Altruists, Egoists, and Hooligans in a Local Interaction Model

By Ilan Eshel, Larry Samuelson, and Avner Shaked*

We study a population of agents, each of whom can be an Altruist or an Egoist. Altruism is a strictly dominated strategy. Agents choose their actions by imitating others who earn high payoffs. Interactions between agents are local, so that each agent affects (and is affected by) only his neighbors. Altruists can survive in such a world if they are grouped together, so that the benefits of altruism are enjoyed primarily by other Altruists, who then earn relatively high payoffs and are imitated. Altruists continue to survive in the presence of mutations that continually introduce Egoists into the population. (*JEL* C70, C78)

An act is altruistic if it confers a benefit on someone else while imposing a cost on its perpetrator. How does costly altruistic behavior survive? Why doesn't utility maximization inexorably eliminate such behavior?

One answer is immediately available: allegedly altruistic acts are not really altruistic. Upon closer examination, they confer net benefits rather than costs. For example, charitable donations may bring benefits such as public recognition or a warm glow that overwhelm the cost of the donation. If we push revealed preference theory to its logical limit, this conclusion becomes as inescapable as it is tautological. If someone commits an "altruistic" act, then this reveals that he prefers doing so.[1]

A second answer is also available: the interaction in which the altruistic act occurs may be repeated. If the interaction is infinitely repeated, then the folk theorem (Drew Fudenberg and Eric Maskin, 1986) ensures that there are equilibria in which Altruists survive, though there are also equilibria in which altruism does not appear.[2] David M. Kreps et al. (1982) show that there are equilibria in which Altruists survive in finitely repeated games with incomplete information, though once again a folk theorem result appears, including equilibria without altruism (Fudenberg and Maskin, 1986).

We do not doubt that people often derive benefits from seemingly altruistic acts, and that many interactions are repeated. However, we also believe that altruistic acts occur for which conventional models do not readily account. Embellishing the models to encompass such acts often leads to utility functions that are uncomfortably exotic or to an uncomfortably strong faith in repetition.

This paper provides an alternative model of altruistic behavior with two key properties. First, we abandon the assumption that people

* Eshel: Department of Statistics, School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel; Samuelson: Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706; Shaked: Department of Economics, University of Bonn, Adenauerallee 24-26, D 53113 Bonn, Germany. We thank Ed Green, Reinhard Selten, and an anonymous referee for helpful comments. Financial support from the National Science Foundation and the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged. Part of this work was done while Eshel and Samuelson were visiting the Department of Economics at the University of Bonn and while Samuelson and Shaked were visiting the Institute for Advanced Studies at the Hebrew University of Jerusalem. We are grateful to both for their hospitality and support.

[1] There are many parallels between discussions of altruism in economics and biology. Biologists frequently argue that once the full effects of actions on reproductive success are recognized, seemingly altruistic acts are privately optimal. For example, there are now numerous explanations for why it may be individually beneficial for a bird to attract a predator's attention by uttering a cry that alerts the bird's flock to the predator's presence. Richard Dawkins (1976) discusses bird alarms as well as other cases of apparent altruism.

[2] In biological contexts, explanations for altruism similar to those that arise in repeated games appear in the guise of reciprocal altruism (Robert L. Trivers, 1971).

are rational agents choosing utility-maximizing actions. Instead, we believe that people must learn which actions work well, and that an important force in learning is imitation. Second, interactions between agents in our model are "local," meaning that altruistic acts are more likely to affect nearby agents than more distant neighbors and that agents are more likely to imitate nearby than more distant neighbors.

To see how these forces can allow altruism to survive, suppose there are two kinds of agents: Altruists, who provide a public good to their neighbors at a cost to themselves, and Egoists, who do not do so. Suppose further that Altruists tend to exist in concentrated groups. Altruists can then earn higher payoffs than Egoists, because Altruists are more likely to enjoy the public goods provided by other Altruists. The imitation-based learning process now prompts other agents to become Altruists. In addition, nearby agents are the ones most likely to imitate the Altruists. This preserves the tendency of Altruists to clump together in groups and hence preserves the conditions needed for altruism to survive.

This argument is unconvincing without some stability analysis. We expect perturbations to occasionally switch the behavior of some agents from Altruist to Egoist or Egoist to Altruist, perhaps because someone has analyzed the model and deduced that it is utility maximizing to be an Egoist, or has made a mistake, or has simply experimented with a new action. An Egoist thrust into the midst of Altruists will thrive on the public goods provided by the latter and will be imitated, while an Altruist thrust in the midst of Egoists will fare poorly and will be ignored. Perturbations or "mutations" that occasionally cause people to switch strategies thus apparently produce a force pushing toward egoism. An explanation of altruistic behavior must demonstrate that altruism can withstand such mutations.[3]

We find that only states composed primarily of Altruists survive in the presence of rare mutations. If a mutation introduces an Egoist in the midst of Altruists, then the Egoist will survive and spread. However, the resulting group of Egoists quickly confronts limits on its ability to expand, as each expansion causes the public goods supplied by neighboring Altruists to be shared among more and more Egoists and hence reduces Egoists' payoffs. Egoists are thus readily introduced but cannot expand beyond small, isolated groups. Isolated Altruists, in contrast, cannot even survive in the midst of Egoists. However, mutations will occasionally introduce a group of Altruists in the midst of Egoists. Such a group of Altruists can expand without bound. Mutations thus more readily lead to large groups of Altruists than Egoists, allowing the former to dominate.

Altruism is not the only type of externality that can arise between agents. We extend the model to consider Hooligans, or agents who benefit from imposing damages on their neighbors. The same forces that allow Altruists to survive in the presence of Egoists also allow Altruists to survive against Hooligans, or Egoists to survive in the midst of Hooligans, though Hooligans will typically not be eliminated entirely. The analysis is then further extended to general 2×2 games, allowing us to examine games with two strict Nash equilibria, one payoff dominant and one risk dominant.

The work most closely related to ours includes Theodore C. Bergstrom and Oded Stark (1993), Lawrence E. Blume (1993), and Glenn Ellison (1993), and papers by Martin A. Nowak and Robert M. May (1992, 1993) and Nowak et al. (1994). Our spatial structure matches one of the models considered by Bergstrom and Stark as well as the simplest case considered by Ellison, while Blume and Nowak and May examine spatial models in which agents are arranged in a plane rather than along a line. We differ from Ellison and Blume in taking imitation, rather than some variant of best-reply dynamics, to be the driving force behind strategy selections. This is crucial, as altruism has no hope in a world of best responders. We differ from Nowak and May in relying on analytical techniques rather than simulations, albeit for a very simple

[3] Group selection models share many of the features of our model of local interaction, but are now widely considered to be implausible as explanations for altruistic behavior, largely because of their inability to withstand mutations. We discuss the relationship of our work to group selection arguments in Section V.

model, and especially in studying the effect of mutations.

Section I presents the model. Section II examines equilibria of the imitation process. There are many possible limiting outcomes, depending upon the initial conditions of the system, but Altruists comprise a significant portion of the population in all but one of these. We establish conditions under which the probability of an initial condition leading to the elimination of altruism shrinks to zero as the population grows large. Section III shows that only those limiting outcomes with a significant proportion of Altruists survive mutations. Section IV pursues a generalization of the model in which agents interact in larger neighborhoods. In doing so, we find that it can be to Altruists' advantage to have a relatively high cost of altruism. A higher cost of altruism ensures that if Altruists survive, then they must do so in larger groups, because only then do they share enough of the public good to compensate for the high cost of being an Altruist. This in turn ensures that if there are any Altruists at all, then there are relatively large groups of Altruists. Section V concludes. Unless otherwise noted, proofs are contained in the Appendix.

I. Altruists and Egoists

We consider a collection of N individuals, where N is finite. Each individual can be either an Altruist or an Egoist. An Altruist provides a public good that contributes one unit of utility to those who receive its benefits. The *net* cost to the Altruist of providing the public good is $C > 0$, so that the combination of enjoying the benefits of his own public good and bearing the costs of its provision reduces the Altruist's utility by C. Egoists provide no public goods and bear no costs. Instead they simply enjoy the benefits of the public goods provided by others.

Time is divided into discrete periods. At the end of each period, after consuming any public good that is available and bearing provision costs (if an Altruist), each agent decides, according to a learning rule, whether to be an Altruist or Egoist in the next period.

The nature of this learning rule is important. One possibility is that the agents are fully rational, though myopic, and the learning rule leads them to adopt single-period, expected-utility-maximizing actions. In this case, they will realize they face a variant of the prisoner's dilemma and will play the strictly dominant strategy, namely Egoist. Instead of choosing best replies, however, our players imitate the strategies of others whom they observe to be earning high payoffs.

At one level, this imitation seems preposterous. How hard can it be to figure out that being an Egoist (or defecting in a prisoner's dilemma) is a strictly dominant strategy? This is indeed a trivial task for a game theorist facing the sterilized 2×2 games with which we often work. However, these games are a simplified representation of a much more complicated reality. The agents who actually play the game may not recognize that they are playing a game, may not know who their opponents are, may not know what strategies are available, and may not know what payoffs these strategies bring. They may then be unable to think like game theorists, or like the agents in game-theoretic models. At the same time, we believe that people are generally able to form a good estimate of others' payoffs, whether these payoffs are measured in terms of money or other units such as social status or prestige, and that people tend to imitate the behavior of those they observe earning high payoffs.

Imitation alone appears to hold out no hope for the survival of altruism. Egoists will enjoy the same public goods as Altruists, while only the latter bear costs. As a result, all Egoists will earn higher payoffs than all Altruists and imitation can only lead players to become Egoists.

