Evolution of Social Behavior: Individual and Group Selection Models

Theodore C. Bergstrom
University of California Santa Barbara

October 1, 2001

> " A selector of sufficient knowledge and power might perhaps obtain from the genes at present available in the human species a race combining an average intellect equal to that of Shakespeare with the stature of Carnera. But he could not produce a race of angels. For the moral character or for the wings he would have to await or produce suitable mutations."
> ... J.B.S. Haldane [28], p. 110

1 Selfishness and Group Selection

What can our evolutionary history tell us about human motivations and social behavior? The genes that influence our own behavior are copies of copies of copies of genes that have guided our ancestors to survive and to reproduce successfully. If selfish behavior leads to reproductive success, then we should expect that our genes urge us to behave selfishly.

Considerations of this kind led Richard Dawkins [19] to assert:

> "If we were told that a man lived a long and prosperous life in the world of Chicago gangsters, we would be entitled to make some guesses as to the sort of man he was.... Like successful Chicago gangsters, our genes have survived, in some cases for millions of years, in a highly competitive world. ... If you look at the way natural selection works, it seems to follow that anything that has evolved by natural selection should be selfish." (pp 2-4)

Michael Ghiselin [25] states this view even more emphatically:

> "Where it is in his own interest, every organism may reasonably be expected to aid his fellows ... Yet given a full chance to act in his own interest, nothing but expediency will restrain him ... Scratch an "altruist" and watch a "hypocrite" bleed.

Sir Alexander Carr-Saunders [14], a sociologist who pioneered studies of demography and social evolution, offered an opposing view. Carr-Saunders maintained that evidence from primitive cultures suggests that prehistoric man was clustered into groups who inhabited well-defined areas, and that migration between groups was infrequent. Moreover, these groups typically managed to avoid overpopulation and the attendant scourges of war, famine, and disease. According to Carr-Saunders, population densities remained roughly constant and close to the levels that would maximize per capita food consumption. Carr-Saunders found that in existing primitive societies,

fertility is deliberately restrained by means of abortion, infanticide, and longterm sexual abstinence. He argued that such reproductive restraint is inconsistent with selfish maximization of individual fertility and must somehow be explained by "group selection."

Thus Carr-Saunders asserted that:

> "Now men and groups of men are naturally selected on account of the customs they practise just as they are selected on account of their mental and physical characters. Those groups practising the most advantageous customs will have an advantage in the constant struggle between adjacent groups over those that practise less advantageous customs. Few customs would be more advantageous than those which limit the number of a group to the desirable number ... There would grow up an idea that it was the right thing to bring up a certain limited number of children and the limitation of the family would be enforced by convention." [14] (page 223)

Carr-Saunders suggested that group selection operates for humans "who have achieved sufficient social organization," but not for more primitive animals. In an encyclopedic study [66] of the spatial dispersion of animal populations, V.C. Wynne-Edwards proposed that this principle has far more ancient roots, and applies to much of the animal kingdom. Wynne-Edwards maintained that in a vast number of species and in a great variety of environments, animals do not reproduce as rapidly as they would if individuals were attempting to maximize their own fertility. He cited many examples of species in which large gatherings assemble just before breeding time. These gatherings, he claimed, allow individuals to determine the existing population density and to adjust their reproductive decisions in such a way as to maintain a relatively constant population. In Wynne-Edwards view, animal species are able to solve the "tragedy of the commons" and to maintain population densities at an "optimal level for each habitat that they occupy."

According to Wynne-Edwards:

> "Where we can still find nature undisturbed by human influence ... there is generally no indication whatever that the habitat is run down or destructively overtaxed. On the contrary the whole trend of ecological evolution seems to be in the very opposite direction, leading towards the highest state of productivity that can possibly be built up within the limitations of the

> inorganic environment. Judging by appearances, chronic over-exploitation and mass poverty intrude themselves only as a kind of adventitious disease, almost certain to be swiftly suppressed by natural selection." [66] (page 8)

Wynne-Edwards argued that group selection was made possible because "animal (and plant) species tend to be grouped into more or less isolated populations" who depend on food resources of a "localized, immobile" character. He argues that territoriality and longevity of local populations are essential for group selection

