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 inescapable as it is tautological. If someone commits an “altruistic” act, then this reveals that he prefers doing so. 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. 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 primarily 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).
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  
\[ N_i^a - C \text{ if } i \text{ is an Altruist and } N_i^a \text{ if } i \text{ is an Egoist, where } N_i^a \in \{0, 1, 2\} \text{ is the number of } i \text{'s Altruist neighbors (excluding himself).} \]

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.  
\[ \text{If an agent and her two neighbors all play the same strategy, be it Altruist or Egoist, then the agent will continue to play that strategy.} \]

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, and David Schneider (1993), Mark S. Gell-Mann and Jack Layard (1993), and P. K. are 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 state 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 \( P_{aa} = 1 \) and \( i \) is a stationary state of the Markov process. An absorbing set may contain more than one state, in which case \( P_{aa} = 0 \) 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 \( C < \frac{1}{2} \). 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 learns to form a circle. From (1) to (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:
  \[ a a a a a \]

- A set of two states, consisting of:
  \[ a a a a a a \]

  \[ a a a a a a \]

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 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 \( \alpha \) and \( \beta \) play the same strategy, either Altruist or Egoist, and if all agents between \( \alpha \) and \( \beta \) play this strategy, then we will refer to agents \( \alpha, \beta \) and the intermediate agents as an interval of
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.

\[
\begin{pmatrix}
K_1 & K_2 \\
C_1 & C_2
\end{pmatrix}
\]

\[
= \begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix} \cdot \begin{pmatrix}
0 & -1 \\
-1 & 0
\end{pmatrix} \cdot \begin{pmatrix}
0 & -1 \\
-1 & 0
\end{pmatrix} = \begin{pmatrix}
1 & -1 \\
-1 & 2
\end{pmatrix} \cdot \begin{pmatrix}
1 & -1 \\
-1 & 2
\end{pmatrix}.
\]

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 incurs 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 incurs 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 \( 2 \times 2 \) 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 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 \( 2 \times 2 \) games. Then an argument analogous to the proof of Proposition 4 shows that the imitation dynamics depends only upon the two numbers:

\[
\alpha = \frac{c - b}{a - b}, \quad \beta = \frac{d - b}{a - b}.
\]

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 \( \alpha \) and \( \beta \), where \( \beta < 1 \) (because we have assumed \( a > d \)). We have:

- Prisoner's Dilemma: \( 0 < \beta < 1 \), \( 1 < \alpha \).
- Coordination Games: \( \beta < 1 \), \( a < 1 \).
- Chicken: \( \beta < 1 \), \( a < 1 \).
- Efficient Dominant Strategy: \( \beta < 0 \), \( a < 1 \).

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. 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 \( a + \beta < 1 \), while the equilibrium \((Y, Y)\) is risk dominant if \( a + \beta > 1 \). The interval in which \( \alpha = 1 + \beta C \), \( \beta = \frac{1}{2} C \), and \( C < 1/4 \) shown in Figure 3, describes the range of Altruist and Egoist games that was analyzed in subsections A and B of this section.

\(^{14}\) 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 easiest to read if we use a lower case a to represent the strategy X.
The American Economic Review

VOL. 81 NO. 3
MARCH 1991

QUESTIONS: ACTIVITY, ROOSTERS, AND NOODLES

III. Model

The model in this paper is an extension of the model in

\[ g > 0 \]

where \( g \) is the growth rate of technology and \( \eta \) is the fraction of workers engaged in the service sector. The model is solved numerically using a finite difference method.

In the model, the allocation of workers between the service and manufacturing sectors is determined by the relative productivity of these sectors. The service sector is more productive and has a higher wage, leading workers to choose jobs in that sector. The model also includes a mechanism for the adjustment of the wage rate in response to changes in the economy.

The results of the model suggest that the growth of the economy is driven by the accumulation of capital and the development of new technologies. The model also shows that the allocation of workers between sectors is determined by the relative productivity of these sectors.

IV. Conclusion

The results of the model suggest that the growth of the economy is driven by the accumulation of capital and the development of new technologies. The model also shows that the allocation of workers between sectors is determined by the relative productivity of these sectors.

In conclusion, the model provides a framework for understanding the dynamics of economic growth and the allocation of workers between sectors. The model is useful for policy makers and economists who are interested in understanding how the economy works.
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 \( \mu \) can be arbitrarily close to one.\(^6\)

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 Egoist 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 \((\mu_1, \mu_2)\) and \((\mu_3, 1)\). Changing the value of \( C \) within each such an interval does not affect the outcome, while we shall see that the two intervals give different behavior.\(^6\)

We let \( \mu = 1 \) throughout.

The investigation of the model with neighborhoods of radius two and costs \( C \in (\mu_1, 1) \) 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

\[ x \text{ stands for a strategy that could be either } A \text{ or } E. \]

\[
\begin{align*}
Eaaa & a xEEE \\
avv & a EEExx \\
avv & a aEEE.
\end{align*}
\]

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

\[
\begin{align*}
EaaaE & E EEEE \\
avv & a EEE \\
avv & a E EEEE.
\end{align*}
\]

These allow us to prove:\(^{21}\)

**PROPOSITION 7:** Let neighborhoods be of radius two and let \( \mu_1 < C < 1 \). 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 A and three 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 \( \mu_2 \).

