnocentric egoists?

Consistent with Riolo et al. (2001), we find that if agents leave offspring in their neighborhood in proportion to fitness, the population becomes increasingly characterized by stronger ethnocentrism under selection pressure (Model 1, 2, and 3 under no mutation at $L \geq 1$). In other words, agents in emergent societies become more parochial to adapt themselves to increasingly homogenized environments. Meanwhile, not only altruists but also unconditional defectors – who are not introduced in Riolo et al.’s model – are weeded out. Since less tolerant agents become more successful than more tolerant ones as tag diversity disappears, agents who cooperate only with identical others finally become dominant in emergent cooperative societies. In spite of a continuous erosion of tolerance, high levels of global cooperation can be maintained in highly homogeneous populations (cf. ‘mechanical solidarity’ in a Durkheimian sense).

We find that emergent societies consisting of less parochial cooperators are easily disrupted by mutant defectors displaying similar markers (Model 1 under mutation). Even the strongest parochial cooperators are vulnerable to mutant defectors as long as they have exactly the same tags. In this way, the stability of parochial cooperation depends on whether or not a shared belief continues that markers predict each other's behavior, given that tolerance is not observable in tag-based societies. Once mutant defectors seriously perturb signaling among parochial cooperators, there occurs a downward spiral of cooperation that leads to societies of mutual betrayal. While there are no mutual benefits for agents in betrayal societies, once discriminating cooperators are clustered with each other to receive a differential share of the benefits of cooperation, they can establish new cooperative societies with new salient markers. These cycles proceed along with "the tide of tolerance" (Sigmund and Nowak 2001) and the formation and dissolution of tag clusters as cultural groups.

In search for alternative cooperative societies marked by society-wide and stable cooperation, we first apply the idea that tags should mutate faster than tolerance (Model 4 under mutation). It is expected that the faster tags mutate than tolerance, the more stable tag-based cooperation. In other words, as parochial cooperators change their signals (e.g. passwords) more frequently, the chance of the invasion by defectors (e.g. hacking passwords) will decrease. We find that if tags mutate faster but it is not sufficient, then emergent societies are yet frequently damaged by cascades of non-cooperation triggered by mutant defectors; and if tags mutate fast enough, parochial cooperation shows high robustness despite the fact that all elements of tags are completely mutable (easily mimicable).

Next, we consider tags and tolerance as culturally transmissible traits rather than as genetically inheritable ones. We construct two new societies, 'caste' societies and 'modern' societies (Model 2 and 3 under mutation). We discover that emergent caste and modern societies are highly parochial, as in societies with completely mutable tags: agents with the minimum tolerance for cooperation are still in the majority of populations although emergent societies tend to be more tolerant as cooperation becomes less costly. In the absence of imitation errors, there is no significant difference in the likelihood of the survival of cooperation in both caste and modern societies. These two societies are, however, significantly different from societies with completely mutable tags. Given persistent heterogeneity of markers, within-group parochial cooperation evolves with between-group non-cooperation. In other words, tagging acts as self-enforcing labeling to make it difficult for tolerance to cross segregated groups with different markers. Finally but most importantly, cooperation is highly stable against mutant defectors displaying tolerably similar identities only in modern societies although modern societies are less parochial than caste societies on average.

We identify the conditions for the stable institutionalization of society-wide cooperation in tag-based societies. First, parochial cooperators should share a minimal number of markers that are difficult for defectors to imitate: the average dissimilarity at the global level should not be too low to obstruct the efficiency of cooperation. As the number of core markers increases, the level of global cooperation declines. In societies with faster mutating tags, parochial cooperators are able to develop new signals by changing tags

rapidly before mutant defectors invade existing cooperative clusters. However, if tags mutate faster than a threshold value of the ratio of tag mutation to tolerance mutation, emergent societies do not reach high levels of cooperation because too much tag heterogeneity is constantly introduced into the population. On the contrary, modern societies reach higher levels of global cooperation than societies where tags mutate faster than tolerance because they have lower degrees of cultural heterogeneity at the global level.

Second, parochial cooperators should have core markers in different dimensions of the cultural space: the average homophily in markers at the local level should be low enough to stabilize cooperation. We observe that although both caste and modern societies retain similar degrees of cultural diversity, cooperation is seriously vulnerable to mutant defectors only in caste societies. Given that only a few cultural groups emerge, caste societies tend to disintegrate when tolerably similar defectors randomly attack either the society as a whole or the dominant group. In modern societies, however, emerging cultural groups are manifold. Consequently, modern societies consisting of multiple small-sized groups loosely coupled with one another can be highly robust against emerging free-riders (cf. Bruggeman 2008: 89-93). Even if discriminators in some clusters accept mutant defectors displaying tolerably similar tags as in-group members, those in other clusters facilitate the restoration of cooperation in attacked clusters through network reciprocation. In this way, multiple cultural groups in modern societies tend to self-organize their group boundaries through 'tagging' even in the presence of cultural perturbation.