This argument is applicable only if the benefits of the public good provided by each Altruist extend to every agent in the population. The prospects for Altruists improve if the public good is a *local* public good. To make this precise, we introduce a neighborhood structure taken from Bergstrom and Stark (1993) and Ellison (1993). Agents in the model are located around a circle.[4] Each agent interacts

[4] A spatial interpretation is convenient, but local interaction structures may arise in other ways. In academia, field of specialization is probably more important than location in determining patterns of interaction.

with his two immediate neighbors, i.e., with one agent to his right and one to his left. If an agent is an Altruist, then his immediate neighbors enjoy the benefit of his public good provision. The payoff of agent i is then given by $N_i^A - C$ if i is an Altruist and N_i^A if i is an Egoist, where $N_i^A \in \{0, 1, 2\}$ is the number of i's Altruist neighbors (excluding himself).

In each period, each agent takes a draw from an independent Bernoulli trial, causing the agent to "learn" with probability $\mu \in (0, 1]$ and to retain her strategy with probability $1 - \mu$. An agent who learns observes her own payoff and the payoff and strategy of each agent in her neighborhood. She then chooses to be an Egoist if the average payoff of the Egoists in her sample exceeds that of Altruists, and chooses to be an Altruist if the average payoff of Altruists exceeds that of Egoists.[5] If an agent and her two neighbors all play the same strategy, be it Altruist or Egoist, then the agent will continue to play that strategy.

We shall concentrate on the case of $\mu = 1$, so that every agent learns in every period. This gives a deterministic learning process which simplifies the derivation and statement of the results. However, Bernardo A. Huberman and Natalie S. Glance (1993) have recently argued that the outcomes of local interaction models can be sensitive to whether all agents adjust their strategies at the same time. We accordingly comment on how each of our results would be modified if $\mu < 1$.

A state is a specification of which agents are Altruists and which are Egoists. Let S be the set of possible states. For states i and j in S, let P_{ij} be the probability that a single iteration of the imitation process changes the system to the state j given that the current state is i. Since the learning process is deterministic (with $\mu = 1$), P_{ij} is either 0 or 1. The collection $\{P_{ij}\}_{i,j \in S}$, along with a specification of the initial state at time zero, is a Markov process on the state space S. We refer to this Markov process as the "imitation dynamics."

II. Equilibrium

A. *Absorbing Sets*

We are interested in the stationary distributions of the imitation dynamics. We say that a set of states is *absorbing* if it is a minimal set of states with the property that the Markov process can lead into this set but not out of it. An absorbing set may contain only one state, say i, in which case $P_{ii} = 1$ and i is a stationary state of the Markov process. An absorbing set may contain more than one state, in which case $P_{ij} = 0$ if i is contained in the absorbing set and j is not, while the Markov process cycles between states in the absorbing set.

For each absorbing set of the Markov process, there is a unique stationary distribution the support of which consists of that absorbing set. We can then learn much about the stationary distribution of the learning process by studying absorbing sets.

We begin by compiling a description of the imitation dynamics. We assume $C < 1/2$.[6] At the end of each period, an agent may either retain her strategy or choose a strategy played by one of the two agents closest to her, depending upon their payoffs. These payoffs in turn depend on the strategies of the next two neighbors. The fate of an individual is then completely determined by the strategies of her four nearest neighbors.

An Egoist who learns by imitating his neighbors can become an Altruist only if at least one of his two nearest neighbors is an Altruist. However, if both of his immediate neighbors are Altruists, then the Egoist earns a payoff of two, more than an Altruist can ever earn, causing the Egoist to retain his strategy. An Egoist can therefore become an Altruist only if exactly one of his neighbors is an Altruist, with the Altruist neighbor earning a higher payoff than the average payoffs of the Egoist and his Egoist neighbor. This payoff inequality holds only if the Altruist has an Altruist neighbor, since otherwise the Altruist receives the lowest possible payoff of $-C$, and if the other Egoist in the neighborhood faces a neighborhood containing only Egoists, so as to bring the average Egoist payoff below $1 - C$. Hence, an Egoist can become an Altruist only if he faces either the following combination of strategies or its mirror image, where "a" represents an Altruist and "E" an Egoist,

(1) $\quad aa \quad E \quad EE$

and where it is the central Egoist who converts to an Altruist.[7] In all other cases, Egoists remain Egoists.

A similar calculation shows that an Altruist will remain an Altruist if and only if one of the following combinations of strategies (or their mirror images) occurs,

(2) $\quad xa \quad a \quad ax$

$\quad\quad aa \quad a \quad EE$

where it is the central Altruist whose fate is in question and where an "x" holds the place of an agent who may be either an Altruist or an Egoist. In all other cases, Altruists change to Egoists.

Conditions (1)–(2) provide a complete description of the individual imitation dynamics. To illustrate some absorbing sets, we represent the agents as being located on a line, where we think of the ends of the line as being joined to form a circle. From (1)–(2), we easily verify that the following are absorbing sets:

- The state in which all are Altruists.
- The state in which all are Egoists.
- A state in which all are Altruists except two adjacent Egoists:

$$\cdots aaaaaaaaEEaaaaaaaa\cdots.$$

- A set of two states, consisting of:

$$\cdots aaaaaaaaaEaaaaaaaaa\cdots$$

$$\cdots aaaaaaaaEEEaaaaaaaa\cdots.$$

In this last case, the imitation dynamics cycle between the two states in the absorbing set. The lone Egoist initially earns the highest possible payoff of 2, inducing his two neighbors to become Egoists and leading to the second state in the cycle. Each of these new Egoists finds himself in the situation described by (1), where he has two Egoists on one side and two Altruists on the other. This causes the new Egoists to switch back to altruism, beginning the cycle anew. We refer to such a cycle as a *blinker*.

The two outside agents in the blinker face a coordination problem. It is an equilibrium for one but not for both to be an Egoist, and the learning scheme causes them to cycle around this equilibrium. We suspect that cycles in behavior do occur, though our simple model captures these cycles in a crude way. The presence of blinkers is a product of setting $\mu = 1$, forcing all agents to assess their strategies in every period. If $\mu < 1$, then blinkers are no longer absorbing sets, since a period will eventually arise in which only one of the two outside agents in the blinker revises her strategy, leading to a pair of adjacent Egoists. All absorbing sets would then be singletons.

These examples, and combinations constructed from them, include all of the possibilities for absorbing sets. Some terms will be useful in making this precise. If agents α and β play the same strategy, either Altruist or Egoist, and if all agents between α and β play this strategy, then we will refer to agents α, β, and the intermediate agents as an interval of

[5] We simplify the analysis by choosing the cost C so that payoff ties do not arise. There are many other plausible learning rules. For example, an agent may simply compare the best Egoist and best Altruist payoff among those payoffs she observes, or may compare the sum of the Egoist and Altruist payoffs, rather than considering averages. Itzhak Gilboa and David Schmeidler (1995, 1996), in the context of their case-based decision theory, examine the difference between considering the sums or the averages of payoffs.

[6] We can always vanquish altruistic behavior by making it too expensive. If $C > 1/2$, we find that the only absorbing sets of the learning process are those consisting of single states in which either all agents are Altruists or all are Egoists. The former contains only itself in its basin of attraction, while the latter attracts the remainder of the state space. In the presence of mutations, only the latter absorbing state survives. The case of $C = 1/2$ leads to inconvenient payoff ties.

[7] We find the displays easiest to read if we use a lower case "a" to represent Altruists, and the text easiest to read if we continue to use "A." We will also often separate agents in whom we are interested by spaces, as in the case of the central Egoist here, though these spaces have no significance other than directing attention to particular agents.

either Altruists or Egoists. We call a maximal such interval a *string*. Notice that strings may be of any length from 1 to N, the length of the circle. We then have (the proof is in the Appendix):[8]

PROPOSITION 1: *Let* $0 < C < \frac{1}{2}$ *and* $\mu = 1$. *Then:*

(1.1) *Absorbing sets consist of (i) the state in which all agents are Egoists, (ii) the state in which all agents are Altruists, and (iii) sets containing states in each of which Altruist strings of length three or longer are separated by Egoist strings of length less than four. These sets are either singletons (in which case all Egoist strings are of length two) or contain two states [in which case any string of length one (three) in one of the states blinks to a string of length three (one) in the other].*

(1.2) *Except for the state in which all agents are Egoists, the proportion of Altruists in an absorbing state, or the average proportion over the two states in an absorbing set, is at least* 0.6.

Proposition 1 indicates that there are many absorbing sets, each of which is the support of a stationary distribution of the imitation process. In all but one of these absorbing sets, the majority of the population is Altruists. Hence, there is no possibility for moderation in altruism. If Altruists survive at all, they must be the majority.

To see what lies behind this result, we first note that a string of Egoists in an absorbing set can never be longer than three. If the length of an Egoist string exceeds three, then the two Egoists at its edges will each have two Egoists on one side and two Altruists on the other, and hence they will become Altruists [cf. (1)], causing the string to shrink. Egoists can thus survive only in strings of length two or strings of length one (where the latter alternate with strings of length three in a blinker). Altruist strings must be at least length three in order to survive, and surviving Altruist strings can expand, since doing so creates more and more high-payoff Altruists. This allows us to conclude that if there are any Altruists at all, then Altruists will occur in strings of length at least three while Egoists occur in strings of at most two (or in blinkers the average length of which is two), and hence there will be at least 60-percent Altruists.

B. *Basins of Attraction*

Because the state in which all agents are Egoists is absorbing, the system may drive Altruists to extinction. To assess the likelihood of such an event, we identify the initial conditions from which the system converges to an absorbing set containing Altruists.

The proof of Proposition 1 shows that any string of Altruists either drops below length three at some point, after which it disappears, or persists forever. We refer to a string of Altruists whose fate is the latter as a "persistent" string. The system will converge to a state in which at least 60 percent of the agents are Altruists if and only if the initial condition contains at least one persistent string.

