

# How Altruism Can Prevail Under Natural Selection

by

Ted Bergstrom and Oded Stark

*University of Michigan and Harvard University*

Current version: March 22, 2002

# How Altruism Can Prevail Under Natural Selection

Ted Bergstrom and Oded Stark

## Introduction

Why have we economists been convinced for so long that our old friend, *homo economicus* must be selfish? No doubt we find considerable support for this hypothesis in the behavior of our colleagues. We might also expect that evolutionary pressures tend to produce selfish behavior—with the notable exception of the relation between parents and offspring. But can we expect natural selection to act in favor of altruistic behavior in other relationships?

Evolutionary biologists have created a theory that predicts altruistic behavior, not only between parents and children, but also among siblings and other close relatives.<sup>1</sup> Richard Dawkins' expression of this view in *The Selfish Gene*, is that the replicating agent in evolution is the *gene* rather than the animal. If a gene carried by one animal is likely to appear in its relatives, then a gene for helping one's relatives, at least when it is cheap to do so, will prosper relative to genes for totally selfish behavior.

This paper presents a series of examples in which natural selection sustains cooperative behavior in single-shot prisoners' dilemma games. In prisoners' dilemma, cooperation always gets a lower payoff for oneself and a higher payoff for one's opponent than defection. Therefore it seems appropriate in this simple case to identify *altruism* with playing *cooperate* in prisoners' dilemma.<sup>2</sup> The reason that cooperative behavior toward siblings can be sustained even where defection is a dominant strategy, is that an individual who has a gene for cooperating with its siblings has a good chance of benefiting from the presence

---

<sup>1</sup> See, for example, William Hamilton (1964a, 1964b), Richard Dawkins (1976), John Maynard Smith (1982), and Robert Trivers (1985).

<sup>2</sup> More subtle questions about the nature of altruistic preferences are left for other investigations. Each of us has done some work of this kind. See B. Douglas Bernheim and Oded Stark (1988), Oded Stark (1989), and Ted Bergstrom (1988), (1989), (1992).

of the same gene in its siblings. Similar reasoning applies to behavior that is imitative rather than genetically inherited if those who share common role models are more likely to interact with each other than with randomly selected members of the population.

## 1. The Game Creatures Play and the Nature of Equilibrium

Individuals will be assumed to play one-period, two-person games of prisoners' dilemma with their siblings or neighbors. In each game that it plays, an individual can choose one of two strategies, *cooperate* or *defect*. The payoffs from this game are listed in the matrix below. If the parameters satisfy the restriction  $S < P < R < T$ , then *defect* will be a dominant strategy for each game. For the game to be called a "prisoners' dilemma", it should also satisfy the restriction that  $S + T < 2R$ .

### Prisoners' Dilemma

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	$R, R$	$S, T$
	Defect	$T, S$	$P, P$

Total payoff to an individual will be the average of its payoffs in the prisoners' dilemma games that it plays. Where behavior is genetically inherited, we assume that the expected number of surviving offspring that an individual produces will be higher, the higher its total payoff. Where behavior is copied from neighbors, the probability that an individual's behavior is copied will depend on its payoff.

A population can have either a *monomorphic equilibrium* or a *polymorphic equilibrium* (or possibly both). In a stable monomorphic equilibrium, only one type of individual is present and if a mutant individual of the other type should arise, it must reproduce less rapidly than normal individuals. In a polymorphic equilibrium, more than one type

of individual is present and each type that is present receives the same expected payoff. Stability of polymorphic equilibrium requires that if one type happens to become more common than the equilibrium proportion, it will have a lower expected payoff than the other type.

## 2. Evolution of Genetically Transmitted Behavior

Since little is known about the environments which shaped our genetic inheritance, the evolutionary hypothesis may not be very informative about many aspects of our preferences. But the fundamental processes of mating, child-rearing and relations between siblings appear to have changed little over the millennia. Accordingly, we may learn a good deal about the “economics of the family” from a look at the evolutionary theory of relations among kin.<sup>3</sup>

### *Altruistic Sororities Without Sex*

Just to help us understand the logic of inheritance, we begin with a toy model that seems unrealistic for humans—*asexual reproduction*. Let us assume that any individual will, if she survives to reproductive age, have exactly two children, whose genes are just like her own (except in the case of rare mutations). A surviving individual with a gene for *cooperate* will have two offspring with genes for *cooperate*. A surviving individual with a gene for *defect* will have two offspring like herself.

To keep the population size constant, we must assume that only half of the individuals who are born will survive to reproductive age and that the probability that any individual will survive to reproduce will be higher, the greater her payoff in the game that she plays with her sister. We claim that the only equilibrium is a population consisting entirely of cooperators.

---

<sup>3</sup> We find ourselves in good company in this heresy. Becker (1976, 1981) and Hirshleifer (1977, 1978), explore genetic explanations for altruism among close family members. Robert Frank (1985) seeks an evolutionary explanation for human emotions, and Arthur Robson (1992a, 1992b) explores an evolutionary explanation of human attitudes toward risk.

