ELSEVIER

Journal of Economic Behavior & Organization
Vol. 57 (2005) 267–286

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

Can stable social groups be maintained by homophilous imitation alone?

Richard Durrett[a], Simon A. Levin[b,*]

[a] *Department of Mathematics, Cornell University, Ithaca, NY 14853, USA*
[b] *Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544-1003, USA*

Received 9 June 2003; accepted 29 September 2003
Available online 15 December 2004

Abstract

A central problem in the biological and social sciences concerns the conditions required for emergence and maintenance of cooperation among unrelated individuals. Most models and experiments have been pursued in a game–theoretic context and involve reward or punishment. Here, we show that such payoffs are unnecessary, and that stable social groups can sometimes be maintained provided simply that agents are more likely to imitate others who are like them (homophily). In contrast to other studies, to sustain multiple types we need not impose the restriction that agents also choose to make their opinions different from those in other groups.

JEL classification: C0; C7; D7

Keywords: Social norms; Mathematical model; Cooperation; Group formation; Group dynamics; Homophily

1. Introduction

Evolutionary ecology has made great advances in the understanding of animal behavior, building on the fundamental assumption that observed behaviors are adaptations to environ-

* Corresponding author. Tel.: +1 609 258 6880; fax: +1 609 258 6819.
E-mail addresses: rtd1@cornell.edu (R. Durrett), slevin@princeton.edu (S.A. Levin).

0167-2681/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2003.09.017

mental conditions. The classical tool for exploring such questions was optimization (West et al., 1997); optimization approaches, however, examine only part of the story. More generally, the evolution of behaviors occurs within a frequency-dependent context, in which game theory provides the more appropriate formalism (Lewontin, 1961; Maynard Smith, 1974, 1982; Oster and Wilson, 1978). In particular, game–theoretic approaches are essential to understanding the evolution of individual interactions, and especially of others regarding preferences (or behavior) (Hoffman et al., 1994).

Of special interest is the evolution of cooperation and collective action. Altruistic behavior was a puzzle even for Darwin (1859). Why indeed should individuals sacrifice their own fitness to aid others when a naive view would suggest that natural selection should weed out such behaviors? Haldane (1948) captured the essence of the answer in his off-cited quip that he would sacrifice his own life for two siblings or for eight cousins; the nature of the relatedness between siblings (1/2) or cousins (1/8) is such that the contribution of genes to future generations just balances in Haldane's equation. Hamilton (1964) formalized this in his classic development of the notion of extended fitness, in which one's own personal fitness is supplemented by the effects of one's actions on relatives discounted appropriately according to their relatedness. In particular, for the haplodiploid insects, males arise from unfertilized eggs; hence, sisters are related by 3/4. In such circumstances, therefore, selection for altruistic behavior is particularly effective, and thus the haplodiploid insects provide the strongest examples of eusociality.

However, the extension of these explanations to human societies clearly is still inadequate to explain altruism, for example, the willingness to give to charity, to serve in the military, or in general to subscribe to the norms of society. How are behaviors sustained that make societies work, but clearly are not in the individual's own genetic interest? Understanding what the relevant forces are is crucial not only to sustaining norms that are in the common good, but that also to efforts to overcome destructive norms such as racism and overconsumption (Veblen, 1902). To this end, it is crucial to recognize that individual behaviors that adhere to norms may not be adaptive; rather, they may represent specific realizations or extensions of more general behavioral syndromes that were, or indeed still are, adaptive under other conditions. Thus, Simon (1990) explains the high degree of loyalty that individuals may show, say, to the companies for which they labor as vestigial behavior from earlier hierarchical societies. More generally, we follow individual or collective rules of conduct that simplify our responses to a variety of situations, without necessarily making a separate calculation in each instance. Bicchieri writes "Once a norm is established in a given context, people will tend to apply it to all contexts that are perceived as relevantly similar to the original one as a 'default rule' instead of making complex cost-benefit calculations" (Bicchieri, 1997, p. 18). In other situations, where cost-benefit calculations might lead individuals to deviate from normative behaviors, societies impose a range of penalties to make such independence costly (Henrich et al., 2001).[1]

The evolution of social behavior occurs on a range of time scales, and cultural evolution in general occurs on a much more rapid time scale than genetic evolution. Cavalli-Sforza and Feldman (1981) have developed an elegant and powerful theory of coupled gene–culture

[1] We thank Sam Bowles for bringing our attention here to the writings of Bernard Mandeville, who emphasized the compatibility of private vices and public benefits.

evolution, but for many of the traits of interest, the genetic and cultural time scales can be separated fairly cleanly. Over long time scales, the general rules of individual behavior are shaped; on shorter time scales, those general rules largely determine how individuals will respond to particular situations and how cultural evolution emerges, often with similar patterns appearing in diverse environments. For example, no one is born with a gene that determines what religious beliefs they will have; that is the result of accidents of birth and cultural context and of individual decisions. However, the tendency for some form of religion to emerge in so many societies suggests that more general evolutionary syndromes facilitate the emergence of formal religions through a process of cultural evolution.

Economists and other social scientists debate what the basic behavioral rules are that shape the emergence of collective dynamics. Are individuals, as much classical economic theory would suggest, members of *Homo economicus*, perfectly or nearly perfectly rational, knowledgeable, and operating in their own self-interest, or are their decisions somehow largely imposed by society, reflecting general syndromes whose adaptive value might not apply to each situation. What role do emotions such as guilt, shame and pride play? Clearly, individuals are both selfish and social animals, and an understanding of human behaviors must address this (Ehrlich, 2000). In particular, we need to explicate how social norms emerge and are sustained, how individuals come to form groups for collective action, and how those groups are sustained and in turn influence individual decisions.

The task laid out in the opening paragraphs is a daunting one, and indeed a rich theory has begun to develop in recent years (Boyd and Richerson, 1985, 1996; Bowles, 2001; Henrich, 2004; Skyrms, 1996; Young, 1993). Ultimately, one must recognize that particular norms evolve within broader systems of justice, for example, and that individual behaviors are indeed the complex creations of their owners who must extrapolate from experiences, imitation and instruction to create personal rules of conduct. In this paper, however, we take only the simplest first steps, endeavoring to understand what sorts of patterns can emerge from imitation only. Indeed, we show that imitation is sufficient to lead to the creation of stable social groups that then can create the context for the institution of collective rules and behaviors.

In human societies, groups (political parties, religions and special interest groups) form, consisting of people who share similar opinions, and they remain reasonably stable over time. One idea that has been extensively studied is that cooperating groups can emerge when individuals help those that are like themselves. Dawkins (1976) introduced the ‘green beard effect’ as a thought experiment in sociobiology. Consider a gene that confers on its bearer not only a green beard, but also the instinct to provide assistance to all other owners of a green beard. Individuals with such a gene would form a self-serving clique, so the gene would spread within the population.

Nowak and Sigmund (1998) and Riolo et al. (2001) have studied similar systems in which a population of agents meet randomly as potential givers and receivers of help. Giving help entails some cost to the donor, but getting help provides a larger benefit to the recipient. If individuals are indistinguishable, cheaters who refrain from helping incur no costs and grow at the fastest rate, eliminating cooperators. However, as cited, if individuals can guess whether recipients are likely to give assistance in return, cooperation based on reciprocation can emerge. Leimar and Hammerstein (2001) argue that the more robust results occur when individuals aim for "good standing" (Sugden, 1986).