The following proposition first characterizes persistent strings. We then suppose that agents' initial identities as either Altruists or Egoists are randomly determined, and investigate the probability that this leads to an initial state containing a persistent string of Altruists. We have:[9]

PROPOSITION 2: *Let* $0 < C < \frac{1}{2}$ *and* $\mu = 1$. *Then:*

(2.1) *A string of Altruists is persistent if and only if (i) the string contains at least five Altruists, (ii) the string consists of four Altruists bordered on at least one end by two Egoists, or (iii) the string consists of three Altruists bordered on each end by two Egoists or bordered on at least one end by three Egoists. All other strings of Altruists are eliminated by period three.*

(2.2) *If agents' initial identities as Altruists or Egoists are determined by independent, identically distributed random variables placing positive probability on Altruist, then as N gets large, the probability of a persistent string of Altruists in the initial state, and hence convergence to an absorbing set containing at least 60-percent Altruists, approaches unity.*

Under randomly determined initial conditions and a large population, the probability that Altruists survive is high because there will almost certainly be an initial group of Altruists large enough to ensure their survival, and hence to ensure that most agents are eventually Altruists. However, a great deal of growth may be required before a single group of Altruists can comprise an appreciable fraction of a large population. How long must we wait before most agents are Altruists?

By period three, any string of Altruists that is not persistent will have been eliminated, and the population will consist of persistent strings of Altruists separated by strings of Egoists. If a string of Egoists is not already of length two or a blinker, then it will contract at a rate of two agents per period, as the Egoists on the two ends of the string switch to altruism. We can accordingly pose our waiting-time question as the following: how long do we expect to wait until a string of agents lying between two persistent Altruist strings has been reduced to length two or to a blinker? But since this waiting time is half of the string's length, plus possibly three periods, we can equivalently ask how long a string of agents we expect to find between two persistent strings of Altruists. There is no reason why strings of Egoists lying between surviving strings of Altruists should be longer in larger populations, and hence no reason for expected waiting times to increase as N increases. However, the expected length of such a string is quite sensitive to the probability that an agent in the initial condition is an Altruist. We can calculate:[10]

PROPOSITION 3: *Let* $0 < C < \frac{1}{2}$ *and* $\mu = 1$, *and let agents' identities in the initial condition be randomly and independently determined, with probability p attached to being an Altruist. Then, in the limit as the population gets arbitrarily large, an upper bound on the expected length of a string of agents between two persistent Altruist strings is given by:*

p	Length
0.5	22
0.4	34
0.3	63
0.2	171
0.1	1140
0.05	8480
0.01	1.01×10^6
0.001	1.00×10^9.

Expected waiting times are thus relatively moderate as long as there is a sufficiently high initial probability that a randomly selected agent is an Altruist. For example, a probability of altruism of 10 percent gives an upper bound of 573 on the expected waiting time (half of the Egoist string's expected length, plus three). On the other hand, persistent strings will be extremely rare if Altruists are very rare, and expected waiting times will be very long.[11]

C. *Hooligans*

Altruism, conferring a benefit on someone else at a cost to oneself, is not the only way that one agent's actions may affect another. At the opposite extreme we have Hooligans, who benefit by imposing harm on others. Notice that hooliganism need not be limited to the psychopathic. Those who litter in order to avoid the cost of disposing of their refuse,

[8] If $\mu < 1$, then this proposition continues to hold, except that blinkers are no longer absorbing sets.

[9] This result holds if $\mu < 1$, with the modification that some strings of two Altruists, as well as strings of the form *aEaaaEEE*, survive with a probability greater than zero but less than one.

[10] If $\mu < 1$, then agents learn less frequently, and expected waiting times will be longer.

[11] We expect to wait longer until Altruists dominate the population when Altruists are rare, but the resulting absorbing states are likely to have higher proportions of Altruists. This result holds because such initial conditions will be characterized by relatively small numbers of long strings of Egoists, who will be transformed into small numbers of short strings of Egoists. When Altruists are more likely, there will be many short but distinct strings of Egoists in the initial condition, leading to an absorbing state with more strings of Egoists.

I		II	
		X	Y
	X	a, a	b, c
	Y	c, b	d, d

FIGURE 1. 2 × 2 GAME

those who pollute rather than take costly abatement measures, and those who shirk in group efforts are all Hooligans.

Our model is easily generalized to accommodate Hooligans. Let there be two types of agents, denoted by 1 and 2. Let type 1 contribute K_1 to the payoff of each of his neighbors at a cost of C_1 to himself. Let type 2 contribute K_2 to each neighbor at a cost C_2 to himself. There is no loss of generality in assuming that $K_1 > K_2$. Our model of Altruists and Egoists is then the special case in which $K_1 = 1$, $C_1 = C$, and $K_2 = C_2 = 0$. The behavior of the model depends only on a single parameter:[12]

PROPOSITION 4: *Let $K_1 > K_2$. Then any variation in the values of K_1, K_2, C_1, and C_2 that preserves*

$$\frac{C_1 - C_2}{K_1 - K_2} \tag{3}$$

gives rise to the same imitation dynamics.

Hence, any two specifications of the payoffs that preserve $(C_1 - C_2)/(K_1 - K_2)$ give rise to the same absorbing sets, basins of attraction, and dynamic paths for the imitation dynamics, and the same limiting distributions in the presence of mutations.

For the Altruist and Egoist model of the previous sections, the ratio (3) was C, which was interpreted as the cost of altruism. Consider the following pairs of types of players. In each case, the first column identifies the effect an agent of type 1 has on his two neighbors and the cost to the agent of that effect, while the second column provides analogous information for an agent of type 2.

$$\begin{pmatrix} K_1 & K_2 \\ C_1 & C_2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ C & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & -C \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ C & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & -1 \\ C & -C \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ -C & -2C \end{pmatrix}.$$

The first specification is the familiar Altruist and Egoist pair from previous sections. The second pair of agents consists of an Egoist and a Hooligan who enjoys (incurs a negative cost from) causing damage of one unit to his neighbors. In case this Hooligan seems too malicious in his enjoyment of the harm he causes, the third pair rewrites this situation as an agent of type 1 who imposes no harm on others but incurs a cost of C to avoid doing so, with a type-2 agent who does not incur the cost and imposes damage of one unit on his neighbors. The fourth pair includes an Altruist and a Hooligan. The last pair has two Hooligans, one of whom causes twice the damage and doubly benefits from doing so. In each of these specifications, the ratio $(C_1 - C_2)/(K_1 - K_2)$ is given by C, and hence these are equivalent models. As long as $C < 1/2$, it is always the first type in each pair that will come to comprise the majority of a large population with a randomly determined initial condition. Hooligans will then be in the minority when facing Egoists or Altruists, though some Hooligans will survive, just as some Egoists survive when paired against Altruists.[13]

Similar insights can be used to extend the analysis to general 2 × 2 symmetric games. Suppose each agent must choose a single strategy to use when playing the game shown in Figure 1 with each of his neighbors. Without

[12] The proof establishes this proposition for a wide class of imitation dynamics including those of the current section as a special case.

[13] The final specification shows that Hooligans will be in the majority if paired with even worse Hooligans.

sacrificing generality, we can assume $a > d$. We will then further concentrate on the case in which $a > b$. This latter assumption excludes some games but retains the common examples of 2 × 2 games. Then an argument analogous to the proof of Proposition 4 shows that the imitation dynamics depends only upon the two numbers:

$$\alpha = \frac{c - b}{a - b}, \quad \beta = \frac{d - b}{a - b}. \tag{4}$$

In light of this, we can transform the payoffs in Figure 1 by subtracting b from each payoff and dividing by $a - b$ to obtain the equivalent representation of the game given in Figure 2, where $\alpha = (c - b)/(a - b)$ and $\beta = (d - b)/(a - b)$.

We can now classify games according to the values of α and β, where $\beta < 1$ (because we have assumed $a > d$). We have:

- Prisoner's Dilemma: $0 < \beta < 1$, $1 < \alpha$.
- Coordination Game: $0 < \beta < 1$, $\alpha < 1$.
- Chicken: $\beta < 0$, $1 < \alpha$.
- Efficient Dominant Strategy: $\beta < 0$, $\alpha < 1$.

This classification is illustrated in Figure 3. An "efficient dominant strategy" game is one in which X is a strictly dominant strategy and the outcome (X, X) is efficient, unlike the prisoner's dilemma. A coordination game has two strict Nash equilibria, given by (X, X) and (Y, Y). Chicken has one mixed-strategy Nash equilibrium and two asymmetric pure strategy equilibria.[14] In the case of a coordination game, (X, X) is the payoff-dominant equilibrium (because $a > d$ and hence $\beta < 1$), and is also risk dominant if $\alpha + \beta < 1$, while the equilibrium (Y, Y) is risk dominant if $\alpha + \beta > 1$. The interval in which $\alpha = 1 + 1/2 C$, $\beta = 1/2 C$, and $C < 1/2$, shown in Figure 3, describes the range of Altruist and Egoist games that was analyzed in subsections A and B of this section.

I		II	
		X	Y
	X	1, 1	0, α
	Y	α, 0	β, β

FIGURE 2. TRANSFORMATION OF GAME IN FIGURE 1

The methods developed in the previous sections to examine Altruists and Egoists can be applied to any other game in this classification. For example, consider coordination games. Let $\alpha + \beta > 1$ so that (X, X) is the payoff-dominant equilibrium but (Y, Y) is the risk-dominant equilibrium. Now consider a boundary between a group of agents playing strategy X and a group playing strategy Y, or[15]

$$\cdots xxxxxxxxxxYYYYYYYYYYY \cdots.$$

The only agents at risk of changing their strategies are the two agents, one playing X and one playing Y, at the ends of their respective strings. Each faces a neighborhood with one X and one Y agent, in addition to themselves. In Ellison's (1993) model, each chooses a best response to his two neighbors. By assumption, Y is risk dominant and hence is a best reply when one neighbor plays X and one plays Y. Hence, the agent playing Y retains his strategy while the agent playing X switches to Y. The string of Y's thus grows while the string of X's shrinks, ensuring that best-reply learning leads to the selection of the risk-dominant equilibrium.

In our imitation model, the X player on the boundary earns a payoff of 1, while the adjacent X player earns 2. The Y player on the boundary earns $\alpha + \beta$ while the adjacent Y player earns 2β. Comparing the average payoffs, we find that the boundary player Y retains his strategy if $\alpha + 3\beta > 2$, while a boundary

[14] The asymmetric pure strategy equilibria of this symmetric game become relevant if agents can condition their strategies on some asymmetry, such as location.