> "Above all, the local stock conserves its resources and thereby safeguards the future survival of its descendants; and no such adaptation could have evolved if the descendants did not normally fall heirs to the same ground ... it is of the greatest importance in the long-term exploitation of resources that local populations should be self-perpetuating. If confirmation were needed of this conclusion, it could be found in the almost incredible facilities of precise navigation developed in all long-distance two-way migrants whether they are birds, bats, fish, or insects, to enjoy the advantages of two worlds, and still retain their life-long membership in the same select local stock. Ideally, localisation does not entail complete reproductive isolation however; we have to consider later the pioneering element also—in most species relatively small—that looks after colonisation and disseminates genes. [66] (page 20)

Some biologists received Wynne-Edwards book as academic heresy. According to Dawkins [19], by "being wrong in an unequivocal way, Wynne-Edwards is widely credited with having provoked people into thinking more clearly about selection." Others have attempted to provide formal underpinnings for his rather loosely argued proposition. A detailed and interesting account of this controversy can be found in Sober and Wilson [56].

George C. Williams [59] launched vigorous attack on what he regarded as the careless use of group-selection arguments by biologists. Williams argued that almost all of the behavior that Wynne-Edwards claims as evidence for group-selection is consistent with individuals maximizing their own long run reproductive interests or those of close relatives. For example, individuals may find it in their long term own reproductive interest to have as many offspring as possible only when the population density is low and to have fewer offspring and to give better care to each when the population density

is high. It also may be that defending territories that are larger than the minimum territory necessary for survival and reproduction in good years is individually worthwhile when the possibility of bad years is accounted for. Williams cites many examples of animal behavior that would be very difficult to rationalize as contributing to the survival prospects of the group.

Williams concluded that

> "... group related adaptations do not, in fact, exist. A *group* in this discussion should be beunderstood to mean something other than a family and to be composed of individuals that need not be closely related." (p. 93)

A more nuanced view is stated by John Maynard Smith [44], who maintains that

> "the argument is quantitative, not qualitative. Group selection will have evolutionary consequences: the only question is how important these consequences have been."

2 Games and Social Interactions

2.1 What is the Game and Who is Playing?

To understand the conflict between the individual and group selection views, it is useful to model social interaction as a game in which the players and the payoffs are explicitly specified. In the language of game theory, the two polar positions can be stated as:

- *Individual selection theory:* To predict social outcomes, we need to examine the game in which the players are individual animals and the payoff to each animal is its expected number of progeny. The outcomes that we expect to see are the Nash equilibria for this game.

- *Group selection theory:* To predict social outcomes, we need to examine the game in which the players are geographically semi-isolated communities of individuals and the payoff is the community's expected reproductive rate. The outcomes we expect to see are Nash equilibria where the players are communities.

A third alternative game formulation is suggested by the work William G. Hamilton [29] on *kin selection theory.* As Dawkins [19] suggests, individuals can be thought of as *survival machines* programmed to make copies of

their programmers, the genes. The organisms that we observe are machines that were built by those genes that have in the past been most successful in getting themselves reproduced. Selfish organisms are not typically the best vehicle for genes to use in their own reproduction. Machines that are designed to care for their offspring and to help their close relatives (who are likely to carry the same genes as their own) will typically do better.

- *Kin selection theory:* To predict social outcomes, we need to examine the game in which the players are genes that operate according to Mendelian rules of replication and that carry specific instructions to the organisms that carry them. The payoffs to these genes are their replication rates.

We shall return to the discussion of kin selection theory later in this paper. In the next sections, we examine the competing models of individual and group selection theory and points between. Taken at face value, these theories have radically different implications for the evolutionary nature of men and beasts.

Individual selection theory suggests a world populated by resolutely selfish *homo economicus* and his zoological (and botanical) counterparts. By contrast, in a world shaped by group selection we would expect to see impeccable socialists with an instinctive "Kantian" morality toward other members of their group. Of course the localism that leads to group selection would also be likely to produce some unsavory impulses towards xenophobia an intertribal warfare.