To see this, consider a population consisting entirely of cooperators. What would happen to a mutant defector that appeared in this population? Since her sister is a cooperator, the mutant gets a payoff of  $T > R$ , and so she is more likely to survive than any other member of the population. But her good fortune will not be sustained by her descendants. Her daughters will both inherit her *defect* gene and will both defect. Sisters in each generation of her descendants will also defect and get  $P < R$ , and hence gradually disappear from the population.

Similarly, in a population of defectors, a mutant cooperator would face a defecting sister and would get a payoff of  $S$ , while the surrounding defectors would get  $P > S$ . Although her survival probability will be lower than the population average, her daughters and their descendants would all be cooperators. Each of them will receive a payoff of  $R > P$  and their numbers would grow relative to those of the defectors.

#### *Diploid Siblings are Sometimes Altruistic (But Not as Often as Their Parents Would Like)*

Diploid parents will not be surprised to discover that in our own species, siblings are not always as cooperative as asexual siblings would be. We will show that there is a rich menu of possible equilibria with diploid siblings. Depending on the parameters of the prisoners' dilemma game, there may be a unique stable monomorphic equilibrium with cooperators only, a unique stable monomorphic equilibrium with defectors only, or there may be two locally stable monomorphic equilibria—one with cooperators only and the other with defectors only. Finally, there are parameter values for which there are no stable monomorphic equilibria, but for these parameters there will be a stable polymorphic equilibrium with some cooperators and some defectors in the population.

We consider a large population which reproduces sexually and has diploid genetic structure. Each individual plays a single-shot game of prisoners' dilemma with each of its siblings. To simplify exposition, we will assume that each individual who survives to mate and reproduce has exactly three offspring. The probability that an individual survives to reproduce will be higher, the higher the total payoff that it gets in the games it plays with

its two siblings.<sup>4</sup> Individuals are able to distinguish their siblings from other members of the population and may use different strategies in games played with siblings from the strategies used with outsiders.

The strategy that any individual uses in play with its siblings is determined by the contents of a single *genetic locus*. This locus contains two genes, one randomly selected from each of its parents' two genes. For the present discussion, we assume that mating is monogamous and random with respect to the genes controlling behavior toward sibling.

We assume that there are two kinds of genes, a *c* (cooperate) gene and a *d* (defect) gene. Then there will be three possible types of individuals, namely type *cc* homozygotes who carry two *c* genes, type *cd* heterozygotes who carry one *c* gene and one *d* gene, and type *dd* homozygotes who carry two *d* genes. Type *cc* homozygotes always play *cooperate* and type *dd* homozygotes always play *defect*. If heterozygotes always defect, then the *d* gene is said to be *dominant* and the *c* gene is said to be *recessive*. If heterozygotes always cooperate, then the *c* gene is dominant and the *d* gene is recessive.<sup>5</sup>

In this paper, we confine our attention to monomorphic equilibria which are stable against invasion by dominant mutant genes. Thus, we will consider whether a population consisting entirely of *cc* homozygotes could be “invaded” by mutant “dominant” *d* genes such that *cd* heterozygotes always play defect. Similarly, we will ask whether a population consisting entirely of *dd* homozygotes could be invaded by dominant *c* genes.<sup>6</sup> The possibility of invasion by recessive mutants leads to an interesting, but rather elaborate

---

<sup>4</sup> A richer model would have  $n$  siblings play a general  $n$ -person game rather than have each individual play two separate two-person games with its two siblings. Such a model could focus on questions of returns to scale within families. The assumption that all individuals who survive to mate and reproduce allows us to sidestep the complications that would arise from reconciling the assumption of monogamous random mating with the assumption that the number of children has depends on ones payoffs in prisoners' dilemma. What should we do if husband and wife have different expected number of offspring?

<sup>5</sup> Alternatively, one could assume that heterozygotes play a mixed strategy with some fixed probabilities of cooperation and defection.

<sup>6</sup> Notice that we do not take the view that either *c* genes or *d* genes must be intrinsically dominant. Instead we ask whether in a monomorphic population, if a dominant mutant of the opposite type should arise, the mutant strain would increase as a proportion of the population or would ultimately disappear.

analysis which will not be pursued here.<sup>7</sup>

First let us ask when a population consisting only of cooperators would be resistant to invasion by mutants with a (dominant) gene for defection. Suppose that the entire population consists of type  $cc$  homozygotes, all of whom cooperate. Now let some individual experience a mutation which changes one of its  $c$  genes to a  $d$  gene which is dominant over the type  $c$  gene. The mutant will therefore be a  $cd$  heterozygote and will play *defect*. In the games it plays with its siblings, this mutant will get a higher payoff than normal members of the population, since its normal siblings cooperate while it defects. Therefore the mutant receives  $T$  in each game while ordinary members of the population receive  $R < T$ . But in order to find out whether the mutant type will invade the population in the long run, we must follow the fortunes of its offspring who inherit the mutant gene.