[15] As with Altruists and Egoists, we find the displays easier to read if we use a lower case x to represent the strategy X.

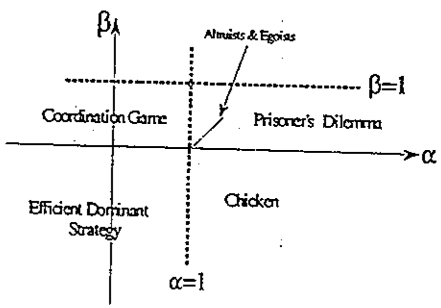

FIGURE 3. CLASSIFICATION OF GAMES

X will turn into Y if $\alpha + \beta > {}^{3}/_{2}$. Hence, there are three subregions of (α, β) values within the region of coordination games in which risk dominance and payoff dominance conflict. In the first region, $\alpha + 3\beta > 2$ and $\alpha + \beta > {}^{3}/_{2}$, and hence both players will play Y in the following period. In the second region, $\alpha + 3\beta < 2$ and $\alpha + \beta < {}^{3}/_{2}$, and hence both will play X in the following period. In the third region, $\alpha + 3\beta > 2$ and $\alpha + \beta < {}^{3}/_{2}$, and both will retain their strategy. In the first region, the string playing the risk-dominant action will grow, in the second it will shrink while the string playing the payoff-dominant action grows, and in the third region each string maintains its length.

If a string of agents playing the risk-dominant action Y is to expand, its payoffs must provide a premium over that required for risk dominance (i.e., $\alpha + \beta > {}^{3}/_{2}$). This is necessary because an agent at the end of a string of X agents compares not whether X or Y is a best reply, but whether the X or Y players in his neighborhood are earning higher average payoffs. One of the X players in his neighborhood is bordered by two other X players, and hence receives an exceptionally high payoff. Risk dominance alone is not enough to overcome this payoff.

If a string of agents playing the *payoff-dominant* action X is to expand (while Y is risk dominant), then we must have $\alpha + 3\beta < 2$ and $\alpha + \beta < {}^{3}/_{2}$. In conjunction with the requirement that $\alpha + \beta > 1$, these inequalities require $\alpha > \beta$. Hence, the payoff to playing strategy Y must be greatest if the opponent plays X, even though (Y, Y) is an equilibrium. This occurs because the neighborhood of an agent who plays Y, and who is located on the end of a string of Y agents, contains a Y player who faces two Y opponents and hence earns a relatively high payoff. The average payoff to X can be highest only if the payoff in equilibrium (Y, Y) is relatively small. This is in turn compatible with risk dominance only if $\alpha > \beta$.

Given that strategy Y must receive a premium over risk dominance in order to expand, and given that strategy X can expand in the absence of this premium only if the additional condition $\alpha > \beta$ holds, then it is no surprise that there are some cases in which neither string will expand. Imitation can then yield peaceful coexistence of the two strategies, unlike best-response behavior. Imitation allows the coordination failures created by coexistence to persist because agents on the boundary of a string, and hence experiencing coordination failures, are most likely to observe other agents with the same strategies who are not facing coordination failures and to observe agents with the other strategy who *are* plagued by such failures. This introduces a force against changing strategies, and builds sufficient inertia into the system to support coexistence.

III. Mutations

We now ask how altruism fares in the presence of mutations. We assume that at the end of each period, and after imitation has occurred, each agent takes a draw from an independent, identically distributed Bernoulli random variable. With probability λ, this agent is a mutant and changes his type, either from Altruist to Egoist or from Egoist to Altruist. With probability $1 - \lambda$, this agent experiences no mutation. We will be interested in the case in which λ is small, so that imitation is the primary force driving strategy revisions. We study this by examining the limiting case as the mutation probability λ goes to zero.[16]

[16] In economics, the common practice is to follow the lead of Michihiro Kandori et al. (1993) and H. Peyton

Let Γ_{ij} be the probability that the combination of imitation and mutation changes the state to j given that the current state is i.[17] Then $\{\Gamma_{ij}\}_{i,j \in S}$ is again a Markov process on the state space S, which we refer to as the "imitation-and-mutation dynamics." Notice that $\Gamma_{ij} > 0$ for all i and j, which is to say that for any two states i and j, there is some combination of mutations capable of changing the system from i to j. Hence, for each fixed mutation rate, the imitation-and-mutation dynamics has a unique stationary distribution. The proportions of states reached along any sample path approach this distribution almost surely, and the distribution of states at time t approaches this distribution as t gets large. (John G. Kemeny and J. Laurie Snell [1960 Theorems 4.1.4, 4.1.6, and 4.2.1].)

We study the limit of these stationary distributions as the probability of a mutation λ gets small, which we refer to as the *limiting distribution*.

PROPOSITION 5: *Let* $0 < C < {}^{1}/_{2}$. *If* $N > 30$, *then the limiting distribution places positive probability only on states contained in absorbing sets of the imitation process in which the proportion of Altruists is at least* 0.6.

The techniques involved in establishing this result, which holds for $\mu \in (0, 1]$, were developed by M. I. Freidlin and A. D. Wentzell (1984) and were introduced into economics by Kandori et al. (1993) and Young (1993). The argument begins by observing that when the mutation rate is small, the system spends virtually all of its time in absorbing sets of the imitation dynamics, and hence the limiting distribution allocates all of its probability to such sets. Movements between absorbing sets of the imitation dynamics can be accomplished only by mutations. The system will allocate most of its probability to absorbing sets of the imitation process that are easy to reach, in the sense that it requires relatively few mutations to reach their basin of attraction from other absorbing sets. The proof involves showing that as long as the population is sufficiently large, it is much easier for mutations to introduce Altruists into a world of Egoists than for mutations to eradicate Altruists from a purely altruistic or mixed world.

One's initial impression might be that mutations should be inimical to Altruists, because a mutant Egoist will thrive and grow when introduced into a collection of Altruists while a lone Altruist will wither and die when introduced into a collection of Egoists. Notice, however, that a small clump of Altruists in the midst of Egoists will not only survive, but will grow. It takes only three adjacent Altruists in a world that is otherwise completely Egoists to ensure that the imitation dynamics lead to an absorbing set containing at least 60-percent Altruists.

It takes only a single mutation to introduce an Egoist into a world of Altruists. However, the resulting Egoist string can grow no longer than three. In light of this, consider an initial state that consists only of Altruists. A mutation creating an Egoist or a clump of Egoists will prompt imitation dynamics leading to an absorbing state with no more than three Egoists. To get additional Egoists, additional mutations are required. These mutations can lead to states where there are many small clumps of Egoists. As these clumps become more numerous, and hence closer together, additional mutations join together previously separated clumps of Egoists. But as long as there are still some strings of Altruists, these newly joined strings of Egoists will shrink, replacing two original strings with a new, shorter string (of length three or less) and ultimately *decreasing* the proportion of Egoists. In order to further increase the proportion of Egoists, mutations must simultaneously eliminate *all* strings of Altruists. But this requires a large number of mutations, if N is large, and hence is extraordinarily unlikely. If N exceeds 30, then it takes at least four mutations to eliminate all Altruists, which suffices for the result.

Altruists can thus invade a world of Egoists with only a local burst of mutation that creates a small string of Altruists, which will then subsequently grow to a large number of

Young (1993) in concentrating on arbitrarily small mutation rates. In biology, the concept of an evolutionarily stable strategy (John Maynard Smith [1982]) is built around the presumption that mutations are arbitrarily improbable compared to the forces of selection.

[17] If Q_{ij} is the probability that mutations change the state to j, given that the current state is i, then $\Gamma_{ij} = \Sigma_k P_{ik} Q_{kj}$.

Altruists. Mutations can create small pockets of egoism, but these pockets destroy one another if they are placed too close together, placing an upper bound on the number of Egoists that can appear. The only possibility for surpassing this bound lies in a "global" mutation combination that simultaneously attacks all strings of Altruists. Mutations thus lead much more readily to absorbing sets with Altruists than absorbing sets without them, and the limiting distribution concentrates all of its probability on the former. This reinforces our finding that absorbing sets containing Altruists are the limiting outcomes in the absence of mutations, as long as there is some initial probability of altruism in a large population. Our model thus differs from many mutation-counting analyses, in that our limiting distribution does *not* depend critically on highly improbable sequences of mutations and hence need not involve extraordinarily long waiting times.

If we consider large populations, mutations will ensure that there are more than 60-percent Altruists in the population, though less than 100 percent. The exact calculation of the limiting distribution is tedious, but we can establish some bounds. The calculation of these bounds is significantly simpler for the case of $\mu < 1$, though μ can be arbitrarily close to one.[18]

The argument proceeds by noting that if the number of Egoist strings is too small, then the Egoist strings will be far apart and most Altruist strings will be long. A mutation will then tend to strike in the midst of Altruists and create a new string of Egoists, increasing the number of Egoists. If there are many Egoists strings, then these strings will be relatively close together, separated by short Altruist strings. Mutations will then often strike sufficiently close to two Egoist strings as to give rise to imitation dynamics that merge the two Egoist strings, thereby reducing the number of Egoists. We thus expect a centralizing tendency in the number of Egoists. We have:

PROPOSITION 6: *Let $\mu < 1$. Then the limit of the limiting distribution, as the population size gets large, restricts probability to absorbing sets in which the proportion of Altruists is between 70 percent and 87 percent.*

IV. Larger Neighborhoods

We have assumed that agents interact only with their immediate neighbors. This section examines an extension of the model that allows us to make the following point: decreasing the cost of altruism can be bad for Altruists.

We consider the case where each Altruist contributes one unit of the public good to each of his *four* closest neighbors. Each agent observes his own payoff and that of his four closest neighbors, and then chooses the strategy from those played by this group with the highest average payoff. We say that neighborhoods are of "radius two" in this case.