When the game being played within communities is prisoners' dilemma, the contrasting predictions of the two theories are particularly stark and simple. Since the payoff from playing defect is always higher than that of playing cooperate, individual selection theory predicts a population of defectors. But since every member of a community of cooperators gets a higher payoff than any member of a community of defectors, group selection theory predicts a population of cooperators.

Using prisoners' dilemma as a research vehicle, biologists, game theorists, and anthropologists have found much interesting territory between the two poles of individual selection and group selection. Although neither of the polar theories would be supported by modern research, the tension between the forces of individual and group selection continues to be the focus of interesting research. The use of prisoners' dilemma to explore this tension has been very instructive and will play an important part in this survey. On the other hand, as we argue in later discussion, most of the really important (and problematic) social interactions in the world are probably not games

with unique Nash equilibria, let alone dominant strategies, but games that have many distinct Nash equilbria among which societies somehow select.

2.2 Prisoners' Dilemma Games

2.2.1 Multi-player prisoners' dilemma

A multi-player prisoners' dilemma is a game in which individuals may take actions that are, in the words of J.B.S. Haldane [28], "socially valuable but individually disadvantageous." Specifically, we consider a game that has two possible strategies for each player, *cooperate* and *defect*, where the payoff to each player depends on her own strategy and the number of other players who play cooperate. In a game with N players, where K of the *other* players cooperate, let $\Pi_C(K, N)$ and $\Pi_D(K, N)$ denote the payoffs to a cooperator and a defector, respectively.

Definition 1 (N-player prisoners' dilemma game). *A game is an N-player prisoners' dilemma game if the payoff functions satisfy the following:*

- *all players are better off if all play* cooperate *than if all play* defect*; that is,* $\Pi_C(N-1, N) > \Pi_D(0, N)$.
- *regardless of what other players do, an individual gets a higher payoff from playing defect than from playing cooperate; that is,* $\Pi_D(K, N) > \Pi_C(K, N)$*, for all K between 0 and $N-1$.*

2.2.2 The Haldane Public Goods Game

In the literature of evolutionary biology, much discussion focuses on the special class of N-person games in which each player's payoff depends linearly on the number of players in the game who cooperate. This class of games was introduced to biologists in 1932 by J.B.S. Haldane, one of the founders of modern population biology.[1]

Definition 2 (The Haldane game). *The Haldane game is an N player game in which each player can play either cooperate or defect. Where x is the fraction of all players who cooperate, the payoff to each cooperator is $bx - c$ and the payoff to each defector is bx.*

[1] Cohen and Eshel [16] credit Haldane [28] (pp 207-210 of the Appendix) as the source of this model. The notation used here is that of Cohen and Eshel rather than of Haldane.

In a Haldane game with N players, if K other players cooperate, a cooperator will get

$$\Pi_C(K,N) = b\left(\frac{K+1}{N}\right) - c = b\left(\frac{K}{N}\right) - c' \tag{1}$$

where $c' = c - \frac{b}{N}$ and a defector will get

$$\Pi_D(K,N) = b\frac{K}{N}. \tag{2}$$

If all players cooperate, each gets a payoff of $b-c$; if all defect, each gets a payoff of 0. Therefore when $b > c$, all players are better off if all cooperate than if all defect. For all K, $\Pi_D(K,N) - \Pi_C(K,N) = c - \frac{b}{N} = c'$. Thus if $c > \frac{b}{N}$, given the number of other players who cooperate, defection always yields a higher payoff than cooperation. It follows that the Haldane game is an N-player prisoners' dilemma whenever $b > c > \frac{b}{N}$.

Economists will recognize the Haldane game as formally equivalent to the linear "voluntary contribution to public goods" game, much studied in experimental economics. (See Ledyard [39] for a good survey of this work) Thus we will sometimes refer to this game as the *Haldane public goods game.* In a Haldane game with N players, a cooperator confers a benefit of $\frac{b}{N}$ on every player, including himself, so that the net cost of cooperating is $c - \frac{b}{N}$. Some writers, such as Wilson [60], analyze a variant of this game in which a cooperator confers expected benefits of $\frac{b}{N}$ on every player *other than himself* at a cost of c to himself. Results for either of these two games translate easily into corresponding results for the other, since Wilson's formulation of the game with costs c is isomorphic to a Haldane game with costs $c + \frac{b}{N}$.