When the mutant  $cd$  type is rare, it will almost certainly mate with a normal type  $cc$ . The mutant's offspring will therefore be of type  $cd$  with probability  $1/2$  and of type  $cc$  with probability  $1/2$ . An offspring of the mutant who carries the mutant gene will be of type  $cd$  and will play *defect*. With probability  $1/2$ , a randomly chosen sibling of this individual will be a type  $cc$  homozygote and with probability  $1/2$ , that sibling will be another type  $cd$ . Therefore with probability  $1/2$ , this individual can exploit a cooperative sibling and receive a payoff of  $T$ , but with probability  $1/2$ , the sibling will also defect. It then follows that the expected payoff to each heterozygote offspring of the mutant is  $(T + P)/2$ . The offspring of normal  $cc$  types will receive a payoff of  $R$  in the games they play with their siblings. It follows that while the mutant gene is rare, carriers of the mutant gene will reproduce more rapidly than normal individuals if  $T + P > 2R$  and less rapidly if  $T + P < 2R$ .

Now let us ask when a population consisting entirely of type  $dd$  individuals could be invaded by a mutant  $c$  gene where the mutant gene is dominant over the normal genes. A single mutating gene would first appear in a  $cd$  heterozygote. Assuming the  $c$  gene is

---

<sup>7</sup> A heterozygote with a recessive mutant gene will act just like the normal population and so there will be no selection either for or against recessive genes until "genetic drift" produces enough mutant heterozygotes so that they occasionally mate, thereby producing homozygotes who act differently from the remaining population. Equilibria which are resistant to invasion both by dominant and by recessive mutants are studied by Bergstrom (1992).

dominant, the mutant individual would cooperate. The mutant individual would receive the low payoff  $S$ , since it plays its siblings who play *defect*. But on average, its offspring will do better than  $S$  and perhaps will do better than the normal population of defectors, all of whom receive  $P$ .

When the mutant type is rare, a mutant will almost certainly mate with a normal  $dd$  type. Half of the mutant's offspring will be  $cd$  heterozygotes, who cooperate, and half of them will be  $dd$  homozygotes, who defect. An offspring of the mutant who carries the mutant gene will be of type  $cd$  and will play *cooperate*. With probability  $1/2$ , a randomly chosen sibling of this individual will be a type  $dd$  homozygote who defects and with probability  $1/2$ , that sibling will be another type  $cd$  who cooperates. Therefore with probability  $1/2$ , an offspring that carries of the mutant gene will be exploited by its sibling and get a payoff of  $S$ , but with probability  $1/2$ , its sibling will also cooperate and each of them will receive a payoff of  $R$ . The expected payoff to a type  $cd$  offspring of the mutant is therefore  $(S + R)/2$ . This payoff will be smaller than the payoff to normal  $dd$  types if  $2P - S - R > 0$  and larger if the inequality is reversed.

As we have shown, there will be a stable monomorphic equilibrium with all cooperators if  $T + P - 2R > 0$  and there will be no such equilibrium if the inequality is reversed. There will be a stable monomorphic equilibrium with all defectors if  $2P - S - R > 0$  and no such equilibrium if the inequality is reversed. It turns that there are prisoners' dilemma games where each of these inequalities takes either sign.

The possibilities are illustrated in Figure 1. In this figure we have normalized the game to set  $S = 0$  and  $T = 1$ .<sup>8</sup> With this normalization, there will be a stable monomorphic equilibrium with type  $cc$  only if  $R > (P + 1)/2 > 0$  and there will be a stable monomorphic equilibrium with type  $dd$  only if  $R < 2P$ . For parameter values in Region C of Figure 1 there is a stable monomorphic equilibrium with cooperators only and no stable equilibrium with defectors only. For parameter values in Region D of Figure 1, there is a stable monomorphic equilibrium with defectors only and no stable equilibrium with cooperators

---

<sup>8</sup> This can be done without loss of generality, since the population dynamics discussed in this paper are invariant to affine transformations of the payoff matrix.

only. For parameter values in Region B, there will be stable monomorphic equilibria of both types, and for parameter values in Region A, there will not be a stable monomorphic equilibrium of either type.

In order for the game to be a prisoners' dilemma, it must also be that  $R > P$  and that  $R > .5$ . The region in Figure 1 above the two dotted lines satisfies these conditions. We note that each of the regions A, B, C, and D can occur with parameters suitable for prisoners' dilemma.

Figure 1 justifies the claims made in the title of this section. For prisoners' dilemma games with parameter values in Region C, diploid siblings will cooperate, even though it is to their selfish advantage to defect. For prisoners' dilemma games with parameter values in Region D, diploid siblings will both defect, although parents who wish them to maximize their joint payoff would prefer them both to cooperate.

It is interesting to consider Regions A and B. For parameter values in Region B, there are two stable equilibria—one with a monomorphic equilibrium of each kind. For parameter values in Region A, there are no stable monomorphic equilibrium. In order to understand these cases, it is necessary to work out the detailed laws of motion for the dynamical system that results from this model. This is done by Bergstrom and Bergstrom (1992), where it is found that for parameter values in Region A, there exists exactly one stable polymorphic equilibrium and for parameter values in Region B, there is one unstable polymorphic equilibrium and no stable polymorphic equilibria.