In the interactions mentioned in the last two paragraphs, cooperating individuals receive rewards. This feature is also involved in studies of cultural transmission (see e.g., Aoki, 2001) where learned behavior can give the individual a higher fitness. In contrast to these studies we investigate here a model without rewards or fitnesses, in which individuals preferentially imitate others who are similar to them. Our basic question asks what copying mechanisms can lead to the formation of stable groups. Our study is related to the study of memes, behaviors and ideas copied from person to person by imitation, (see e.g., Blackmore, 2000). However, we concentrate on the formation of groups rather than on the propagation of information. Our focus on what can result from imitation alone, of course, is not to deny that there are rewards from cooperation, but simply to make clear that imitation alone can lead to group formation, which in turn may serve as a matrix for cooperation. In real situations, the processes of imitation and cooperation are likely to be complementary and synergistic, strengthening each other as forces for the emergence and maintenance of groups and norms.

2. Models of imitation

The classical model of imitation is the voter model of Holley and Liggett (1975), in which an individual has no strongly held opinions and where at each time step, an individual chosen at random changes his or her opinion to match that of a randomly chosen neighbor, unless they are already in agreement. In the *majority voter model*, the individual takes a poll of all neighbors in a neighborhood, adopting the view of the majority. All interactions take place on a lattice that is stretched onto a torus so that everyone has the same number of neighbors. If the lattice is two-dimensional, eventually everyone has the same opinion.

The voter model is our starting point, but we will modify it in a variety of ways to reflect social processes more correctly. For the purpose of this paper, we will assume still that all interactions take place on a two-dimensional lattice with its edges identified to avoid boundary effects. Ultimately it will be crucial to consider more complex webs of interaction, such as small-world networks (Watts and Strogatz, 1998), possibly with time-varying connectivities. Furthermore, in real societies, power structures arise and are crucial to the spread of influence, so pairwise interactions need not be symmetric or homogeneous. Networks are not in general fixed, but can themselve evolve. As groups form, they may find ways to discourage members from interacting with members of other groups or from paying much attention to them. One way to begin to consider such influences is to allow individuals to have several bits of information, such as their group affiliation(s) and one or more bits that describe their opinions on various topics.

To be specific, in the models that we will consider in the rest of this paper, an individual will be represented by a string of k binary bits, the jth bit indicating that person's opinion on that particular issue. Our first and simplest model does not have group affiliations. Its dynamics are based on the notion of homophily: individuals are more likely to be influenced by those whose opinions on other issues are similar. For the moment we imagine a homogeneously mixing population with N individuals and formulate the dynamics in continuous time as follows: at rate one, each individual in the population decides to update his or her beliefs, and picks at random one of the k issues, call it j, for possible change. That is, in a small time step h, the probability an individual will decide to update is $h + 0(h)$, where $0(h)$ is small with respect to h. To determine the possible new opinion, the focal individual (call him Fred) chooses an individual (call her Ethel) at random from the population. If Ethel's jth bit agrees with Fred's, nothing happens. If Ethel's jth bit is different from Fred's and i of her remaining $k - 1$ bits agree with Fred's, then Fred adopts her opinion with probability $c + (1 - c)i/(k - 1)$. Fig. 1 shows what happens in a population of 10,000 individuals if we start the three-bit model with 1/2 of the population all 1's and 1/2 of the population all 0's, and where $c = 0.1$. As the reader can see, the opinions very quickly become randomized and reach an equilibrium in which all 8 possible opinion strings have the same probability. This conclusion is generic for all $c > 0$; there is, of course, a discontinuity at $c = 0$, because there the all-0 and the all-1 states are absorbing.

One of the problems with the breakdown of the groups in this simple model is that the symmetry leads to a dissipation of group integrity. If Fred has opinion 111, chooses to update opinion 2, and finds that Ethel has opinion 101, then he is convinced to change. To try to identify opinions with specific groups, we will add a 0th bit to the string to identify the individual's group affiliation (e.g., 1 = Democrat and 0 = Republican). Having introduced this asymmetry, we use 1 and 0 to denote the opinions on the remaining issues (e.g., gun control, abortion, etc.) that are associated with the two groups. Issues labeled 1 are associated with group 1, and issues labeled 0 are associated with group 0. Of course, on longer time

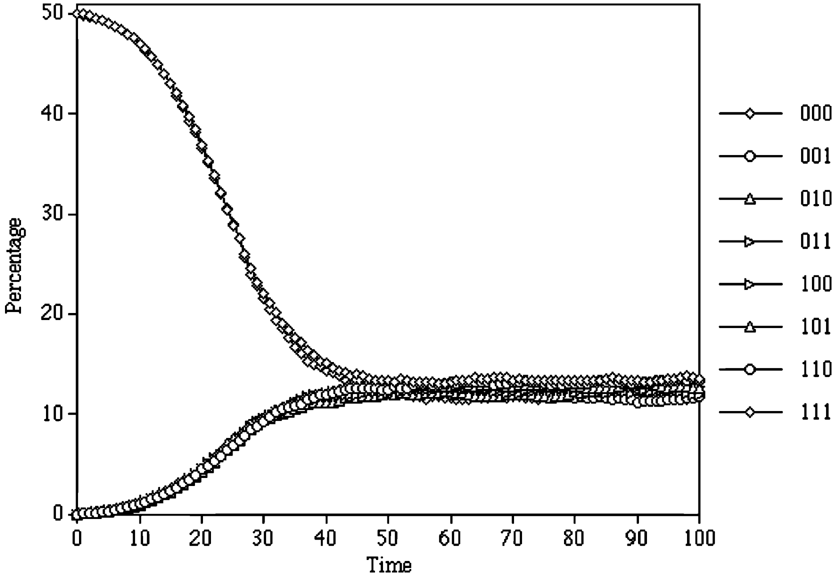

Fig. 1. Simulation of the three-bit model in which the probability of adopting a neighbor's opinion is a linear function of the number of bits that agree. In this case, opinions quickly become randomized.

scales, groups themselves will shift their normative attitudes, but we leave discussion of this for a future paper and do not permit such shifts here. As before, if Fred is a member of the population, then at rate one Fred decides to update his beliefs and picks one of the k bits at random, call it j, for possible change.

If $j = 0$ (i.e., the party bit is chosen), then Fred counts the number of his opinions about issues $j > 0$ that agree with the party line, and if that number is m, then he changes parties with probability q_m.

If $j > 0$, then Fred chooses an individual at random from the population, call her Ethel. As before if Ethel's jth bit agrees with Fred, no change occurs. Let $a = 1$, if Ethel's party is the same as Fred, and $a = 0$ otherwise. Let $b = 1$, if Fred imitating Ethel will bring that opinion back to his party line, and $b = 0$ otherwise. The probability Fred changes is $r_{a,b}$ where we assume:

$$r_{1,1} > r_{1,0} > r_{0,1} = r_{0,0}. \tag{1}$$

The first inequality here is certainly natural; it states that one is more easily convinced by a fellow party member to return to the party line than to defect from it. The equality in the final position may be regarded as providing a null model; it states that members of the other party are as likely to convince one to be faithful to the party line as to stray. The middle inequality simply reflects the likelihood that a fellow party member will be more persuasive in general than a member of the other party.