2.3 Evolutionary Dynamics and Altruism

2.3.1 Prisoners' Dilemma in a Panmictic Population

Let us consider the evolutionary dynamics of a population in which all individuals are "programmed" (perhaps genetically, perhaps by cultural experience) to play one of two strategies, *cooperate* or *defect* in a symmetric multi-person prisoners' dilemma game played by the entire population. We will assume that the dynamics satisfy *payoff-monotonicity*, (Weibull [58]) which means simply that the proportion of the population that plays the strategy with the higher payoff will increase.[2] If the game is prisoners'

[2] A much-studied special case of payoff monotone dynamics is *replicator dynamics* in which the growth rate of the population share using a strategy is proportional to the

dilemma, the payoff to cooperators will necessarily be lower than to defectors, so the proportion of cooperators in the population must decline over time and eventually converge to zero.[3]

Garrett Hardin, in *The Limits of Altruism* [31] expressed the same idea as he argued the futility of utopian dreams in which tribalism and parochialism are abandoned in favor of a panmictic population.

> Competition is severe and total whenever members of the same species are brought together in One World ... Conceivably some conscientious members of the community might eat less than their share of the food, but the resources they thereby released would soon be absorbed by others with less conscience. Some animals might refrain from reproducing, but the space so freed would soon be be occupied by those who were less conscientious. ... Adapting a phrase of the economist David Ricardo, we can speak of the Iron Law of the Overwhelming Minority. It is silly to dream dreams of a heaven on earth that presume a value of zero for the size of the disruptive minority.

Two of the founders of modern population biology, J.B.S. Haldane [28] and Sewall Wright [65], proposed that altruistic behavior is more likely to evolve in a population where group interaction takes place within relatively small subpopulations, (sometimes called *demes*) between which there is occasional, but relatively infrequent migration.

John Maynard Smith [43] was among the first to explore this idea with a formal model. Maynard Smith motivates this model with a charming story of mice in a hayfield:

> "... suppose that there exists a species of mouse which lives entirely in haystacks. A single haystack is colonized by a single fertilized female, whose offspring form a colony that lives in the haystack until next year, when new haystacks are available for colonization. At this time, mice migrate, and may mate with members of other colonies before establishing a new colony. The population consists of aggressive A individuals and timid a individuals, timidity being due to a single Mendelian recessive;"

difference between the average payoff to that strategy and the average payoff in the entire population [58]. The results found in this paper do not require the special structure of replicator dynamics.

[3]The result that the proportion of cooperators will decline monotonically is obvious. The result that it must converge to zero is less obvious. A proof can be found in Weibull [58]. Weibull credits this result to John Nachbar [46].

In the haystack model, all of the inhabitants of a haystack are descendants of two mice, the fertilized foundress and her mate. These mice breed and interact over the course of the summer. The timid mice are the "altruists", pursuing a strategy that is socially valuable but individually disadvantageous. Thus, within any haystack, the timid mice reproduce less rapidly than the aggressive mice, but at the end of the season, haystacks that are made up entirely of timid mice have more inhabitants than those that include aggressive mice. At season's end, the haystacks are demolished and the inhabitants scramble out and mate with mice randomly selected from the entire population. A random sample of the fertilized females is then selected to populate the next season's haystacks.

Cohen and Eshel study a closely related model in which a group may have N founders, but where reproduction is asexual. As in Maynard Smith's model, there are altruists and selfish players. A founding group and its descendants are isolated from the rest of the population for a fixed period of time, after which existing groups are disbanded and survivors are randomly selected to form new groups for the next period. Models with this pattern of group formation, interaction, and dispersal have come to be known as *haystack models.*

If group formation is assortative, so that an altruist is more likely to have altruist neighbors than a selfish individual, then it is easy to see that altruism can be maintained in the population. For example, in the extreme case where group formation is perfectly assortative so that groups consist either entirely of altruists or entirely of selfish, altruists will always receive higher payoffs than selfish individuals and altruists would eventually constitute the entire population.