References

- Alexander, Richard D. 1987. *The biology of moral systems*. Hawthorne, N.Y.: Aldine de Gruyter.
- Axelrod, Robert. 1984. *The evolution of cooperation*. New York: Basic Books.
- Bowles, Samuel, and Herbert Gintis. 2004. Persistent parochialism: trust and exclusion in ethnic networks. *Journal of Economic Behavior and Organization* 55: 1-23.
- Boyd, Robert, and Peter J. Richerson. 1989. The evolution of indirect reciprocity. *Social Networks* 11: 213-36.
- Bruggeman, Jeroen. 2008. *Social networks: an introduction*, London: Routledge.
- Dawkins, Richard. 1976. *The selfish gene*, Oxford: Oxford University Press.
- Edmonds, Bruce, and David Hales. 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>
- _____. 2005. Computational simulation as theoretical experiment. *Journal of Mathematical Sociology* 29: 209-32.
- Efferson, Charles, Rafael Lalive, and Ernst Fehr. 2008. The coevolution of cultural groups and ingroup favoritism. *Science* 321: 1844-9.
- Eshel, Ilan, and L. L. Cavalli-Sforza. 1982. Assortment of encounters and evolution of cooperativeness. *Proceedings of the U.S. National Academy of Sciences* 79: 1331-5.

- Grim, Patrick, Stephanie Wardach, and Vincent Beltrani. 2006. Location, location, location: The importance of spatialization in modeling cooperation and communication. *Interaction Studies* 7(1): 43-78.
- Hales, David. 2004. Understanding tag systems by comparing tag models. *Second Model-to-Model Workshop (M2M2)* at the Second European Social Simulation Association Conference (ESSA'04) Valladolid, Spain 16-19th Sept 2004.
- Hamilton, William D. 1964. The genetical evolution of social behavior. *Journal of Theoretical Biology* 7: 17-52.
- Hammond, Ross A., and Robert Axelrod. 2006. The evolution of ethnocentrism. *Journal of Conflict Resolution* 50(6): 926-36.
- Heylighen, Francis, and Donald T. Campbell. 1995. Selection of organization at the social level: obstacles and facilitators of metasystem transitions. *World Futures: the Journal of General Evolution* 45: 181-212.
- Holland, John H. 1993. The effect of labels (tags) on social interactions. *SFI Working Paper 93-10-064*. Santa Fe Institute, Santa Fe, NM.
- _____. 1995. *Hidden order: How adaptation builds complexity*, Reading, Mass.: Addison-Wesley.
- Kim, Jae-Woo. 2010. A tag-based evolutionary Prisoner's Dilemma game on networks with different topologies. *Journal of Artificial Societies and Social Simulation* 13(3)2. <http://jasss.soc.surrey.ac.uk/13/3/2.html>
- Krebs, Dennis. 1987. The challenge of altruism in biology and psychology. In *Sociobiology and psychology: Ideas, issues, and applications*. edited by C. Crawford, M. Smith, and D. Krebs. Hillsdale, N.J.: Erlbaum. pp. 81-118.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lopreato, Joseph. 2001. Sociobiological thinking. In *Handbook of sociological theory*, edited by J. H. Turner. New York: Kluwer Academic/ Plenum Publishers. pp. 405-33.
- McElreath, Richard, Robert Boyd, and Peter J. Richerson. 2003. Shared norms and the evolution of ethnic markers. *Current Anthropology* 44(1): 122-9.
- Miller, John H., Carter T. Butts, and David Rode. 2002. Communication and cooperation. *Journal of Economic Behavior and Organization* 47: 179-95.
- Nettle, Daniel, and Robin I. M. Dunbar. 1997. Social markers and the evolution of reciprocal exchange. *Current Anthropology* 38(1): 93-9.
- Nowak, Martin A. 2006. Five rules for the evolution of cooperation. *Science* 314: 1560-3.
- Nowak, Martin A., and Robert M. May. 1992. Evolutionary games and spatial chaos. *Nature* 359: 826-9.
- Nowak, Martin A., and Karl Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393: 573-7.
- _____. 2005. Evolution of indirect reciprocity. *Nature* 437: 1291-8.
- Riolo, Rick L., Michael D. Cohen, and Robert Axelrod. 2001. Evolution of cooperation without reciprocity. *Nature* 414: 441-3.
- Robson, Arthur J. 1990. Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *Journal of Theoretical Biology* 144(3): 379-96.
- Sigmund, Karl, and Martin A. Nowak. 2001. Tides of tolerance. *Nature* 414: 403-4.

- Skyrms, Brian. 2004. *The stag hunt and the evolution of social structure*. Cambridge; New York: Cambridge University Press.
- Tajfel, Henri. 1974. Social identity and intergroup behavior. *Social Science Information* 13(2) : 65-93.
- Traulsen, Arne, and Martin A. Nowak. 2007. Chromodynamics of cooperation in finite populations. *PLoS ONE* 2(3): e270.
- Trivers, Robert L. 1971. The evolution of reciprocal altruism. *The Quarterly Review of Biology* 46(1): 35-57.
- Turner, John C. 1982. Towards a cognitive redefinition of the social group. In *Social identity and intergroup relations*, edited by H. Tajfel. Cambridge: Cambridge University Press. pp.15-40.
- van der Berghe, Pierre L. 1981. *The ethnic phenomenon*. New York: Elsevier.
- Wilensky, Uri. 1999. NetLogo. Evanston, IL. <http://ccl.northwestern.edu/netlogo>