### **3. When Children Imitate their Parents or Teachers**

Here we study a model in which behavior is acquired by imitation, rather than genetically. The model discussed here is a variant of models of cultural transmission which were developed by Cavalli-Sforza and Feldman (1980), and Boyd and Richardson (1985).

We assume that each individual has two siblings and plays a game of prisoners' dilemma with each of them. We will also assume that the probability that any individual survives to mate and reproduce is proportional to the average payoff that it receives in the games it plays with its siblings.

Assume that with probability  $v$ , a child adopts the strategy that was used by a randomly chosen one of its two parents and with probability  $1 - v$  it adopts the strategy used by a nonparent, randomly selected from the entire population.

We assume that marriage is monogamous, so that all siblings share the same mother and father. Parent-couples can be one of three possible types; two-cooperator couples, “mixed couples” with one cooperator and one defector, and two-defector couples. Mating is said to be *assortative* if adults always mate with individuals of their own type.<sup>9</sup> Let  $x$  be the fraction of the adult population who are cooperators. If marriage is purely random, the fraction of all marriages which are mixed couples will be  $2x(1 - x)$ . We define a parameter  $m$  where  $0 \leq m \leq 1$  in such a way as to allow mating patterns that lie between the polar cases of purely random ( $m = 0$ ) and purely assortative ( $m = 1$ ) mating. In the population at large, the proportion of mixed couples is  $2(1 - m)x(1 - x)$ , the proportion of two-cooperator couples is  $x^2 + mx(1 - x)$ , and the proportion of two-defector couples is  $(1 - x)^2 + mx(1 - x)$ .<sup>10</sup>

Given the proportions of couples of each type, we can determine the proportions of all sibling pairs consisting, respectively, of two cooperators, one cooperator and one defector, and two defectors. This enables us to determine not only the proportion of offspring of each type, but also the expected payoffs to offspring of each type, since we will know the probability that a randomly chosen sibling of an individual of each type will be a cooperator or a defector. With this information, we are able to determine the relative growth rates of the population of cooperators and of defectors. The details of this process are worked out in the Appendix of this paper.

This model turns out to have a remarkably convenient mathematical structure. The rate of change of the number of surviving individuals of each type in any generation turns

---

<sup>9</sup> If mating requires mutual consent, if types are costlessly recognizable and search costs are negligible, this would be a natural outcome, since cooperators can expect more offspring if they mate with other cooperators than if they mate with defectors so long as  $2R > S + P$ .

<sup>10</sup> Cavelli-Sforza and Feldman attribute this parameterization of assortative mating to Sewall Wright (1921). These proportions would be achieved if couples were first randomly matched and then the fraction  $m$  of the mixed couples were broken up and the freed individuals paired with persons of their own type.

out to be a linear function of the fraction of the population in the parent generation who are cooperators. Therefore the difference between the growth rates of the number of individuals of the two types is also a linear function of the fraction of the population who are cooperators.<sup>11</sup>

In particular, the difference between the growth rates of the population of cooperators and the population of defectors is expressed by  $D(x) = A + Bx$ , where

$$A = v^2(1 + m)(R - S) - 2(P - S)$$

and where

$$A + B = v^2(1 + m)(T - P) - 2(T - R).$$

Depending on the parameter values,  $v$ ,  $m$ ,  $S$ ,  $P$ ,  $R$ , and  $T$ , the dynamics of this system falls into one of the following four qualitatively distinct cases.

- **Case i.** If  $A > 0$  and  $A + B > 0$ , then the only stable equilibrium is a monomorphic equilibrium in which the entire population consists of cooperators. This situation is illustrated in Figure 2a.
- **Case ii.** If  $A > 0$  and  $A + B < 0$ , then there are two stable equilibria, one in which the entire population consists of cooperators and another in which the entire population consists of defectors. There is also an unstable polymorphic equilibrium in which the proportion of cooperators is  $-A/B$ . This situation is illustrated in Figure 2b.
- **Case iii.** If  $A < 0$  and  $A + B > 0$ , then there are two unstable monomorphic equilibria, one in which the entire population consists of cooperators and another in which the entire population consists of defectors. The only stable equilibrium is a polymorphic equilibrium in which the proportion of cooperators is  $-A/B$ . This situation is illustrated in Figure 2c.
- **Case iv.** If  $A < 0$  and  $A + B < 0$ , then the only equilibrium is a monomorphic equilibrium in which the entire population consists of defectors. This situation is illustrated

---

<sup>11</sup> This linearity does not hold in models with diploid siblings. With diploid inheritance, the equation for the difference in expected growth rates is typically quadratic or cubic. See Bergstrom and Bergstrom, 1992.

in Figure 2d.

It is interesting to look at some special cases. Suppose that there is perfectly assortative mating,  $m = 1$ , and that children always imitate their parents,  $v = 1$ . Then the model is formally the same as the model of asexual reproduction discussed above. In this case,  $A = A+B = 2(R-P)$ . For every prisoners' dilemma game,  $R > P$ , so the only equilibrium for prisoners' dilemma would be a population consisting of cooperators only.

Another simple special case is where  $v = 1$  and there is random mating, so that  $m = 0$ . In this case,  $A = R + S - 2P$  and  $A + B = 2R - T - P$ . In this case, the parameter values corresponding to each of the four cases are exactly the same as those characterizing the four possible cases for a diploid population, as displayed in Figure 1.