Fig. 2 shows a simulation of the model with two opinion bits, $q_0 = 1$, $q_1 = 0.55$, $q_2 = 0.1$, $r_{1,1} = 1.0$, $r_{1,0} = 0.5$, $r_{0,1} = r_{0,0} = 0.2$. The new scheme does a somewhat better job of stabilizing the groups. In the first 25 units of time the system reaches a quasi-stationary state that it maintains until about time 100, when the Republicans begin to take over the system. In the limit the system consists entirely of Republicans with state 000 and Democrats with state 100, who transform from Republicans by spontaneous party changes.

This inspires our third change in the model. We add introspective or idiosyncratic changes. Each individual and each of their opinions changes away from the party line at rate s and back toward the party line at rate σs. Fig. 3 shows the fraction of Republicans in simulations of the previous model for various values of s. If $s = 0$ we have the previous result. The frequency approaches a little over 90% in equilibrium. As s increases, the frequency in equilibrium decreases until, for sufficiently large s, a 50–50 mix is stable. The last result, polarization of society into two stable groups, has also been obtained by Macy et al. (2004) for a model that was inspired by Hopfield's (1982) neural nets.

3. Analytical results

The simulations discussed in the previous section suggest that when the rate of introspective or idiosyncratic changes is high, then the symmetric equilibrium is stable, but it loses its stability as s decreases. In this section we will derive analytical results to help clarify this picture. We restrict ourselves to the simplest possible case: one party bit and one opinion bit. Letting u, v, x, and y be the frequencies of 11, 10, 01, and 00 in the population and

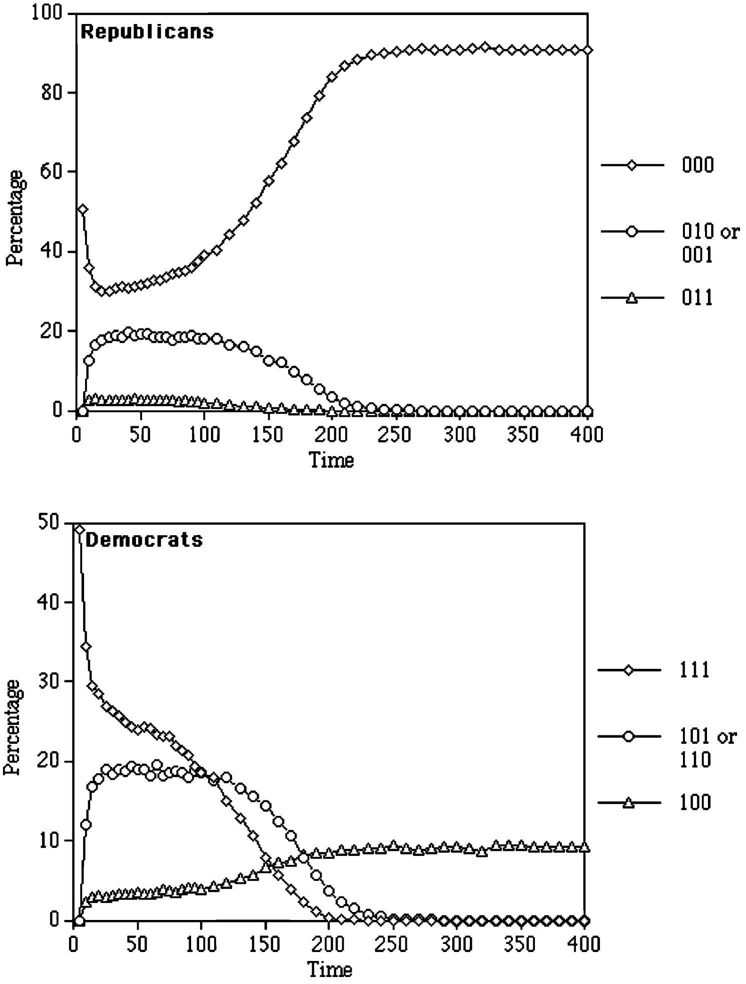

Fig. 2. Simulation of a three-bit model in which the first bit gives the person's political party. Individuals prefer to listen to people from their own party and to return toward the party line. Groups are somewhat stable, but eventually one party takes over.

letting the population size tend to infinity, we arrive at the following differential equations:

$$
\begin{aligned}
\frac{\mathrm{d}u}{\mathrm{d}t} &= -u(q_1 + s + r_{10}v + r_{00}y) + v(\sigma s + r_{11}u + r_{01}x) + xq_0 \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= -y(q_1 + s + r_{10}x + r_{00}u) + x(\sigma s + r_{11}y + r_{01}v) + vq_0 \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -v(q_0 + \sigma s + r_{11}u + r_{01}x) + u(s + r_{10}v + r_{00}y) + yq_1 \\
\frac{\mathrm{d}x}{\mathrm{d}t} &= -x(q_0 + \sigma s + r_{11}y + r_{01}v) + y(s + r_{10}x + r_{00}u) + uq_1.
\end{aligned}
\tag{2}
$$

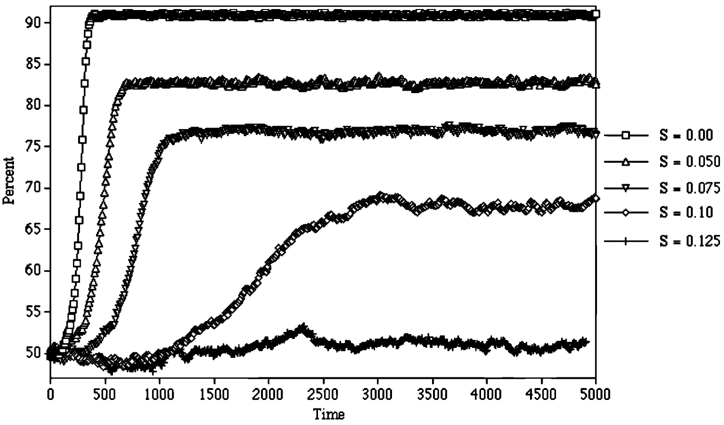

Fig. 3. A modification of the party model in which individuals make introspective changes (i.e., not based on a neighbor's opinion) that favor a return to the party position. If the rate of such changes is large enough, there are two stable parties of roughly equal size.

Furthermore, note that the equations are invariant under an interchange of u and y, and v and x. Recall from (1) that $r_{0,1} = r_{0,0} = r_0$ and write $r_1 = r_{1,1} - r_{1,0} > 0$.

Symmetry dictates that the line $u = y$, $v = x$, $u + v + x + y = 1$ is invariant. A little algebra (see Appendix A for these and other details) shows that there is a unique fixed point with $2u \in (0, 1)$ given by:

$$v_0 = \frac{-b - \sqrt{b^2 - 4ac}}{4a}, \qquad v_0 = \frac{1}{2} - v_0, \qquad y_0 = u_0, \qquad x_0 = v_0, \tag{3}$$

where a, b and c are given in Appendix A. A stability analysis of this fixed point shows that it is stable if

$$r_1(q_1u_0 - q_0v_0) + s(2q_1\sigma + 2q_0) > 0 \tag{4}$$

which holds if s is large enough.