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:

\[
\begin{align*}
\ldots & aaaaEEaaaaaaaEEEEaaa \\
\ldots & aaaaaEEaaaaaaaEEEEeeaaa \\
\ldots & aaaaaEEaaaaaaaEEEEeeaa \\
\ldots & aaaaaEEaaaaaaaEEEEee
\end{align*}
\]

This has \( \mu_2 \) 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 \( C \in (\mu_1, \mu_2) \) than in the case of higher costs, being \( \mu_2 \) rather than \( \mu_1 \). 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

\[ \text{When agents interacted only with their immediate neighbors, the only relevant cost consideration was whether } C \text{ was larger or smaller than } \mu_2. \]

\[ \text{If } C > \mu_2, \text{ then altruism is so costly that only Egoists survive. The results for } 1 < C < \mu_2 \text{ are qualitatively similar to those for } \mu_1 < C < 1, \text{ with the quantitative differences noted below. Cost levels } \mu_3 < C < \mu_2 \text{ give results similar to those of } \mu_1 < C < \mu_2. \]

\[ \text{Costs } C < \mu_2 \text{ give noticeable different and more complicated behavior that we do not investigate here. Cost levels that lie at the boundary between these various intervals create complications arising out of cases in which the average payoffs to Altruists and Egoists are equal.} \]

\[ \text{The "generally" in this statement allows us to avoid values of } C \text{ within the interval } (\mu_1, 1) \text{ that create payoff ties between Altruists and Egoists. For } 1 < C < \mu_2 (C \neq \mu_2), \text{ we have the same characterization of absorbing sets, except that blinkers must be separated by at least six Altruists, making the minimal percentage of Altruists } 0.6. \]

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

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 1. Because the string can increase in length only if it borders a segment of three E’s (from (11)), 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 \( \tau \) such that the number of A strings at time \( \tau \) 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 \( \tau \); and all A strings in subsequent periods are length three or longer.

---

**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. If 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 \( \bar{A} \) on its outside. This gives us four possibilities to consider. First, suppose each E has an A on its outside. Then from (1) – (2), the system proceeds as follows:

\[
\ldots, aE aaaa aE, EEE, aaaa aE, EE Eaaa xE, EEEEE, xE aaaa aEx
\]

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 outside while the other may have an A on its outside. This gives us the following case and its mirror image:

\[
\ldots, aE aaaa aE, EEE, aaaa aE, EE Eaaa xE, EEEEE, xE aaaa aEx
\]

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

\[
\ldots, aE aaaa aE, EEE, aaaa aE, EE Eaaa xE, EEEEE, aEx aaaa xE
\]

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:

\[
aaaaaE, EEEaaaE, aaaaEE, EEEaaaE, aaaaEE
daE.
\]
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 \( X(n, m) \) for \( n' + m' < n \).

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 \( T \). 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 \( \lfloor k/5 \rfloor \) —the integral value of \( k/5 \)—mutations.\(^{27}\) 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 \( (S, P) \), where \( S \) is the state space and \( P \) is the transition matrix, and let the Markov process induced by the imitation-and-mutation dynamics by \( (S, P) \), where \( S \) is the transition matrix. We say that an agent chosen to assess her strategy, under the random imitation dynamics, has "received the learn draw."\(^{28}\)

Step 1: This step shows that instead of examining \( (S, P) \), we can work with a simpler Markov process \( (K, \Delta) \). 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 \( (S, P) \).\(^{29}\) The transition matrix is \( \Delta \), where \( \Delta \) is the probability that a single mutation in \( (S, P) \), followed by the imitation dynamics, leads from an absorbing set with \( j \) 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 change the number of Egoist strings to change by at most one. The proportion of Altruists in the limiting distribution of \( (K, \Delta) \) matches the proportion in the limiting distribution of \( (S, P) \).

Step 2: We now examine \( (K, \Delta) \). 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 \( \delta^* \) of a birth-death process must satisfy the detailed balance condition:

\[
\frac{\delta^*(k)}{\delta^*(k+1)} = \frac{\Delta_{k+1}}{\Delta_{k}}.
\]

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 \( \Delta_{k+1}/\Delta_{k} > 1 + \varepsilon \), and if \( 2k/N \leq 0.30 \), then \( \Delta_{k+1}/\Delta_{k} < 1 - \varepsilon \). In particular, this ensures (from \( A2 \)) that the ratio \( \delta^*(k)/\delta^*(k+1) \) is bounded below by some \( \varepsilon > 0 \) for all \( k \), and bounded above by some \( \varepsilon < 1 \) for all \( k \). As \( N \) grows, the number of pairs \( (k, k+1) \) with \( 2k/N \leq 0.13 \) and \( 2k/N \geq 0.30 \) grows, and hence the number of pairs for which these bounds on the stationary distribution hold, approaches infinity. This can only occur if the probability attached by \( \delta^* \) to states \( k \) such that \( 2k/N \in (0.13, 0.30) \) approaches unity.\(^{30}\)

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 \( \Delta_{k+1} \). 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 string of Altruists, or

\[
\frac{\delta^*(k)}{\delta^*(k+1)} = \frac{\Delta_{k+1}}{\Delta_{k}} \geq \frac{1}{N} (N - 10k).
\]

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:

\[
\frac{\delta^*(k)}{\delta^*(k+1)} = \frac{\Delta_{k+1}}{\Delta_{k}} \leq \frac{1}{N} (N - 5k - 3).
\]

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

\[
\Delta_{k+1} \leq \frac{1}{N} (N - 5k).
\]

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

\[
\Delta_{k+1} = \frac{1}{N} (N - 5k).
\]

We use these calculations to obtain:

\[
\Delta_{k+1} = \frac{1}{N} (N - 10k) > 1 + \varepsilon
\]

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

\[
\Delta_{k+1} = \frac{1}{N} (N - 5k - 3) < 1 - \varepsilon
\]

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

**REFERENCES**