If  $v = 0$ , then  $A = 2(S - P) < 0$  and  $A + B = 2(R - T) < 0$ . In this case, for a prisoners' dilemma game, the only equilibrium is a population consisting only of defectors.

Notice that the parameters  $m$  and  $v$  influence equilibrium only through their influence on the expression,  $(1 + m)v^2$ . An increase in  $(1 + m)v^2$  will increase both  $A$  and  $B$  for given payoff parameters,  $S$ ,  $P$ ,  $R$ , and  $T$ . This means that the larger is  $(1 + m)v^2$ , the larger is the set of payoff parameters for which there is a monomorphic equilibrium with all cooperators and the smaller the set of payoff parameters for which there is an equilibrium with all defectors. That is to say, the more likely children are to imitate their parents, and the more likely their parents are to be the same as each other, the more likely cooperative behavior is to prevail.

#### **4. When does Provincialism Promote Cooperation?**

In the evolutionary examples that we looked at, there is a good chance that one will play a game with an opponent whose behavior is inherited from the same parent as one's own behavior. We can expect similar effects in spatial models, where neighbors interact in games of prisoners' dilemma and where behavior comes from imitation of relatively successful neighbors. This effect has been documented in a series of computer simulations carried out by Nowak and May (1992), carried out on a two-dimensional grid. Nowak and May show not only that cooperative behavior can be sustained, but also that a great

variety of cycles and waves can occur. Here we will show that cooperation can be sustained in interesting ways even in a one-dimensional model that is simple enough to be studied with a pad and paper.

Imagine a road which runs around a lake. Along this road live several farmers, each of whom has one neighbor on his left and one on his right. Each farmer plays a game of prisoners' dilemma with his neighbors and his total income is the sum of his payoffs from these games. The farmers' sons grow up, observing the actions of their fathers and their neighbors. When the fathers die, their sons take over the farms and decide whether to be cooperators or defectors. The sons choose their strategies after observing the actions by and the payoffs received by their fathers and their neighbors. For this discussion, let us consider prisoners' dilemma games for which  $2P < S + R$  and  $2R > T + P$ . (For example  $S = 0$ ,  $P = 1/4$ ,  $R = 3/4$ ,  $T = 1$ .) A variety of interesting patterns emerge. The nature of equilibrium will depend on the details of neighbors interact, which neighbors are observed by the sons, and how the sons choose there behavior.

Consider first the case where each farmer plays prisoners' dilemma with his two immediate neighbors and where the sons imitate the behavior of their father or one of his neighbors, depending on who receives the highest payoff. This setup leads to a relatively tranquil outcome in which there are many possible stable configurations. In fact, any arrangement of defectors and cooperators which consists of clusters of 3 or more cooperators and clusters of 2 or more defectors will be stable.<sup>12</sup> Cooperation in clusters smaller than two will disappear.<sup>13</sup> An isolated cooperator will get a higher payoff than either of his

---

<sup>12</sup> Consider a cluster of 3 or more cooperators which abuts a cluster of 2 or more defectors. The sons of farmers in the interior of cooperator cluster can see only their fathers and two cooperative neighbors, so they will cooperate. The son of a cooperator on the boundary of a cluster sees a cooperator neighbor who receives  $2R$ , his father who receives  $S + R$ , and a defector neighbor who receives a payoff of  $T + P$ . By assumption the largest of these payoffs is  $T + P$ , so he cooperates as his father did. The son of a defector in the interior of a defector cluster sees only defect and will defect. The son of a defector on the boundary of a cluster sees a cooperator neighbor who receives  $S + R$ , his defecting father who receives  $T + P$ , and a defector neighbor who receives  $2P$ . Since for a prisoners' dilemma game  $S < P < R < T$ , his father will have a higher payoff than the cooperator neighbor, so the son chooses to defect, just as his father did.

<sup>13</sup> An isolated cooperator gets  $2S$  and his neighbors get at least  $T + P > 2S$ . A pair of cooperators surrounded by defectors will each get  $R + S$ , while the adjacent defectors will each get at least  $T + P > R + S$ . So in each case the cooperators' sons will defect.

neighbors and so will be imitated by his son and by the sons of both of his neighbors.

Something more exciting happens if we change the preceding model so that the sons pay no attention to their fathers, but imitate their fathers' most prosperous neighbor. In this case, we see some remarkable cultural patterns which seems to “pick up their feet and walk down the road.” For example, suppose that somewhere along the road there is a grouping of five farmers consisting of a cooperator with a defector on his right, followed by a string of three cooperators to the right of the defector, making a pattern CDCCC. Suppose that all other farmers on the road are defectors. It is not hard to show that with this configuration, every son along the road will adopt the behavior of his father's neighbor on the left. This means that the behavior cluster CDCCC moves one farm to the right in each generation.<sup>14</sup> An observer who watched the behavior of the resident of a single farm over a long period of time would see cycles, in which a spell of defections would be interrupted by a cooperation, then a defection, then three cooperations and then a return to defection. Other similar patterns which walk down the road can be constructed from any block of three or more cooperators followed by a “tail” of an arbitrary number of alternating cooperators and defectors.