But is it possible for altruism to be sustained in a haystack model if new groups are formed at random from the population at large? When groups are formed by independent random draws, the proportions in each group will not mirror the proportions in the population at large. Random selection results in some groups that have disproportionately many altruists and some that have disproportionately many selfish individuals. Within each group, the altruists get lower payoffs and hence reproduce less rapidly than the selfish. But there is a countervailing effect. Groups that contain more altruists grow more rapidly. Can this between -group effect overwhelm the within-group effect and cause the proportion of altruists *in the overall population* to increase over time? Or does Hardin's "Iron Law" extend to populations randomly selected into groups? The next sections offer a partial answer to this question.

2.4 The Iron Rule of Selfishness

David S. Wilson [60], in a pioneering study of group selection, showed that for his particular model, "random" formation of groups must result in the elimination of altruism. In a survey article called *Natural, Kin and Group Selection* [27], Alan Grafen states that "with random grouping there is no selection for altruism." However, Maynard Smith, Eshel [20], Cohen and Eshel [16], and Matessi and Jayakar [42] seem to have contrary results. Although mating is random in Maynard Smith's haystack model, for some parameter values, there is a stable equilibrium in which the entire population consists of altruists. Eshel [20] asserts that "for any altruistic trait, there is a critical level of demographic mobility under which selection would always operate for the establishment of the altruist." In Cohen and Eshel's models [16], there is "random distribution of altruist and selfish in small founder groups" and it turns out that if groups remain together long enough before being dispersed, there exists a stable equilbrium consisting entirely of altruists (as well as another stable equilibrium consisting entirely of selfish.)

To establish the circumstances under which Grafen's claim of no-altruism-with-random-sorting is correct, we need to specify the reproductive dynamics that we have in mind, as well as what we mean by altruism, and by random mating. In this section altruism is defined as playing altruist in an N-person prisoners' dilemma game in which a player's payoff is her reproduction rate. As we will later discuss, this does not exhaust the forms of behavior that might reasonably be called altruistic.

We define a *generalized haystack model*, as a model in which there is a large population of individuals, who are programmed for one of two strategies, altruist or selfish. At the beginning of each time period, these individuals are "randomly" partitioned into groups (possibly of different sizes). Each member produces (asexually) a number of offspring equal to her payoff in a game that she plays with other members of her own group. Offspring are programmed to use the same strategy as their parent. At the end of each time period, all groups are disbanded and new groups are randomly selected from the combined population of the disbanded groups.

Let $p_A(K, N)$ be the probability, conditional on being an altruist, that a player is assigned to a group of size N in which K of the *other* group members are altruists. Let $p_S(K, N)$ be the probability, conditional on being selfish, that one is assigned to a group of size N in which K of the other members are altruists. We define group formation to be *non-assortative with respect to strategy* if when new groups are assigned from the offspring

of the previous groups, altruists and selfish individual offspring have equal chances to be selected to join one of the new groups, and the probability distribution of group size and the number of other group members who are altruists is the same whether one is programmed to use the altruist strategy or the selfish strategy.

Definition 3 (Non-assortative matching process). *A matching process is* non-assortative *between types if*

- *In each period, the number of individuals of each type who are selected to join the new groups is proportional to the total number of offspring of that type who were produced in the previous period.*
- *In each period, for all K and N, $p_A(K,N) = p_S(K,N)$.*

A simple example of a non-assortative matching process is an urn model in which there is a fixed number of locations, each with a given capacity, and where each location is populated by independent random draws from the total population.

Theorem 1 (Iron Rule of Selfishness). *In a generalized haystack model, if groups are formed by a matching process that non-assortative and if the game that determines reproduction rates is an N-player prisoners' dilemma played with members of one's own group, then the proportion of altruists in the population will approach zero as the number of periods gets large.*

Proof. In each period at the time when new groups are formed, the expected numbers of offspring produced by each selfish individual and each altruist of the previous generation are, respectively:

$$\sum_N \sum_{K=1}^{N-1} p_S(K,N)\Pi_S(K,N) \quad \text{and} \quad \sum_N \sum_{K=1}^{N-1} p_A(K,N)\Pi_A(K,N). \tag{3}$$