When $s = 0$ the second term disappears and we are left to consider the sign of $q_1u_0 - q_0v_0$. This is positive if and only if:

$$\frac{q_0}{q_0 + q_1} < \frac{r_1 - 2r_0 + \sqrt{r_1^2 + 4r_0^2}}{2r_1} \tag{5}$$

(see Appendix A). In this case, the symmetric fixed point is stable even when $s = 0$. Conversely, if the inequality in (5) is reversed, the symmetric fixed point is unstable for small s.

To understand (5), consider first the null case $q_0 = q_1$, which means that an individual is equally likely to switch parties whether or not his or her opinion matches the party line. In this case (5) is always satisfied, and the symmetric fixed point is always stable. More realistically, however, $q_1 < q_0$; and as q_1 is decreased sufficiently, a threshold must be reached where the equilibrium is destabilized. That is, an increasing tendency to stick with a party if one agrees with its principles leads to the destabilization of the symmetric equilibrium.

To have concrete examples of the two alternatives, let $r_1 = 0.5$, $r_0 = 0.2$, and $q_0 = 0.1$. In this case,

$$\frac{r_1 - 2r_0 + \sqrt{r_1^2 + 4r_0^2}}{2r_1} = 0.740.$$

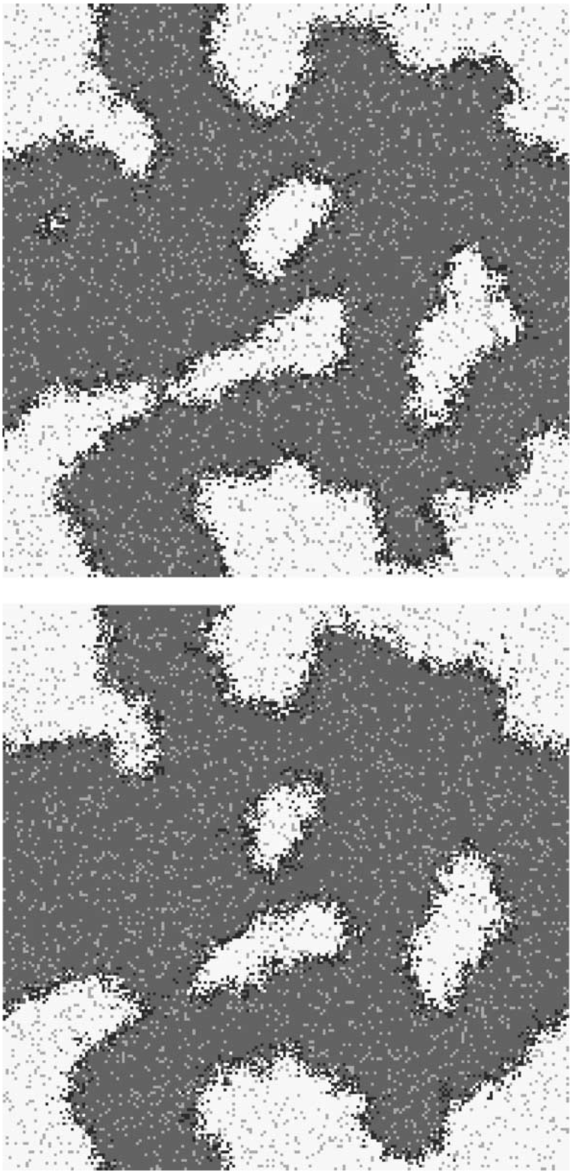

Fig. 4. Two simulations of a spatial version of the party model with introspective changes. Isolated islands tend to shrink, and boundaries become straighter by a process known as motion by mean curvature.

Exact balance happens if:

$$q_1 = \frac{0.1 - 0.0740}{0.740} = 0.035.$$

For smaller values of q_1 the symmetric fixed point is unstable for small s. For larger values of q_1 the symmetric fixed point is always stable.

If $s = 0$, there is also a fixed point with $u = q_0/(q_0 + q_1)$, $v = 0$, $x = q_1/(q_0 + q_1)$ and $y = 0$. Thus, the equilibrium consists of only 11s and 01s and all that happens are spontaneous party changes. This is what we observed in Fig. 2. Analysis of the asymmetric fixed points of this equation for $s > 0$ is algebraically difficult. However, a numerical study of our concrete example suggests the following: If the symmetric fixed point is unstable when $s = 0$ then the two asymmetric fixed points exist and are locally attracting up to the point where the symmetric fixed point becomes stable, at which point the three fixed points merge into one. That means that which party is dominant depends on initial conditions, or “frozen accidents.” If the stable fixed point is stable when $s = 0$, then the asymmetric fixed point with $u > y$ has negative values of v and y when $s > 0$. We will elaborate on the importance of these results in the next section.

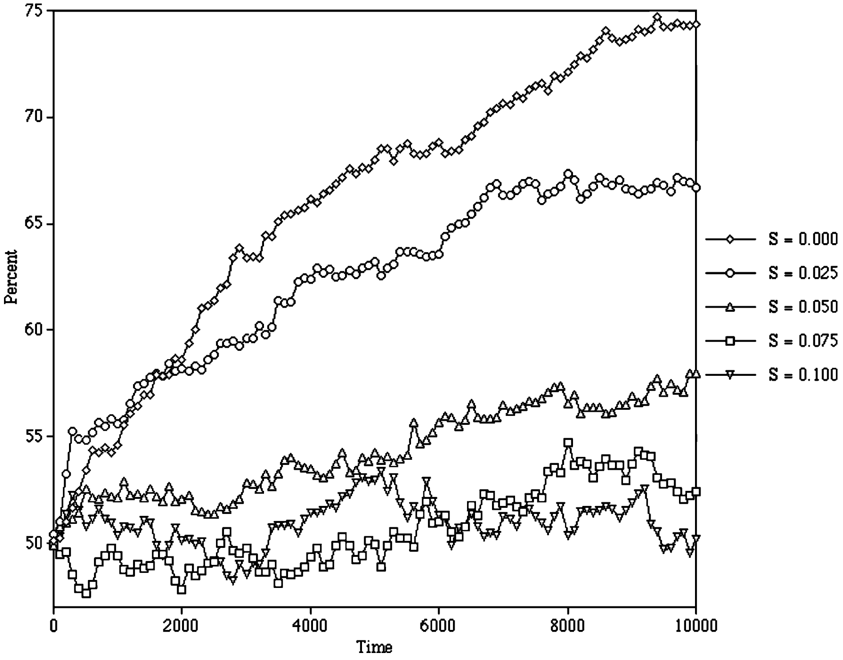

Fig. 5. Frequencies in the spatial model vs. time. Again, if the rate of introspective changes is large enough, there are two stable parties.

3.1. Spatial models

In this section, we will investigate the changes that occur when our interactions take place in a spatially structured population rather than a homogeneously mixing one. We will consider only the case of a rectangular grid. Nakamaru and Levin (2004) have examined other possibilities such as small-world graphs and networks with power-law connections. We begin with the pure imitation model, which had a trivial limit of equally likely opinions in a homogeneously mixing population. This model is an instance of Case 1 of Durrett and Levin (1994), so their results suggest that the spatial model will have an equilibrium state with short-range correlations and global frequencies like those of the equilibrium ODE.