Another example of interest is the case where each farmer observes two neighbors to his left and two neighbors to his right and plays prisoners' dilemma with all four of them. Each son copies the most prosperous farmer from the set which includes his father and his father's four nearest neighbors. For this case, also, there are equilibria in which cooperation is sustained. All such equilibria have blocks of cooperators punctuated by patterns of defection of one of two kinds: 1) Stable pairs of defectors surrounded by cooperators. 2) “Blinkers”, which cycle in the following way. At one stage, there is a single defector, surrounded by cooperators. This defector does better than any of his neighbors and is imitated by the sons of all the farmers who can see him, making a cluster of five defectors in the next generation. The sons of the outer two of these five defectors then cooperate—leaving a cluster of three defectors. The sons of the outer two of these three defectors will

---

<sup>14</sup> Similar phenomena occurring in the well-known cellular automaton game of “Life” are known as “gliders”. Nowak and May also find gliders in their two-dimensional simulations.

cooperate, leaving an isolated defector. Then the cycle resumes.

It would be nice to have general theorems that would allow us to classify spatial games of this type and to predict the patterns of outcomes for broad classes of games. At this point, all we have are examples which show that cooperation can be sustained and that regular cycles of alternating cooperation and defection are possible.

## 5. Maximizers and Imitators

There is a striking formal similarity between a genetic model of behavior towards siblings and a model in which some individuals are imitators and others are rational maximizers who take into account the behavior of imitators. In the diploid genetic model discussed above, successful genes must “take account of” the fact that an individual with a gene for treating his sibling in a given way will, with probability  $1/2$ , be faced with a sibling who treats his sibling in the same way.

It has been suggested by Donald Cox and Oded Stark (1992) that even selfish people would be kind to their aged parents because much human behavior is “imprinted” during childhood. That is, children observe how their parents behave and later adopt these behaviors without knowing why. Suppose that an adult couple believe that their behavior toward their parents will be imprinted on their children, so that when they are old, their children will treat them as they treated their parents. Then, even if they were entirely selfish, they would treat their parents as they would like to be treated when they are old.

But it would be very odd to assume that the parents in the middle generation are “free to choose”, rationally according to their self-interest, while the behavior of their children is predetermined by imprinting. To make this story internally consistent, we allow the possibility that any individual may be either an imitator or a maximizer, with some probability between 0 and 1. Parents can not tell whether a young child is going to be an imitator or a maximizer. Imitator children will treat aged parents exactly as their parents treated their own parents. Maximizers will choose their behavior to maximize their self-interest, but with the awareness that their actions may be imitated by their children.

To simplify the formal treatment, let us study the case of single-parent families with

one mother and one daughter. Let us assume that maximizers seek to maximize a von Neumann-Morgenstern utility function  $U(x, y)$ , where  $x$  is the maximizer's actions toward her mother and  $y$  is the action of her daughter toward her when she is old. If the maximizer were certain that her daughter would be an imitator, she would choose the "Kantian"  $x$  that maximizes  $U(x, x)$ . But if she believes that her daughter may be a maximizer rather than an imitator, then she will not be so generous to her mother. If a parent chooses action  $x$  toward her mother, then an imitating daughter will choose action  $x$  toward her, but a maximizing daughter will choose an action  $y$  which is independent of her mother's choice of  $x$ .

Let us assume a stationary environment such that the planning problem faced by each generation is the same as that faced by its successor. Suppose that the probability that a child is an imitator is  $\pi$  and suppose that the action taken by a maximizing child toward her mother is  $y$ . Then a mother who chooses action  $x$  toward her parent will have an expected utility of

$$\pi U(x, x) + (1 - \pi)U(x, y).$$

Let  $\bar{x}(y)$  be the choice of  $x$  that maximizes the above expression. Since the environment is stationary, if the parent's daughter is a maximizer, she will face the same maximization problem as her mother. Therefore her action  $y$  toward her mother will be the same as the action  $\bar{x}(y)$  of the parent toward *her* mother. It follows that in any generation, a maximizing parent will choose  $\bar{x}$  so that the value of  $x$  that maximizes  $\pi U(x, x) + (1 - \pi)U(x, \bar{x})$  is  $\bar{x}$ .

Where the utility function  $U$  is differentiable, the first-order necessary condition for maximizers is found by calculating the derivative of  $\pi U(x, x) + (1 - \pi)U(x, \bar{x})$  with respect to  $x$ . This first-order condition is  $U_1(\bar{x}, \bar{x}) + \pi U_2(\bar{x}, \bar{x}) = 0$ , where  $U_i(x, y)$  is the partial derivative of  $U$  with respect to its  $i$ th argument. In equilibrium, according to this condition, maximizers will choose  $x$  so that the marginal cost  $-U_1(\bar{x}, \bar{x})$  of kindnesses to their parents is equal to  $\pi$  times the marginal benefits of kindness received from their children.