As in the previous case, the cost of altruism plays a crucial role in shaping the results. We study two intervals for the parameter C, namely $(^3/_4, ^5/_6)$ and $(^5/_6, 1)$.[19] Changing the value of C within such an interval does not affect the outcome, while we shall see that the two intervals give different behavior.[20] We let $\mu = 1$ throughout.

The investigation of the model with neighborhoods of radius two and costs $C \in (^5/_6, 1)$ begins with a calculation of transition rules. The nontrivial conditions under which an Altruist will remain an Altruist are the following cases and their mirror images, where the Altruist in the center is the agent in question and

[18] The advantage of $\mu < 1$ is that all of the absorbing sets under the random imitation dynamics are then singletons. This makes it easier to do the necessary calculations (and reduces the number of calculations) of the probabilities that mutations transform a given absorbing set into the basin of attraction of another.

[19] When agents interacted only with their immediate neighbors, the only relevant cost consideration was whether C was larger or smaller than $^1/_2$.

[20] If $C > ^3/_4$, then altruism is so costly that only Egoists survive. The results for $1 < C < ^5/_4$ are qualitatively similar to those for $^5/_6 < C < 1$, with one quantitative difference noted below. Cost levels $^1/_2 < C < ^3/_4$ give results similar to those of $^3/_4 < C < ^5/_6$. Costs $C < ^1/_2$ give noticeably different and more complicated behavior that we do not investigate here. Cost levels that lie at the boundaries between these various intervals create complications arising out of cases in which the average payoffs to Altruists and Egoists are equal.

x stands for a strategy that could be either A or E:

$$\begin{array}{llll} (5) & Eaaa & a & xEEE \\ & aaaa & a & EEEx \\ & aaaa & a & aEEE. \end{array}$$

The cases in which an Egoist will become an Altruist are the following, as well as their mirror images:

$$\begin{array}{llll} (6) & aaaE & E & EEEx \\ & Eaaa & E & EEEE \\ & aaaa & E & EEEx. \end{array}$$

These allow us to prove:[21]

PROPOSITION 7: *Let neighborhoods be of radius two and let $^5/_6 < C < 1$. Then absorbing sets generically consist of (i) the state in which all agents are Egoists and the state in which all agents are Altruists, and (ii) sets containing states in which strings of Altruists of length five or more are separated by either strings of three E's or blinkers, where blinkers consist alternately of one E and five E's or consist alternately of two E's and six E's. With the exception of the state in which all agents are Egoists, the proportion of Altruists is at least $^5/_9$.*

The proof mimics that of Proposition (1) and is omitted. To obtain the minimal proportion of Altruists in the stable sets that contain Altruists, we note that we can pack blinkers that alternate between two and six E's next to each other with five A's between them, in the following way:

$$\begin{array}{ll} (7) & \cdots aaaaaEEaaaaaEEEEEEaaaaa \cdots \\ & \cdots aaaEEEEEEaaaaaEEaaaaaaa \cdots. \end{array}$$

This guarantees a maximum of Egoists, and here we have a proportion of $10/18 = ^5/_9$ Altruists.

For costs in the interval $(^3/_4, ^5/_6)$, we have the following:

PROPOSITION 8: *Let neighborhoods be of radius 2 and let $^3/_4 < C < ^5/_6$. Then generically, absorbing sets include (i) the state in which all agents are Egoists and all are Altruists, and (ii) sets containing states in which strings of three or more Altruists are separated by strings of exactly three Egoists, or by blinkers which alternate between one and five or between two and six Egoists.[22] Except for the state in which all agents are Egoists, the proportion of Altruists in an absorbing set is a least $^1/_2$.*

The proof again mimics that of Proposition (1). To obtain the lower bound on the number of Altruists, note that it is possible to arrange blinkers in the following way:

$$\begin{array}{ll} (8) & \cdots aaaEaaaEEEEEaaaEaaaEEEEEaaaEaaaaa \cdots \\ & \cdots aEEEEEaaaEaaaEEEEEaaaEaaaEEEEEaaa \cdots. \end{array}$$

This has $^1/_2$ of the population as Altruists. There is no denser way to arrange blinkers.

In each case a result analogous to Proposition 2 holds, establishing that if agents' initial identities are independently determined and may be altruistic, then as the population grows, the probability of convergence to a state in which altruism survives approaches unity.

The lower bound on Altruists is lower for $C \in (^3/_4, ^5/_6)$ than in the case of higher costs, being $^1/_2$ rather than $^5/_9$. In this sense, it can be disadvantageous for Altruists to have their altruism come too cheaply. The forces behind this result are revealed by comparing (7) and (8). Example (7) reflects the fact that when

[21] The "generically" in this statement allows us to avoid values of C within the interval $(^5/_6, 1)$ that create payoff ties between Altruists and Egoists. For $1 < C < ^5/_4$ $(C \neq ^7/_6)$, we have the same characterization of absorbing sets, except that blinkers must be separated by at least six Altruists, making the minimal percentage of Altruists 0.6.

[22] If two one/five blinkers are separated by a string of only three Altruists, then the blinkers must be out of phase, so that the state in which one of the blinkers has five Egoists is the state in which the other blinker has only one Egoist. A two/six blinker requires at least five Altruists on each side.

costs are relatively high, strings of Altruists must be at least five Altruists long in order for Altruists to receive payoffs high enough to survive. Example (8) reflects the fact that for lower costs, Altruists' payoffs are higher and shorter strings (of length three) of Altruists can survive. The imitation dynamics can then lead to outcomes in which islands of Egoists are separated by strings of only three rather than five Altruists, and hence a smaller proportion of Altruists.

V. Conclusion

We have shown that if players choose their strategies in games by imitating successful players, and if there is a local or neighborhood structure to both the interaction between agents and their learning, then altruistic behavior can survive.

Imitation and the local nature of the interactions are both important to this result. Best-response learning would immediately lead agents to adopt the dominant strategy of Egoist. Agents who choose their strategies by imitating others will imitate Altruists, but only if the latter happen to be earning relatively high payoffs. The role of the local interaction structure is to allow Altruists to huddle together in concentrated groups. The benefits of the public goods supplied by Altruists are then enjoyed primarily by Altruists, leading to higher payoffs than those of Egoists, who tend to be surrounded by other Egoists.

A group of Altruists is always a ripe target for invasion by a mutant Egoist, who will thrive on the public goods provided by the Altruists. For this reason Altruists can survive, but they generally cannot conquer. Instead, the Altruists will be riddled with pockets of Egoists. However, there are limits to the expansion of egoism. As more and more Egoists try to free ride on nearby Altruists, the payoffs of Egoists fall and imitators become Altruists. The result is the preservation of altruism in coexistence with egoism. We can hope for altruism, but not for a perfect world of altruism.

Attention is drawn to the importance of the mutations in a local interaction model by results from biological studies of group selection. Group selection models typically assume that agents are arranged in isolated groups created by either spatial separation or kinship relationships.[23] Payoffs are measured in terms of expected numbers of offspring, and the proportion of the population playing a relatively high-payoff strategy increases because the agents playing that strategy have relatively large numbers of offspring. As in our model of learning by imitation, the important property of this reproduction-based dynamic process is that if the agents in a location tend to be predominately Altruists, then altruism can spread to nearby locations, as the concentrated Altruists earn high payoffs and hence produce many offspring that spill over into neighboring locations.

Kin selection theories are now widely accepted as explanations for some seemingly altruistic behavior.[24] However, group selection models that are not based on kinship relationships have been criticized [e.g., Williams (1966); Dawkins (1976)], initially because the mechanism that caused some groups to grow faster than others was not specified, and subsequently because the combination of small groups, rare mutations, and infrequent migration required to support altruism is thought to be implausible.[25]

Our model differs from a typical biological model of group selection in that our agents are arranged in overlapping rather than isolated groups. This overlapping-neighborhoods structure opens the possibility that mutations introducing Egoists into our model can favor Altruists while being detrimental to Egoists, by disrupting and eliminating strings of Altruists and hence causing two groups of Egoists to join and then shrink to a single, small group. This contrasts with typical biological group selection models, where migration and especially mutation work relentlessly against altruism and one must struggle to find a plausible explanation for why the migration and mutation rates are sufficiently small to allow altruism to survive. In addition, mutations in our model can introduce Altruists in the midst of Egoists. In biological models with isolated groups, such a possibility is thought to involve mutation rates that are unrealistically high and group sizes that are unrealistically small. In the presence of overlapping groups, relatively small numbers of mutations allow Altruists to gain a local foothold from which they can spill over into nearby locations. We could thus reinterpret our analysis as a biological model with local interactions and dynamics based on reproduction and emigration, providing a new explanation for altruism that exploits the overlapping group structure.

Our model of agents occupying locations around a circle is very simple. What happens if they are placed in a plane, or in a higher-dimension structure? To gain some insight into these cases, recall that Altruists fare poorly when exposed to many Egoists, while Egoists fare well when exposed to many Altruists. Taking agents to be arranged along a circle ensures that any group of A's cannot have too many Altruists who are on the boundary and hence are exposed to Egoists, and ensures that any group of Egoists cannot have too many members exposed to Altruists. This in turn produces conditions under which Altruists are likely to thrive. Moving to the plane or to richer spaces raises the possibility that groups of Altruists will appear that are irregularly shaped and that expose virtually all of their members to Egoists. These Altruists may then not survive, and the persistence of altruism appears to be less certain.

The extensive simulations of Nowak and May (1992, 1993) and Nowak et al. (1994) suggest that in the absence of mutations, there are many initial conditions from which a significant proportion of Altruists persist. Once again, Egoists in their model do well in the midst of Altruists while Altruists do poorly in the midst of Egoists, and concentrated groups of Altruists can then expand. The dynamics are much more complicated than in our simple model, but altruistic behavior typically survives.