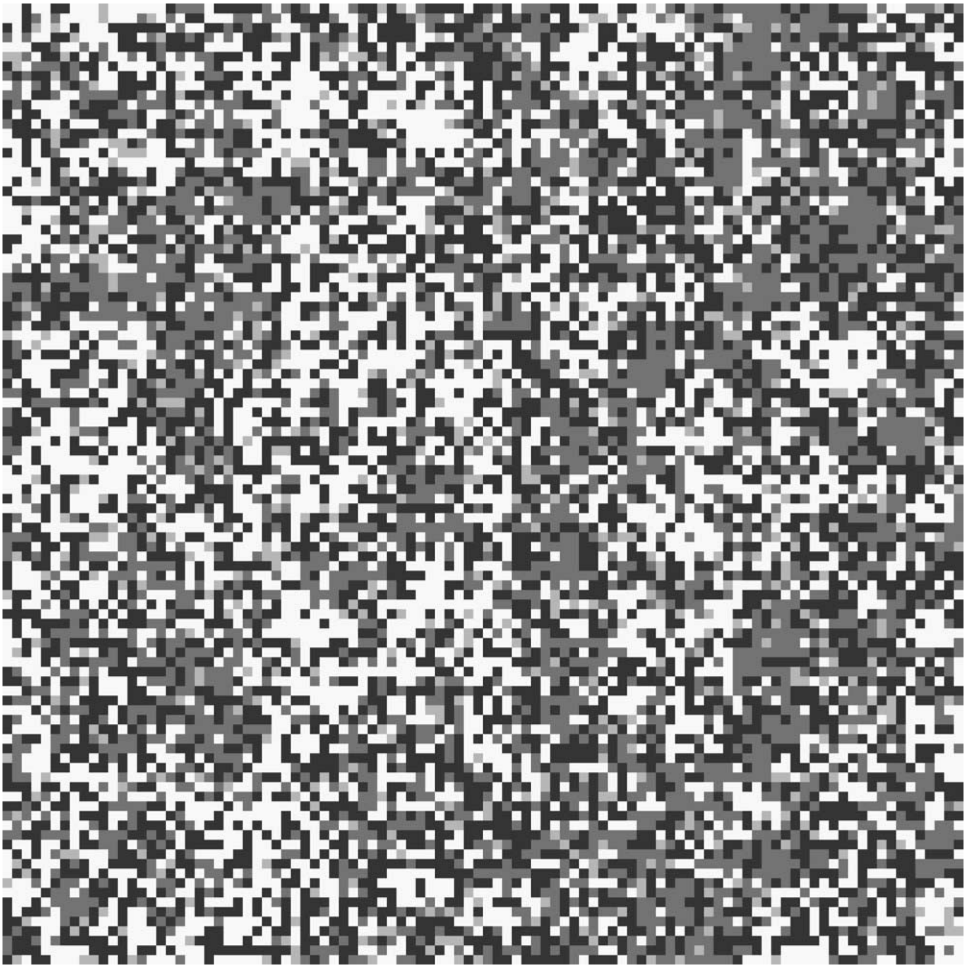

Fig. 6. A snapshot of the stationary distribution when there are two stable parties. As theory predicts the correlations between the states of sites decay rapidly with distance.

Axelrod (1997) has considered a similar model in his Chapter 7. His individuals have five cultural traits that can be any single digit number $0, 1, \ldots, 9$. An individual is chosen at random and picks a neighbor at random with probability proportional to the number of traits they share, and then imitates one of the traits chosen at random. There are two differences with our model. A minor one is that the neighbor is chosen first and then the opinion is chosen. A major one is that there is no probability of choosing a neighbor who is completely different from you (in our notation $a = 0$). This small change, setting $a = 0$, makes a big difference. Now the regions of cultural similarity grow until they cover almost all of the space. We say almost all, since it is possible to end up with small islands of individuals who agree but have no opinions in common with their neighbors (see Axelrod, 1997, p. 157).

The difference between the properties of our model with $a = 0$ and $a > 0$ is well-known in the case of one opinion, which is the voter model of interacting particle systems. Holley and Liggett showed that for finite range models in two dimensions, clustering occurs (i.e., the regions that share the same opinion grow as time goes on). In contrast, the model with $a > 0$ reaches an equilibrium distribution (see Griffeath, 1978). There are a number of mathematical results concerning the size of clusters in the model with $a > 0$ (see Sawyer, 1979), and the rate of clustering in the model with $a = 0$ (see Cox and Griffeath, 1986). However, it does not seem possible to extend these results to the case of more than one opinion. Turning now to the more complicated model with the party bit, the ordinary differential equation associated with the homogeneously mixing version of our model has either (Case 1) a globally attracting fixed point or (Case 2) two locally attracting fixed points. Results of Durrett and Levin (1994) suggest that in the second case the behavior of the system can be predicted by considering the behavior of a front separating two half spaces of individuals with all 1s and all 0s. If the front moves to favor 1s then they will become the dominant type in the system, while if the front moves to favor 0s, then they will become dominant.

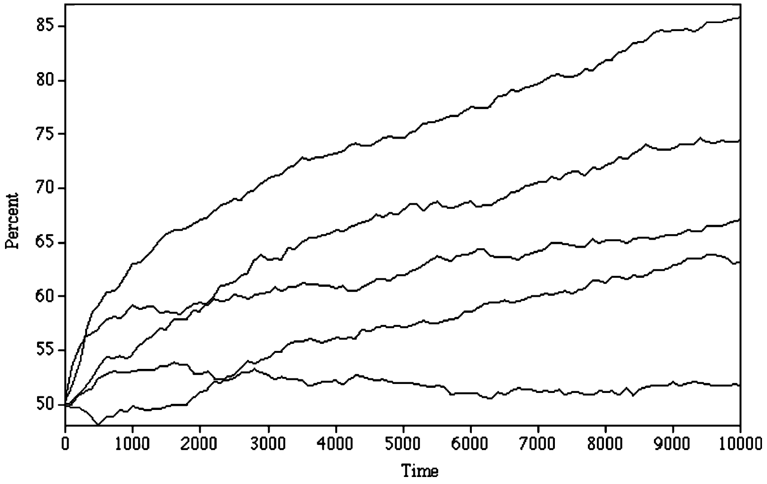

Fig. 7. Growth of the dominant party in five simulations of the spatial model with no introspective changes. Even though the parameter values are identical, the rates of growth are different.