## 6. Conclusion

We have seen several environments in which an individual will certainly receive a higher payoff from defecting than from cooperating and where “copies” of an individual are more likely to appear, the higher the higher her payoff. Even in such unpromising soil, we see that cooperation can persist and flourish. The reason is that both genetic inheritance and cultural inheritance are blunt instruments. With genetic inheritance a gene for behavior that is inherited by one individual is likely to appear in its siblings. Similarly, in many environments, cultural norms are likely to simultaneously influence both players in the games in which they interact.

## Appendix—Mathematics of Cultural Evolution

Pairs of individuals can be of three types. A type 1 pair consists of two cooperators, a type 2 pair consists of one cooperator and one defector, and a type 3 pair consists of two defectors. If the fraction of cooperators in the population is  $x$ , and the assortative mating parameter is  $m$ , then the fractions of parent pairs of the  $i$ th type is given by the  $i$ th entry in the column vector

$$\vec{p}(x) = (x^2 + mx(1-x), 2(1-m)x(1-x), (1-x)^2 + m(1-x))'.$$

Assume that a child imitates a randomly chosen parent with probability  $v$  and a randomly chosen member of the population at large with probability  $1-v$ . The probability that a randomly chosen pair of offspring from a type  $i$  parent-pair is a type  $j$  sibling pair is given by the  $ij$ th entry of the following matrix,  $M(x)=$

$$\begin{pmatrix} (v + (1-v)x)^2 & (\frac{v}{2} + (1-v)x)^2 & (1-v)^2x^2 \\ 2(v + (1-v)x)(1-v)(1-x) & 2(\frac{v}{2} + (1-v)x)(1 - \frac{v}{2} - (1-v)x) & 2(1-v)xv(1-x) \\ (1-v)^2(1-x)^2 & (1 - \frac{v}{2} + (1-v)x)^2 & (v + (1-v)(1-x))^2 \end{pmatrix}.$$

Given that the fraction  $x$  of the  $n$ th generation are cooperators, the probability that a randomly chosen pair of siblings from the  $n + 1$ st generation are of type  $i$  is given by the  $i$ th entry of the column vector  $\vec{s}(x) = M(x)\vec{p}(x)$ . Calculation shows that  $\vec{s}(x) = (s_1(x), s_2(x), s_3(x))$ , where

$$\begin{aligned} s_1(x) &= x \left( \frac{v^2(1+m)}{2}(1-x) + x \right), \\ s_2(x) &= 2x(1-x) \left( \frac{v^2(1+m)}{2} - 1 \right), \\ s_3(x) &= (1-x) \left( \frac{v^2(1+m)}{2}x + 1 - x \right). \end{aligned}$$

Cooperators in type 1 sibling pairs will get payoffs of  $R$  and cooperators in type 2 sibling pairs will get payoffs of  $S$ . Defectors in type 2 sibling pairs will get payoffs of  $T$  and defectors in type 3 sibling pairs will get payoffs of  $P$ .

The probability that any individual survives to reproduce is assumed to be proportional to the average payoff that it receives in the games it plays with its siblings. This means that the total number of surviving cooperators in the second generation will be proportional to  $2s_1(x)R + s_2(x)S$  and the total number of offspring of competitors in the second generation will be proportional to  $s_2(x)T + 2s_3(x)P$ . Where  $x$  is the proportion of the mating population in generation  $n$ , the ratio of the number of surviving cooperators in generation  $n+1$  to the number of cooperators in generation  $n$  must be  $\rho_c(x) = \kappa(2s_1(x)R + s_2(x)S)/x$  and the corresponding ratio for defectors must be proportional to  $\rho_d = \kappa(s_2(x)T + 2s_3(x)P)/(1-x)$  for some common factor of proportionality  $\kappa > 0$ . Examining the above expressions for  $s_1(x)$ ,  $s_2(x)$ , and  $s_3(x)$ , we see that  $\rho_c(x)$  and  $\rho_d(x)$  are both linear expressions in  $x$ . In fact, the difference between the two growth rates is just  $\kappa D(x)$ , where  $D(x)$  is as defined in the text of the paper.