What if the game contains more than two strategies? Altruists can survive in our model because the Altruists near the end of a long string of Altruists earn higher payoffs than do the Egoists near the end of long strings of Egoists. When we extended the argument to general 2×2 games, the criterion for expansion turned out to be a mixture of efficiency and risk dominance. In games with more than two strategies, the criteria for whether strategy x can expand at the expense of y, i.e., for whether agents on the boundary between strings of x and y will switch to x, involve pairwise efficiency and risk-dominance considerations. If there is a strategy that is relatively efficient and does not fare too badly in pairwise risk-dominance comparisons with all other strategies, then the system can converge to states featuring primarily that strategy. However, cyclic behavior can also appear in which strategy x expands at the expense of y, y expands at the expense of z, and z expands at the expense of x.

Imitation may often be important, but adopting the strategy observed to earn the highest average payoff is a very simple decision process. What about imitation rules other than simply comparing average payoffs? For example, agents may base their choices not only on average payoffs but also on the number of agents they observe playing each strategy. This introduces elements familiar from the literature on strategies for the infinitely repeated prisoner's dilemma. Tit for tat, for example,

[23] See, for example, Vero Copner Wynne-Edwards (1962, 1986), W. D. Hamilton (1964, 1972), George C. Williams (1966), Eshel (1972), David Sloan Wilson (1975, 1987), Dan Cohen and Eshel (1976), C. Matessi and S. D. Jayakar (1976), Maynard Smith (1976), and Eshel and L. L. Cavalli-Sforza (1982).

[24] For example, kin selection arguments have been used to explain the behavior of several species of tropical butterflies, some of which incur a cost to develop a bitter taste that discourages birds from preying on others [Lincoln Pierson Brower and Jane Van Zandt Brower (1964), Brower (1969), Woodruff W. Benson (1971), and Eshel (1972)]. Kinship relationships play an especially important role in explaining the behavior of the social insects.

[25] Robert Boyd and Peter J. Richerson (1985) suggest that group selection arguments may be applicable in explaining the evolution of altruistic behavior among humans. They examine a model in which people have a taste for conformity, so that altruism is a *strict* best response as long as sufficiently many other people cooperate. In biology, group selection arguments are often invoked to explain the inefficiency of weapons used in competition for mates, such as excessively branched or curved horns (Konrad Lorenz, 1963), and are used to explain self-imposed limits on reproductive ability when a population is threatened by overpopulation. Wynne-Edwards (1962) suggests examples of the latter phenomenon, and evidence is provided by Frank Fenner (1965) (Mixoma viruses, which infect rabbits), John J. Christian (1970) (rodents and small mammals), and Frank M. Stewart and Bruce R. Levin (1984) (bacterial viruses).

simply adopts the previous strategy it has observed. The implications once again depend upon the behavior of an agent located at the end of a string of similar agents. An Altruist at the boundary between sufficiently long strings of Altruists and Egoists observes equal numbers of Altruists and Egoists among her opponents, as does the adjacent Egoist. If observing the Altruists in this sample makes agents more likely to cooperate by being Altruists, then our results are reinforced. If observing the Egoists makes the agents more likely to retaliate by being Egoists, and if this is sufficient to overcome the average payoff advantages of Altruists, then our results will be reversed and altruism will vanish. The success of altruism then depends upon whether agents facing both Altruists and Egoists tend to see their glasses as half full or half empty.

A great deal of work remains to be done in extending the analysis to larger games as well as more complicated spatial structures and learning rules. It is clear, however, that dynamics driven by imitation can differ significantly from the familiar best-reply dynamics and that imitation coupled with local interactions opens the possibility for altruistic behavior to survive.

Appendix

PROOF OF PROPOSITION 1:

It is immediate that the states in which all agents are Altruists or all agents are Egoists are absorbing states, because imitation cannot introduce Egoists into a world in which there are only Altruists, or vice versa.

To find the remaining absorbing sets, consider what happens to a string of A's as the imitation dynamics proceed. From (2), any A string of length one immediately disappears. Similarly, if we have an A string of length two, the two A's in this string immediately become E's. In the process, however, the adjacent E's may switch to A's. What happens to these adjacent E's? There are four possibilities. The following transitions describe the fate of the E (the center agent in each case) that initially sits just to the left of the string of two A's. A similar analysis holds for the E on the right. An "x" holds the place of an agent whose type we do not have sufficient information to ascertain (see below).

EE E aa	aE E aa	Ea E aa	aa E aa
xE a EE	xE E EE	EE E EE	xE E EE
EE E EE	xx E EE	EE E EE	xx E EE

Moreover, the x's in the final line can be A's only if there existed a string of three or more A's to the left of our segment, to which these agents have now become attached. Hence, any A string of length two disappears after two periods without creating any new A strings.

What of A strings that are of length three or longer? From (1)–(2), the A's at the end of such string are the only potential candidates for becoming E's, and the only way that such a string can increase in length is for a single adjacent E at an end to change to A. Hence, such a string may undergo a change in length of $\{-2, -1, 0, 1, 2\}$. Because the string can increase in length only if it borders a segment of three E's [from (1)], the string cannot merge with any other A strings of length three or more. There are then only two possible fates for such a string. It can persist forever as a distinct string, perhaps varying in length, or its length can fall below three at some point, causing it to be eliminated within the next two periods without giving birth to new strings. We thus have that strings of A's can be destroyed but cannot be created.

Together, these results give: *There exists a time τ such that the number of A strings at time τ is less than or equal to the number of A strings of length three or more at time zero; the number of A strings in any subsequent period is equal to the number at time τ; and all A strings in subsequent periods are length three or longer.*

What can we say about E strings? First, notice that the number of A and E strings must be equal. Next, suppose that time τ has been reached, so that all A strings have length at least three. Then from (1), any E string the length of which is more than two declines in length by two, a string of length two retains its length, and a string of length one increases in length by two. Hence, we will eventually have Egoist strings of length two or blinkers, alternating between lengths one and three, but no longer strings, giving: *There exists a time τ' after which the number of E strings is less than the number of E strings in the initial state and is constant, and E strings either remain at length two or alternate between lengths one and three.* This gives Proposition 1.1. It is now an easy calculation to check that since A strings occur in lengths at least three, and since E strings occur in either length two or alternations between length one and three, that the proportion of A's, if there are to be any A's at all, must be at least 0.6.

PROOF OF PROPOSITION 2:

It is immediate that the system must converge to a state containing Altruists, and hence a state containing at least 60-percent Altruists (by Proposition 1) if there exists a persistent string, and that the probability of a persistent string approaches unity as N gets large if initial identities are randomly, independently determined, with positive probability on Altruist. It then remains to verify the characterization of persistent strings. We examine the case of a string containing at least five adjacent A's. Showing that the remaining strings identified in (2.1) are persistent, and that any other string is eliminated by period three, involves straightforward variations on this argument. (Proposition 1 has already shown that every string of length two or less is eliminated within two periods.)

We show that a string of A's, the length of which is at least five, cannot disappear. In particular, we show that if there exists a string of five A's at time t, then either all five of these agents must also be Altruists at time $t+1$ or they must all be Altruists at time $t+2$. This holds regardless of the strategies played by other agents in the system.

Suppose we have a string of five or more A's bordered on each end by an E. Each of these two E's must have either an A or E on its other side. This gives us four possibilities to consider. First, suppose each E has an A on its other side. Then from (1)–(2), the system proceeds as follows:

$\cdots aE \quad aaaaa \quad Ea \cdots$

$\cdots EE \quad EaaaE \quad EE \cdots$

$\cdots xE \quad aaaaa \quad Ex \cdots.$

As usual, an x holds the place of an agent who may be either an Altruist or an Egoist. For convenience, the original string of five A's is separated by spaces. A similar result clearly holds if the original string contains more than five A's.

Alternatively, one of the E's on the end of the string of A's may have an E on its other side while the other may have an A on its other side. This gives us the following case and its mirror image:

$\cdots aE \quad aaaaa \quad EE \cdots$

$\cdots EE \quad Eaaaa \quad xE \cdots$

$\cdots xE \quad aaaaa \quad xx \cdots.$

Finally, the E's on both ends of the string of A's may be bordered by E's. Then we have:

$\cdots EE \quad aaaaa \quad EE \cdots$

$\cdots Ex \quad aaaaa \quad xE \cdots.$

In each case, the result is that any string of at least five Altruists persists.

PROOF OF PROPOSITION 3:

Let there be countably many agents, denoted by the integers. Consider the initial state, and suppose that, in this state, agent 0 is the rightmost agent of one of the following sequences of agents:

$aaaaaE \quad EEaaaaE \quad EEEaaaE$

$aaaaEE \quad EEaaaEE$

$aaaEEE.$

Let such a string be called a "persistent-string-with-right-boundary." Then we know that contained within agents $\{-6, \cdots, -1\}$ is a persistent Altruist string.

We now let τ be an integer with the property that, in the initial state, the set of agents $\{0, \cdots, \tau\}$ contains a persistent string. Our task is to calculate an upper bound on the expected value of τ.[26] Because the identities of the agents in $\{0, 1, \cdots\}$ are determined independently, each with probability p of being an Altruist, we can describe the initial condition, and hence the expected value of τ, by defining a new Markov process as follows. Let there be 17 states, denoted by $\{1, 2, \ldots, 16, T\}$, where we think of the first 16 of these states as being associated with the following sequences:

1: *aE*	7: *aEaa*	13: *aEaaaE*
2: *EE*	8: *EEaa*	14: *aEaaaa*
3: *EEE*	9: *EEEaa*	15: *aEaaaEE*
4: *aEa*	10: *aEaaa*	16: *aEaaaaE*.
5: *EEa*	11: *EEaaa*	
6: *EEEa*	12: *EEaaaE*	

We define the state at step 0 as being one of states 1, 2, or 3, depending upon whether agents $-2, -1$, and 0 are characterized by xaE ($x \in \{a, E\}$), aEE, or EEE. We define the state at step $t > 0$ to be state T if any agent from the set $\{0, \ldots, t\}$ is the rightmost agent in one of the following sequences:

aaaaa	*EEaaaa*	*EEEaaa*
	aaaaEE	*EEaaaEE*
		aaaEEE.