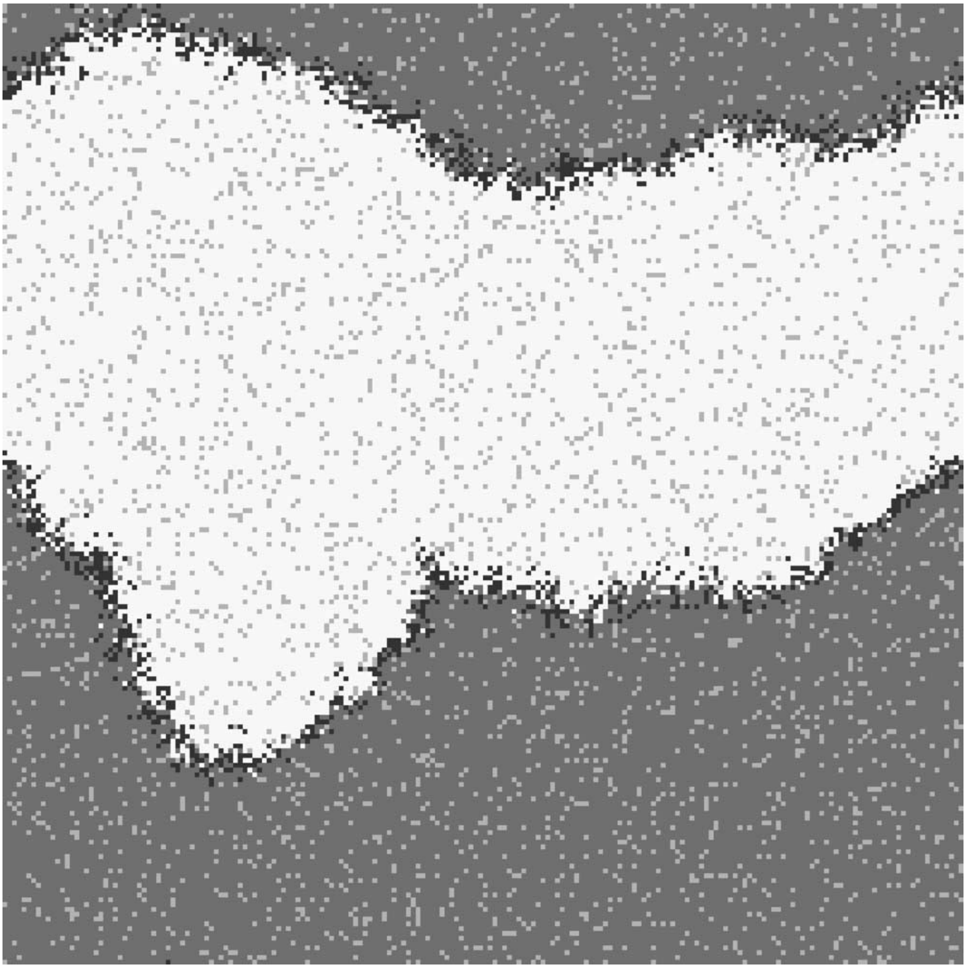

Fig. 8. Snapshot from a simulation of one of the slowly growing cases of Fig. 7. The interface has wrapped around the system. Neither region is an island, so the interface can become straight; the result is that it takes a long time for one party to be eliminated.

In our case, the model under consideration is symmetric under interchange of 0s and 1s, so the front speed is 0. In this case, it has been shown that the interface dynamics will be governed by motion by mean curvature. See results of Katsoulakis and Souganidis (1997) for the long-range Ising model of DeMasi et al. (1994) and Gandhi et al.'s (1999) study of a symmetric case of the model of Durrett and Levin (1994). In words, the interface between the 1 phase and the 0 phase will move in the direction that will straighten it; this causes islands to shrink. Since regions whose boundaries are almost straight must be large, it follows that as time increases the regions that are dominated by one type will grow. Fig. 4 gives two snapshots of the process with $s = 0$, which should help to explain the notion of motion by mean curvature.

To check these predictions we have simulated a spatial version of the model with a neighborhood that is a 5 × 5 square centered at the individual. The dynamics are the same as in the homogeneously mixing model, except that Fred chooses Ethel at random from the other people in his neighborhood. Fig. 5 shows what happens for the parameters we used in Fig. 3. As before, if s is large enough (in this case, $s \geq 0.075$), the 50–50 mix is stable. Fig. 6 shows a picture of the stationary distribution in the case $s = 0.1$, which is consistent with the Case 1 prediction of short-range correlations.

When $s = 0$, the frequency of the dominant type grows over time, but as Fig. 7 shows the rates vary dramatically among different simulation runs, and in the lowest case the frequency seems to not grow at all. Fig. 8 gives a picture of the simulation that shows the problem. To avoid problems with boundaries, we use periodic boundary conditions in our simulations (i.e., in a $L \times L$ system the displacement between two sites is computed modulo L in order to define the neighborhood). In the realization shown in Fig. 8 the boundary has wrapped around. In this situation, the boundary can become flat without eliminating one of the regions. In this case, the two opinions can happily coexist until fluctuations cause the two boundaries to collide and result in one component being bounded. In the long run, either the light area or the dark area will dominate; however, due to spontaneous changes, both groups are always present.

4. Conclusions

The evolution of cooperation among unrelated individuals has been studied from a large number of perspectives. However, most of these investigations were pursued in a game–theoretic context and incorporate payoffs for cooperating individuals. Here, we studied the emergence of stable social groups in an agent-based model without reward and punishment. Our individuals have several bits of information that describe their group affiliation and their opinions about one or more issues. Individuals are more likely to imitate the opinions of others in their own group, especially when the imitation brings them back to the party line. This mechanism is sometimes able to produce two stable social groups of roughly equal size. However, in all cases if we add a mechanism that causes individual opinions to revert to those of the group, then stability occurs when these introspective changes occur at a large enough rate. The reader should note that in contrast to the results of Macy et al. (2004), we obtain polarization without xenophobia, the tendency to adopt opinions opposite from one's neighbors.

Acknowledgements

We are pleased to acknowledge the support of the National Science Foundation, through Grant DMS-0202935 to RTD, and the David and Lucile Packard Foundation, through awards 2000-02688 to the Institute for Advanced Study and 2000-01690 to the Santa Fe Institute. This research was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, foundations, and corporate partners. As always, we acknowledge the computational wizardry of Linda

Buttel who performed the simulations and made the figures. We also gratefully acknowledge numerous communications with Janet Best, Samuel Bowles, Partha Dasgupta, Paul Ehrlich, Isha Ray and Desmond McNeill, and the stimulating environments of the Beijer Institute and the Santa Fe Institute.

Appendix A. Detailed calculations

To analyze (2) we can rewrite the equations as:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -u(q_1 + s) + v\sigma s + xq_0 + r_1 uv + r_0(vx - uy)$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -y(q_1 + s) + x\sigma s + vq_0 + r_1 xy + r_0(vx - uy)$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -v(q_0 + \sigma s) + us + yq_1 - r_1 uv - r_0(vx - uy)$$

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -x(q_0 + \sigma s) + ys + uq_1 - r_1 xy - r_0(vx - uy).$$

Since $u + v + x + y = 1$ there are only three independent equations here. It will be convenient for us to work with the following three combinations:

$$\frac{\mathrm{d}(u - y)}{\mathrm{d}t} = -(u - y)(q_1 + s) - (v - x)(q_0 - \sigma s) + r_1(uv - yx) \tag{A.1}$$

$$\frac{\mathrm{d}(v - x)}{\mathrm{d}t} = -(v - x)(q_0 + \sigma s) - (u - y)(q_1 - s) - r_1(uv - yx) \tag{A.2}$$

$$\begin{aligned}\frac{\mathrm{d}(u + y - v - x)}{\mathrm{d}t} &= -2(u + y)(q_1 + s) + 2(x + v)(q_0 + \sigma s) + 2r_1(uv + xy) \\ &\quad + 4r_0(vx - uy).\end{aligned} \tag{A.3}$$