Figure 1

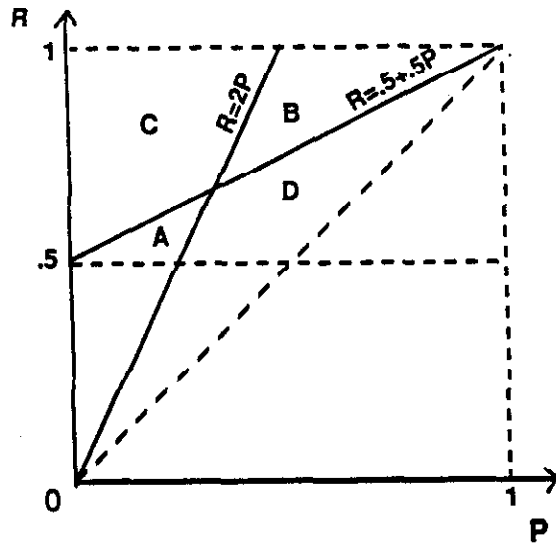
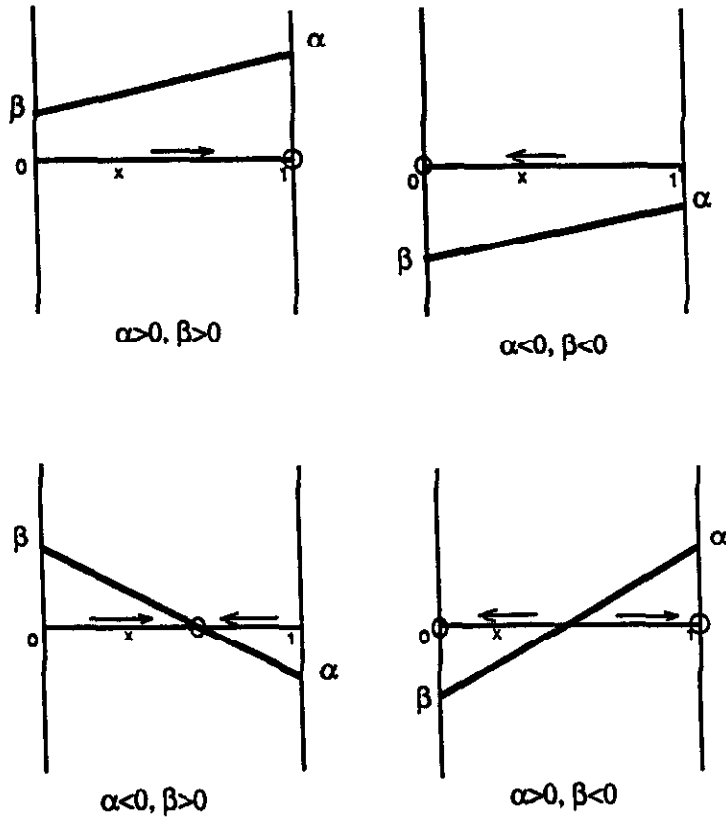


Figure 2



Stable equilibria are circled. Arrows indicate direction of movement.

## References

- Becker, Gary (1976) "Altruism, Egoism, and Fitness: Economics and Sociobiology," *Journal of Economic Literature*, **14**, .
- Becker, Gary (1981) *A Treatise on the Family*. Cambridge, Ma.: Harvard University Press.
- Bergstrom, Carl and Bergstrom, Ted (1992) "The Evolution of Altruism among Diploid Siblings who Play Games Like Prisoners' Dilemma," University of Michigan Working Paper.
- Bergstrom, Ted (1989) "Love and Spaghetti, The Opportunity Cost of Virtue," *Journal of Economic Perspectives*, , .
- Bergstrom, Ted (1992) "On the Evolution of Altruistic Ethical Rules for Siblings," University of Michigan working paper.
- Bergstrom, Ted (1988) "Systems of Benevolent Utility Interdependence," University of Michigan working paper.
- Bernheim, B. Douglas and Stark, Oded (1988) "Altruism within the Family Reconsidered: Do Nice Guys Finish Last?," *American Economic Review*, , 1034-1045.
- Boyd, Robert and Richerson, P. (1985) *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Cavalli-Sforza, Luigi and Feldman, Marcus (1981) *Cultural Transmission and Evolution*. Princeton, N. J.: Princeton University Press.
- Cox, Donald and Stark, Oded (1992) "Intergenerational Transfers and the Generation Effect," Harvard University working paper.
- Dawkins, Richard (1976) *The Selfish Gene*. New York: Oxford University Press.
- Frank, Robert (1988) *Passions within Reason*. New York: Norton.

- Hamilton, William D. (1964a) "The Genetical Evolution of Social Behavior. I," *Journal of Theoretical Biology*, **7**, 1-16.
- Hamilton, William D. (1964b) "The Genetical Evolution of Social Behavior. II.," *Journal of Theoretical Biology*, **7**, 17-52.
- Hirshleifer, Jack (1977) "Economics from a Biological Viewpoint," *Journal of Law and Economics*, , 1-52.
- Hirshleifer, Jack (1978) "Natural Economy Versus Political Economy," *Journal of Social and Biological Structures*, **1**, 319-337.
- Maynard Smith, John (1982) *Evolution and the Theory of Games*. New York: Cambridge University Press.
- Nowak, Martin and May, Robert (1992) "Evolutionary Games and Spatial Chaos," *Nature*, **359**, 826-829.
- Robson, Arthur (1992a) "Status, the Distribution of Wealth, Private and Social Attitudes Toward Risk," *Econometrica*, **60**, 837-858.
- Robson, Arthur (1992b) "The Biological Basis of Expected Utility, Knightian Uncertainty, and the Ellsberg Paradox," Working Paper, Economics Department, University of Western Ontario, London Ontario.
- Stark, Oded (1989) "Altruism and the Quality of Life," *American Economic Review*, **79**, 86-90.
- (in press)
- Stark, Oded (1992) "Nonmarket Transfers and Altruism," *European Economic Review*, , .
- Trivers, Robert (1985) *Social Evolution*. Menlo Park, Ca.: Benjamin/Cummings.
- Wright, Sewall (1921) "Systems of Mating," *Genetics*, **6**, 111-178.