If the state at step t is not T, then the state is given by $i \in \{1, \ldots, 16\}$ if agent t is the rightmost agent in the sequence corresponding to state i, and the same is true of no other $i \in \{1, \ldots, 16\}$ with a longer sequence. Intuitively, the system begins at step zero in one of states 1, 2, or 3, which are the sequences of Altruists and Egoists that must lie at the right end of a persistent-string-with-right-boundary. We then examine individuals 1, 2, 3, and so on, in each case using that individual's identity as either Egoist or Altruist to define a transition of the new Markov process. The latter enters state T (for "terminal") whenever an individual has been encountered who allows us to confirm the existence of a persistent string. States 1 through 16 are fragments of persistent strings.

The transition probabilities for this new Markov process are calculated on the basis of the assumption that agents in the initial state are independently chosen to be Altruists with probability p and Egoists with probability $1 - p$. It is then straightforward to calculate the expected number of steps to absorption in state T from each of states 1, 2, and 3 (see E. Seneta, 1981 Theorem 4.5). Numerical calculations produce a table matching that given in Proposition 3, where these numbers are the maximum of the expected number of steps from the three initial conditions given by states 1, 2, and 3.[27] These figures are upper bounds on the expected number of agents between the end of a persistent-string-with-right-boundary and the end of the next persistent string to the right. The boundary of the persistent-string-with-right-boundary may contain up to three Egoists, but a persistent string must contain at least three Altruists, so these numbers are also upper bounds on the number of agents *between* two persistent strings of Altruists.

PROOF OF PROPOSITION 4:

Let the types of players be denoted by 1 and 2. Let each player i have a set of players whom he potentially imitates, called his learning neighborhood, and a set of players with whom he interacts, called his interaction neighborhood. In particular, player i's imitation rule is

[26] The assumption that N is infinite makes its appearance here. It allows us to calculate the expected value of τ while ignoring the possibility that the set of agents is exhausted before encountering another persistent string, with our search for such a string having taken us around the circle and back to our point of departure at agent 0.

[27] The details of these calculations are avaiable on request. Differences in the expected number of steps for these three initial conditions are small. The maximum such difference appears for large values of p, and is 2.1 for $p = 0.5$ but only 0.3 for $p = 0.1$.

to adopt the strategy that receives the highest average payoff of the strategies represented in his learning neighborhood.[28] In our model of Altruists and Egoists, the interaction neighborhood of player i included his two nearest neighbors, while his learning neighborhood included these two nearest neighbors and himself. Let N_i^{L1} and N_i^{L2} be the sets of type-1 and type-2 players in agent i's learning neighborhood. Let N_i^{I1} and N_i^{I2} be the sets of type-1 and type-2 players in agent i's interaction neighborhood. Let n_i^{L1}, n_i^{L2}, n_i^{I1}, and n_i^{I2} be the numbers of players in the sets N_i^{L1}, N_i^{L2}, N_i^{I1}, and N_i^{I2}. Let P_i be player i's payoff, and let $n_i^{I1} + n_i^{I2} = n^I$ be the size of the interaction neighborhood. According to the imitation rule, player i will become type 1 if and only if:

$$\text{(A1)} \qquad \frac{1}{n_i^{L1}} \sum_{j \in N_i^{L1}} P_j > \frac{1}{n_i^{L2}} \sum_{j \in N_i^{L2}} P_j.$$

When player j is of type s, then his payoff is given by

$$P_j = K_1 n_j^{I1} + K_2 n_j^{I2} - C_s$$
$$= (K_1 - K_2) n_j^{I1} + K_2 n^I - C_s,$$

and the imitation rule (A1) becomes:

$$\frac{K_1 - K_2}{n_i^{L1}} \sum_{j \in N_i^{L1}} n_j^{I1}$$
$$> (C_1 - C_2) + \frac{K_1 - K_2}{n_i^{L2}} \sum_{j \in N_i^{L2}} n_j^{I2}.$$

It is now obvious that when $K_1 - K_2 > 0$, the dynamics will be identical for all pairs of types (K_s, C_s) for which $(C_1 - C_2)/(K_1 - K_2)$ is the same.

PROOF OF PROPOSITION 5:

Let $\mathcal{E}$ denote the state in which all agents are Egoists. Let $\mathcal{A}$ be the state in which all agents are Altruists. Let $\mathcal{X}(n, m)$ denote the collection of absorbing sets with the property that in any state contained in such an absorbing set, at least some agents are Altruists and all agents are Altruists except n strings of Egoists of length two and m blinkers, where $n \geq 0$ and $m \geq 0$. We define $\mathcal{X}(n, m)$ only for values of (n, m) for which $\mathcal{X}(n, m)$ is nonempty. Then $\mathcal{A}$ is the unique element in $\mathcal{X}(0, 0)$ and every absorbing set other than $\mathcal{E}$ is contained in some $\mathcal{X}(n, m)$.

It suffices to show that $D(\mathcal{A}) < D(\mathcal{E})$, where D is defined in Lemma 3 of Samuelson (1994). For this, it suffices to show that:

- Three mutations suffice to transform $\mathcal{E}$ into a state in the basin of attraction, under the imitation process, of a state in $\mathcal{X}(n, m)$ for some (n, m).
- Given any absorbing set in $\mathcal{X}(n, m)$ with $(n, m) \neq (0, 0)$, there exists a state in the absorbing set which a single mutation can transform into a state in the basin of attraction, under the imitation process, of an absorbing set in $\mathcal{X}(n', m')$ with $n' + m' < n + m$ or with $n' < n$ and $m' \leq m + 1$.
- Given any state in any absorbing set in $\mathcal{X}(n, m)$ for any (n, m), it takes at least $N/10$ mutations to reach a state in the basin of attraction, under the imitation process, of $\mathcal{E}$.

A state is in the basin of attraction, under the imitation process, of an absorbing set, if the deterministic imitation process (without mutations) leads from the state to the absorbing set.

To establish the first condition, we need only note that if three mutations introduce three adjacent Altruists into state $\mathcal{E}$, then Proposition 2 ensures that we then have a state in the basin of attraction $\mathcal{X}(n, m)$ for some (n, m). To establish the second condition, consider an absorbing set S' in $\mathcal{X}(n, m)$. If $m > 0$, then we need only choose a state in S' which at least one blinker has only one Egoist. A mutation switching this Egoist to an Altruist then produces a state in absorbing set in $\mathcal{X}(n, m - 1)$. Hence, consider an absorbing set in $\mathcal{X}(n, 0)$. Now let a mutation switch an Egoist to an Altruist. The result is an isolated Egoist (that was adjacent to the Egoist affected by the mutation). The next iteration of the imitation

[28] Player i is excluded from his interaction neighborhood and included in his learning neighborhood. In the case of Altruists and Egoists, this simply reflects our choice to measure the costs of altruism as the net costs, after any benefits of one's own public good provision have been realized.

process will produce a string of three Egoists. If all Altruist strings are still of length at least three, then we have a blinker and a state in an absorbing set contained in $X(n-1, 1)$. If instead at least one Altruist string is now of length only two, then (from the proof of Proposition 1) that string of Altruists will disappear, while no new string can appear, yielding a state in an absorbing set in $X(n', m')$ with $n' + m' < n$.

Finally, we calculate a lower bound on the number of mutations required to convert a state in an absorbing set in $X(n, m)$ into a state in the basin of attraction of $\mathcal{E}$. The mutations must eliminate all of the strings of Altruists in the original state. We first notice that in order to eliminate a string of A's of length k, we must have at least $[k/5]$—the integral value of $k/5$—mutations.[29] A lower bound on the number of mutations needed to eliminate all string of A's is then $N/10$, which arises in the case in which there are strings of A's of length nine (which are the longest that can still be eliminated by a single mutation) with blinkers at the end of the string, where the blinkers are in phase and there are nine Altruists in the string when each blinker consists of a single Egoist. For sufficiently large N, and in particular for N exceeding 30, this number exceeds three, giving the result.

PROOF OF PROPOSITION 6:

Fix the population size N. Let the Markov process induced by the imitation dynamics be $(S, \mathbf{P})$, where S is the state space and $\mathbf{P}$ is the transition matrix, and let the Markov process induced by the imitation-and-mutation dynamics by (S, Γ), where Γ is the transition matrix. We say that an agent chosen to assess her strategy, under the random imitation dynamics, has "received the learn draw."

Step 1: This step shows that instead of examining (S, Γ), we can work with a simpler Markov process (K, Δ). To construct this simpler process, we let $[N/5]$ denote the integral value of $N/5$ and let the state space $K = \{0, 1, \ldots, [N/5]\}$. We interpret a state $k \in K$ as identifying the number of Egoist strings in an absorbing state of $(S, \mathbf{P})$.[30] The transition matrix is Δ, where Δ_{ij} is the probability that a *single* mutation in (S, Γ), followed by the imitation dynamics, leads from an absorbing set with i Egoist strings to an absorbing set with j Egoist strings. Notice that a mutation can create at most one new Egoist string or can destroy at most one string, and hence can cause the number of Egoist strings to change by at most one. The proportion of Altruists in the limiting distribution of (K, Δ) matches the proportion in the limiting distribution of (S, Γ).

Step 2: We now examine (K, Δ). This is a birth-death process, since from state k, there is positive probability of moving only to states $k-1$, k, and $k+1$. The stationary distribution δ^* of a birth-death process must satisfy the detailed balance condition:

$$\text{(A2)} \qquad \frac{\delta^*(k)}{\delta^*(k+1)} = \frac{\Delta_{k+1,k}}{\Delta_{k,k+1}}.$$

To complete the proof, it suffices to show that there is $\varepsilon > 0$ such that for any N, if $2k/N \leq 0.13$ (recall that each Egoist string contains two Egoists), then $\Delta_{k,k+1}/\Delta_{k+1,k} > 1 + \varepsilon$, and if $2k/N \geq 0.30$, then $\Delta_{k,k+1}/\Delta_{k+1,k} < 1 - \varepsilon$. In particular, this ensures (from A2) that the ratio $\delta^*(k)/\delta^*(k+1)$ is bounded below one when $2k/N \leq 0.13$ and bounded above one when $2k/N \geq 0.30$. As N grows, the number of pairs $(k, k+1)$ with $2k/N \leq 0.13$ and $2k/N \geq 0.30$, and hence the number of pairs for which these bounds on the stationary distribution hold, approaches infinity. This can occur only if the probability attached by

[29] This number is calculated by observing that if an Egoist is placed in the midst of a string of Altruists, the result is a blinker, with three Egoists in the next period. In order to eliminate a string of A's, enough Egoists must be inserted so that after a period has passed and each Egoist given rise to a string of three Egoists, with blinkers possibly also converting the A's at each end of the string into E's, all remaining strings of A of the original string must be at most of length two. This requires at least $[k/5]$ mutations.