Symmetry dictates that the line $u = y$, $v = x$, $u + v + x + y = 1$ is invariant. On this line the right-hand sides of the first two equations vanish and $v = x = 1/2 - u$, so the condition for an equilibrium is:

$$\begin{aligned}0 &= -2(2u)(q_1 + s) + 2(1 - 2u)(q_0 + \sigma s) + r_1(2u(1 - 2u)) + r_0((1 - 2u)^2 - (2u)^2) \\ &= c + b(2u) + a(2u)^2\end{aligned} \tag{A.4}$$

where $c = r_0 + 2(q_0 + \sigma s)$, $b = -2(q_1 + \mathrm{s}) - 2(q_0 + \sigma s) + r_1 - 2r_0$, and $a = -r_1$. We have $a < 0$, $c > 0$, and $a + b + c = -r_0 - 2(q_1 + s) < 0$, so as claimed in (3), there is a unique fixed point with $2u \in (0, 1)$ given by:

$$v_0 = \frac{-b - \sqrt{b^2 - 4ac}}{4a}, \qquad v_0 = \frac{1}{2} - u_0, \qquad y_0 = u_0, \qquad x_0 = v_0.$$

To check the stability of this fixed point we use (A.1) and (A.2) to conclude that if we start from $u = u_0 + \varepsilon$, $v = v_0 + \delta$, $x = v_0 - \delta$, $y = u_0 - \varepsilon$, then $uv - yx = 2\varepsilon v_0 + 2\delta u_0$ and

$$\begin{aligned} \frac{\mathrm{d}\varepsilon}{\mathrm{d}t} &= -\varepsilon(q_1 + s) - \delta(q_0 - \sigma s) + r_1(u_0\delta + v_0\varepsilon) \\ \frac{\mathrm{d}\delta}{\mathrm{d}t} &= -\delta(q_0 + \sigma s) - \varepsilon(q_1 - s) - r_1(u_0\delta + v_0\varepsilon). \end{aligned} \tag{A.5}$$

We can rewrite the last system as:

$$\begin{aligned} \frac{\mathrm{d}\varepsilon}{\mathrm{d}t} &= \varepsilon(v_0 r_1 - q_1 - s) + \delta(u_0 r_1 - q_0 + \sigma s) \\ \frac{\mathrm{d}\delta}{\mathrm{d}t} &= \varepsilon(-v_0 r_1 - q_1 + s) + \delta(-u_0 r_1 - q_0 - \sigma s). \end{aligned}$$

To begin the stability analysis we note that

$$\frac{\mathrm{d}(\varepsilon + \delta)}{\mathrm{d}t} = -2q_1\varepsilon - 2q_0\delta,$$

so the matrix

$$A = \begin{pmatrix} v_0 r_1 - q_1 - s & u_0 r_1 - q_0 + \sigma s \\ -v_0 r_1 - q_1 + s & -u_0 r_1 - q_0 - \sigma s \end{pmatrix}$$

always has one negative eigenvalue. The determinant of A is:

$$\begin{aligned} &= (v_0 r_1 - q_1 - s)(-u_0 r_1 - q_0 - \sigma s) - (u_0 r_1 - q_0 + \sigma s)(-v_0 r_1 - q_1 + s) \\ &= 2r_1(q_1 u_0 - q_0 v_0) + s(2q_1\sigma + 2q_0). \end{aligned} \tag{A.6}$$

Recalling that the determinant is the product of the eigenvalues, we have the result claimed in (4).

When $s = 0$, the second term disappears and we are left to consider the sign of $q_1 u_0 - q_0 v_0$. If we let f_{ux} be the flow from 11 to 01 in the symmetric equilibrium (i.e., the rate of jumps 11 → 01 minus the rate of jumps 01 → 11) and define the other flows similarly, then

$$\begin{aligned} f_{ux} &= u_0 q_1 - x_0 q_0 = u_0 q_1 - v_0 q_0 \\ f_{vu} &= v_0(r_{11}u_0 + r_{01}x_0) - u_0(r_{10}v_0 + r_{00}y_0) = r_1 u_0 v_0 + r_0(v_0 x_0 - u_0 y_0) \\ f_{yv} &= f_{ux},\ f_{xy} = f_{vu}. \end{aligned}$$

In equilibrium, the flows in and out of each vertex have to be balanced. By symmetry this is equivalent to $f_{vu} = f_{ux}$. A very special way to achieve this is to have detailed balance (i.e., all of the flows equal to 0). Since $u_0 + v_0 = 1/2$, we have $f_{ux} = 0$ if and only if:

$$u_0 = \frac{q_0}{2(q_1 + q_0)} \qquad v_0 = \frac{q_1}{2(q_1 + q_0)}.$$

Setting $4f_{vu} = 0$ and using symmetry leads to a quadratic equation:

$$0 = r_1(2u_0)(1 - 2u_0) + r_0((1 - 2u_0)^2 - (2u_0)^2) = r_0 + (r_1 - 2r_0)(2u_0) - r_1(2u_0)^2$$

which can be solved to give:

$$2u_0 = \frac{2r_0 - r_1 - \sqrt{4r_0^2 + r_1^2}}{-2r_1}.$$

In order for the system to satisfy detailed balance the two formulas for u_0 must agree, i.e.:

$$\frac{q_0}{q_0 + q_1} = \frac{r_1 - 2r_0 + \sqrt{r_1^2 + 4r_0^2}}{2r_1}. \tag{A.7}$$

If $r_0 = 0$, the right-hand side is $1 > q_0/(q_0 + q_1)$ and $f_{vu} = r_1 u_0 v_0 > 0$. From this we see that if < holds in (A.7) the flows are positive and the symmetric fixed point is stable even when $s = 0$. Conversely, if > holds, the symmetric fixed point is unstable for small s. This is the result given in (5).

To look for asymmetric fixed points, we add Eqs. (A.1) and (A.2) to conclude:

$$0 = -2q_1(u - y) - 2q_0(v - x)q_0.$$

Using $u + v + x + y = 1$ gives two equations:

$$\begin{aligned} v + x &= 1 - (u + y) \\ q_0(x - v) &= q_1(u - y) \end{aligned} \tag{A.8}$$

which can be solved to express v and x in terms of u and y.

$$\begin{aligned} x &= \frac{1}{2}(1 - (u + y)) + \frac{q_1}{2q_0}(u - y) \\ v &= \frac{1}{2}(1 - (u + y)) - \frac{q_1}{2q_0}(u - y). \end{aligned} \tag{A.9}$$

Plugging into (A.1) gives:

$$\begin{aligned} \frac{\mathrm{d}(u - y)}{\mathrm{d}t} &= -(u - y)(q_1 + s) + (q_0 - \sigma s)\frac{q_1}{q_0}(u - y) \\ &\quad + r_1\left(\frac{u - y}{2}(1 - (u + y)) - (u + y)\frac{q_1}{2q_0}(u - y)\right). \end{aligned}$$

From this we see that if $u \neq y$, then:

$$0 = -(q_1 + s) + (q_0 - \sigma s)\frac{q_1}{q_0} + \frac{r_1}{2}\left(1 - (u + y) - \frac{q_1}{q_0}(u + y)\right).$$

Rearranging, we have:

$$\frac{2s}{r_1}\left(1 + \sigma\frac{q_1}{q_0}\right) = 1 - (u + y)\left(1 + \sigma\frac{q_1}{q_0}\right).$$

Solving gives:

$$u + y = \frac{q_0}{q_0 + q_1} - \frac{2s}{r_1} = \frac{q_0 + \sigma q_1}{q_0 + q_1}. \tag{A.10}$$

Having found $u + y$, our final step is to compute $u - y$ using (A.3). Letting $w = u + y$ and $z = u - y$, using (A.9) and noting that $-4uy = z^2 - w^2$,

$$4(vx - uy) = (1 - w)^2 = \frac{q_1^2}{q_0^2}z^2 + z^2 - w^2.$$

(A.8) implies that $x + v = 1 - w$ and $x - v = (u - y)q_1/q_0$, so we have:

$$2(uv + xy) = (u + y)(v + x) + (u - y)(v - x) = w(1 - w) - \frac{q_1}{q_0}z^2.$$

Using the last two equations in (A.3) we have:

$$\begin{aligned} &-2w(q_1 + s) + 2(1 - w)(q_0 + \sigma s) + r_1 \\ &\quad = \left(w(1 - w) - \frac{q_1}{q_0}z^2\right) + r_0\left((1 - w)^2 - w^2 + \left(1 - \frac{q_1^2}{q_0^2}\right)z^2\right) = 0. \end{aligned} \tag{A.11}$$

From the last equation it follows that there are at most two fixed points different from the symmetric one. If $s = 0$, we have $w = q_0/(q_0 + q_1)$. Plugging this into (A.11), we have:

$$r_1\frac{q_0 q_1}{(q_0 + q_1)^2} + r_0\frac{q_1 - q_0}{q_0 + q_1} + \left(r_0\left(1 - \frac{q_1^2}{q_0^2}\right) - r_1\frac{q_1}{q_0}\right)z^2 = 0.$$

For this equation one can guess and verify that $z = w$ is a solution. Since $u + y = u - y$, it follows that $y = 0$. Plugging the values for z and w into (A.9), we find $x = q_1/(q_0 + q_1)$, $v = 0$. Thus, the equilibrium consists of only 11s and 01s and all that happens are spontaneous party changes.

References

Aoki, K., 2001. Theoretical and empirical aspects of gene–culture coevolution. Theoretical Population Biology 59, 253–261.

Axelrod, R., 1997. The Complexity of Cooperation. Princeton University Press, Princeton.

Bicchieri, C., 1997. Learning to cooperate. In: Bicchieri, C., Jeffrey, R., Skyrms, B. (Eds.), The Dynamics of Norms. Cambridge University Press, Cambridge, pp. 17–46.

Blackmore, S., 2000. The Power of Memes. Scientific American, October 2000, pp. 53–54.

Bowles, S., 2001. Individual interactions, groups conflicts, and the evolution of preferences. In: Durlauf, S., Young, P. (Eds.), Social Dynamics. MIT Press, Cambridge, pp. 155–190.

Boyd, R., Richerson, P.J., 1985. Culture and Evolutionary Process. University of Chicago Press, Chicago.

Boyd, R., Richerson, P.J., 1996. Why culture is common, but cultural evolution is rare. In: Runciman, W.G., Smith, J.M., Dunbar, R.I.M. (Eds.), Evolution of Social Behavior Patterns in Primates and Man. Oxford University Press, Oxford, pp. 77–93.

Cavalli-Sforza, L.L., Feldman, M.W., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.

Cox, J.T., Griffeath, D., 1986. Diffusive clustering in the two dimensional voter model. Annals of Probability 14, 347–370.

Darwin, C., 1859. The Origin of Species. Murray, London.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, New York.

DeMasi, A., Orlandi, E., Presutti, E., Triolo, L., 1994. Glauber evolution with K_{ac} potentials. I. Mesoscopic and macroscopic limits, interface dynamics. Nonlinearity 7, 633–696.

Durrett, R., Levin, S.A., 1994. The importance of being discrete (and spatial). Theoretical Population Biology 46, 363–394.

Ehrlich, P.R., 2000. Human Natures: Genes, Culture, and the Human Prospect. Island Press, Washington, DC.

Gandhi, A., Levin, S., Orszag, S., 1999. Nucleation and relaxation from meta-stability in spatial ecological models. Journal of Theoretical Biology 200, 121–146.

Griffeath, D., 1978. Additive and Cancellative Interacting Particle Systems. Springer Lecture Notes in Math 724. Springer–Verlag, New York.

Haldane, J.B.S., 1948. The theory of cline. Journal of Genetics 48, 227–284.

Hamilton, W.D., 1964. The genetical evolution of social behavior. Journal Theoretical Biology 7, 1–52.

Henrich, J., 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. Journal of Economic Behavior and Organization 53, 3–35.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of *Homo economicus*: behavioral experiments in 15 small-scale societies. American Economic Review 91, 73–78.

Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994. Preferences, property rights and anonymity in bargaining games. Games and Economic Behavior 7, 346–380.

Holley, R., Liggett, T.M., 1975. Ergodic theorems for weakly interacting systems and the voter model. Annals of Probability 3, 643–663.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proceeding of the National Academy of Science USA 79, 2554–2558.

Katsoulakis, M.A., Souganidis, P.E., 1997. Stochastic Ising models and anisotropic front propogation. Journal of Statistical Physics 87, 63–89.

Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. Proceedings of the Royal Society of London B 13, 745–753.

Lewontin, R.C., 1961. Evolution and the theory of games. Journal of Theoretical Biology 1, 382–403.

Macy, M.W., Kitts, J.A., Flache, A., 2004. Polarization in dynamic networks: a Hopfield model of emergent structure. In: Breiger, R., Carley, K., Pattison, P. (Eds.), Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers. The National Academy Press, Washington, DC, pp. 162–173.

Maynard Smith, J., 1974. The theory of games and the evolution of animal conflicts. Journal of Theoretical Biology 79, 19–30.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Nakamaru, M., Levin, S.A., 2004. Spread of two linked social norms on complex interaction networks. Journal of Theoretical Biology 230, 57–64.

Nowak, M.A., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.

Oster, G.F., Wilson, E.O., 1978. Caste and Ecology in the Social Insects. Princeton University Press, Princeton.

Riolo, R.L., Cohen, M.D., Axelrod, R., 2001. Evolution of cooperation without reciprocity. Nature 414, 441–443.

Sawyer, S., 1979. A limit theorem for patch sizes in a selectively neutral migration model. Journal of Applied Probability 16, 482–495.

Simon, H.A., 1990. A mechanism for social selection and successful altruism. Science 250, 1665–1668.

Skyrms, B., 1996. Evolution of the Social Contract. Cambridge University Press, New York.

Sugden, R., 1986. The Economics of Rights, Co-operation and Welfare. Blackwell, Oxford.

Veblen, T., 1902. The Theory of the Leisure Class: An Economic Study of Institutions. MacMillan, New York.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. Nature 39, 440–442.

West, G.B., Brown, J.H., Enquist, B.J., 1997. A general model for the origin of allometric scaling laws in biology. Science 276, 122–126.

Young, H.P., 1993. The evolution of conventions. Econometrica 61, 57–84.