[30] The details of this construction, as well as the calculations from Step 3, are available on request. Since any such string must contain at least two Egoists and must be separated from other Egoist strings by at least three Altruists, there can be at most $[N/5]$ such strings.

δ^* to states k such that $2k/N \in (0.13, 0.30)$ approaches unity.[31]

Step 3: This step verifies the required inequalities. Recall that absorbing states consist of strings of two Egoists separated by strings of three or more Altruists. We first calculate a lower bound on $\Delta_{k,k+1}$. A mutation creates a new string of Egoists with probability one if it converts to egoism an Altruist who is bordered by at least four Altruists on each side; with probability between zero and one if the Altruist is bordered by three Altruists on one side and at least four on the other; and otherwise with probability zero. In light of this, we can find a lower bound on the probability of increasing the number of Egoist strings by arranging agents so that there are eight Altruists between each Egoist string, leaving one longer string of leftover Altruists, and assuming that a mutation inserting an Egoist between three Altruists on one side and four on the other never creates a new string of Altruists. The probability of introducing a new Egoist string is then bounded below by the probability that a mutation strikes an agent more than four Altruists away from the end of the long string of Altruists, or

$$\text{(A3)} \qquad \underline{\Delta}_{k,k+1} = \frac{1}{N}(N - 10k).$$

A similar calculation shows that the probability of introducing a new Egoist string is maximized if strings of Egoists are separated by strings of only three Altruists, giving an upper bound of:[32]

$$\text{(A4)} \qquad \bar{\Delta}_{k,k+1} = \frac{1}{N}(N - 5k - 3).$$

We now turn to the probability of eliminating Egoist strings. An upper bound on the probability of eliminating such a string is:

$$\text{(A5)} \qquad \bar{\Delta}_{k,k-1} = \frac{1}{N}5k.$$

A lower bound on the probability of eliminating Egoist strings is given by:

$$\text{(A6)} \qquad \underline{\Delta}_{k,k-1} = \frac{1}{N}\frac{8}{9}2k.$$

We use these calculations to obtain:

$$\frac{\underline{\Delta}_{k,k+1}}{\bar{\Delta}_{k+1,k}} = \frac{N - 10k}{5(k+1)} > 1 + \varepsilon$$

if $k/N \leq 0.065$ (and hence there are no more than 13-percent Egoists), N is sufficiently large, and $\varepsilon < 0.075$. Similarly,

$$\frac{\bar{\Delta}_{k,k+1}}{\underline{\Delta}_{k+1,k}} = \frac{N - 5k - 3}{2(k+1)}\frac{9}{8} < 1 - \varepsilon$$

if $k/N \geq 0.15$ (and hence there are at least 30-percent Egoists), N is sufficiently large, and $\varepsilon < 0.06$. This gives the result.

[31] See Ken Binmore and Samuelson (1997) for a similar argument.

[32] The "-3" reflects a three-Altruist buffer at both ends of any long string of A's in which a mutation cannot create a new string of Egoists.

REFERENCES

Benson, Woodruff W. "Evidence for the Evolution of Unpalatability by Kin Selection in the Heliconiinae (Lepidoptera)." *American Naturalist*, May–June 1971, *105*(943), pp. 213–26.

Bergstrom, Theodore C. and Stark, Oded. "How Altruism Can Prevail in an Evolutionary Environment." *American Economic Review*, May 1993 (*Papers and Proceedings*), *83*(2), pp. 149–55.

Binmore, Ken and Samuelson, Larry. "Muddling Through: Noisy Equilibrium Selection." *Journal of Economic Theory*, June 1997, *74*(2), pp. 235–65.

Blume, Lawrence E. "The Statistical Mechanics of Strategic Interaction." *Games and Economic Behavior*, July 1993, *5*(3), pp. 387–424.

Boyd, Robert and Richerson, Peter J. *Culture and the evolutionary process*. Chicago: University of Chicago Press, 1985.

Brower, Lincoln Pierson. "Ecological Chemistry." *Scientific American*, February 1969, *220*(2), pp. 22–29.

Brower, Lincoln Pierson and Brower, Jane Van Zandt. "Birds, Butterflies and Plant Poisons: A Study in Ecological Chemistry."

Zoologica, November 2, 1964, *49*(3), pp. 137–59.

Christian, John J. "Social Subordination, Density and Mammalian Evolution." *Science*, April 3, 1970, *168*(3927), pp. 84–90.

Cohen, Dan and Eshel, Ilan. "On the Founder Effect and the Evolution of Altruistic Traits." *Theoretical Population Biology*, December 1976, *10*(3), pp. 276–302.

Dawkins, Richard. *The selfish gene*. Oxford: Oxford University Press, 1976.

Ellison, Glenn. "Learning, Local Interaction, and Coordination." *Econometrica*, September 1993, *61*(5), pp. 1047–72.

Eshel, Ilan. "On the Neighbor Effect and the Evolution of Altruistic Traits." *Theoretical Population Biology*, September 1972, *3*(3), pp. 258–77.

Eshel, Ilan and Cavalli-Sforza, L. L. "Assortment of Encounters and Evolution of Cooperativeness." *Proceedings of the National Academy of Sciences*, February 1982, *79*(4), pp. 1331–35.

Fenner, Frank. "Myxoma Virus and Cryctologus Cuniculus: Two Colonizing Species," in H. G. Baker and G. Ledyard Stebbins, eds., *The genetics of colonizing species*. New York: Academic Press, 1965, pp. 485–99.

Freidlin, M. I. and Wentzell, A. D. *Random perturbations of dynamical systems*. New York: Springer-Verlag, 1984.

Fudenberg, Drew and Maskin, Eric. "The Folk Theorem in Repeated Games with Discounting and Incomplete Information." *Econometrica*, May 1986, *54*(3), pp. 533–54.

Gilboa, Itzhak and Schmeidler, David. "Case-Based Decision Theory." *Quarterly Journal of Economics*, August 1995, *110*(3), pp. 605–40.

———. "Case-Based Optimization." *Games and Economic Behavior*, July 1996, *15*(1), pp. 1–26.

Hamilton, W. D. "The Genetic Evolution of Social Behavior." *Journal of Theoretical Biology*, January 1964, *7*(1), pp. 1–52.

———. "Altruism and Related Phenomena." *Annual Review of Ecological Systems*, 1972, *3*, pp. 193–232.

Huberman, Bernardo A. and Glance, Natalie S. "Evolutionary Games and Computer Simulations." *Proceedings of the National Academy of Sciences*, August 15, 1993, *90*(16), pp. 7716–18.

Kandori, Michihiro; Mailath, George J. and Rob, Rafael. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica*, January 1993, *61*(1), pp. 29–56.

Kemeny, John G. and Snell, J. Laurie. *Finite Markov chains*. Princeton, NJ: D. Van Nostrand, 1960.

Kreps, David M.; Milgrom, Paul; Roberts, John and Wilson, Robert J. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, August 1982, *27*(2), pp. 245–52.

Lorenz, Konrad. *On aggression*. Harcourt, Brace & World, 1963.

Matessi, C. and Jayakar, S. D. "Conditions for Altruism Under Darwinian Selection." *Theoretical Population Biology*, June 1976, *9*(3), pp. 360–87.

Maynard Smith, John. "Group Selection." *Quarterly Review of Biology*, June 1976, *51*(2), pp. 277–83.

———. *Evolution and the theory of games*. Cambridge: Cambridge University Press, 1982.

Nowak, Martin A.; Bonhoeffer, Sebastian and May, Robert M. "More Spatial Games." *International Journal of Bifurcation and Chaos*, February 1994, *4*(1), pp. 33–56.

Nowak, Martin A. and May, Robert M. "Evolutionary Games and Spatial Chaos." *Nature*, October 29, 1992, *359*, pp. 826–29.

———. "The Spatial Dilemmas of Evolution." *International Journal of Bifurcation and Chaos*, February 1993, *3*(1), pp. 35–78.

Samuelson, Larry. "Stochastic Stability in Games With Alternative Best Replies." *Journal of Economic Theory*, October 1994, *64*(1), pp. 35–65.

Seneta, E. *Non-negative matrices and Markov chains*. New York: Springer-Verlag, 1981.

Stewart, Frank M. and Levin, Bruce R. "The Population Biology of Bacterial Viruses: Why Be Temperate?" *Theoretical Population Biology*, August 1984, *26*(1), pp. 93–117.

Trivers, Robert L. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, March 1971, *46*(1), pp. 35–57.

Williams, George C. *Adaptation and natural selection*. Princeton, NJ: Princeton University Press, 1966.

Wilson, David Sloan. "A Theory of Group Selection." *Proceedings of the National Academy of Sciences*, January 1975, *72*(1), pp. 143–46.

———. "Altruism in Mendelian Populations Derived from Sibling Groups: The Haystack Model Revised." *Evolution*, September 1987, *41*(5), pp. 1059–70.

Wynne-Edwards, Vero Copner. *Animal dispersion in relation to social behavior*. Edinburgh: Oliver and Boyd, 1962.

———. *Evolution through group selection*. Oxford: Blackwell Scientific Publications, 1986.

Young, H. Peyton. "The Evolution of Conventions." *Econometrica*, January 1993, *61*(1), pp. 57–84.