The difference between the growth rate of the number of altruists and the growth rate of the number of selfish individuals is proportional to the difference between these two rates. Since matching is non-assortative, $p_A(K,N) = p_S(K,N)$. Therefore the difference between the two reproduction rates in (3) is

$$\sum_N \sum_{K=1}^{N-1} p_S(K,N)\left(\Pi_S(K,N) - \Pi_A(K,N)\right) \tag{4}$$

Since the game is an N-player prisoners' dilemma game, it must be that $\Pi_S(K,N) - \Pi_A(K,N) > 0$ for all K and N, and hence the expression in

(4) must be positive. It follows that the growth rate of the population of selfish individuals exceeds that of the population of altruists at all times. Therefore, the limiting value of the proportion of altruists in the population is zero. □

2.4.1 Where *Not* To Look

It is important to understand that this "Iron Rule" does not tell us that evolutionary selection cannot sustain altruistic behavior. The usefulness of Theorem 1 is that it tells us where *not* to look for the evolutionary foundations of such behavior. If we are looking for environments in which cooperative behavior is sustained by group selection, we should expect that at least one of the following is NOT true.

- The game that determines long term reproduction rates is an N-person prisoners' dilemma.
- The matching process that forms groups is "random."

In the next sections, we will exhibit models in which altruistic behavior can be sustained in equilibrium and we will show how it is that in these models one or both of the above conditions are violated.

3 "Haystack Models" of Group Selection

3.1 Maynard Smith's Mice

Maynard Smith's haystack model [43] appears to be the first formal model of "group selection" in which altruistic behavior can overpower individual selection for selfishness, when groups are formed by a non-assortative matching process. Maynard Smith presented his model as one with sexual diploid reproduction, in which "the population consists of aggressive A individuals and timid a individuals, timidity being due to a single Mendelian recessive; $a|a$ are timid, and $A|a$ and $A|A$ are aggressive." However he made a simplifying assumption that makes his model mathematically equivalent to a model with asexual reproduction. To simplify exposition and to make this model directly comparable with the later extensions by Cohen and Eshel, I will present an asexual haystack model that is formally equivalent model to Maynard Smith's sexual diploid model.

There is a meadow with many haystacks, each of which is colonized every year by exactly two mice. These two mice and their descendants interact

and reproduce asexually for the entire season, until harvest time when the haystacks are cleared. The dislodged resident mice scramble out into the meadow, and when new haystacks are built in the next year, exactly two mice are randomly selected from the meadow to colonize each new haystack. If the number of surviving mice is more than twice the number of haystacks, the extra mice are consumed by predators.

In a haystack settled by two timid mice, all of the descendants are timid and in a haystack settled by two aggressive mice, all of the descendants are aggessive. If a haystack is settled by one aggressive mouse and one timid mouse, the aggressive descendants of the aggressive mouse will eliminate the descendants of the timid mouse from the haystack population. The number of mice at season's end is assumed to be the same in haystacks colonized by one mouse of each type as in haystacks colonized by two aggressive mice.

Although timid mice do poorly when matched with aggressive mice, haystacks inhabited entirely by timid mice produce more surviving offspring at the end of the season than haystacks inhabited by aggressive mice. Specifically, it is assumed that a haystack colonized by two timid mice will produce $1 + K$ times as many surviving mice as a haystack colonized by one or more aggressive mouse.

We can study the dynamics of the meadow's mouse population by measuring the proportions of the population of each type at the time the haystacks are cleared. Let x_t be the fraction of timid mice in the population at harvest time in year t and $\xi(t) = x_t/(1 - x_t)$ the ratio of the numbers of timid mice and aggressive mice.

With random mating, in year $t + 1$ the fraction of haystacks colonized by two timid mice is x_t^2 and the fraction colonized by at least one aggressive mouse is $1 - x_t^2$. Let the number of mice who emerge at the end of the season be r if there was at least one aggressive mice among the founders and $(1 + K)r$ if the founders were both timid. If $0 < x_t < 1$, then at harvest time in year $t + 1$, the ratio the number of timid mice to that of aggressive mice will be

$$\xi_{t+1} = \frac{x_t^2(1 + K)r}{(1 - x_t^2)r} \tag{5}$$

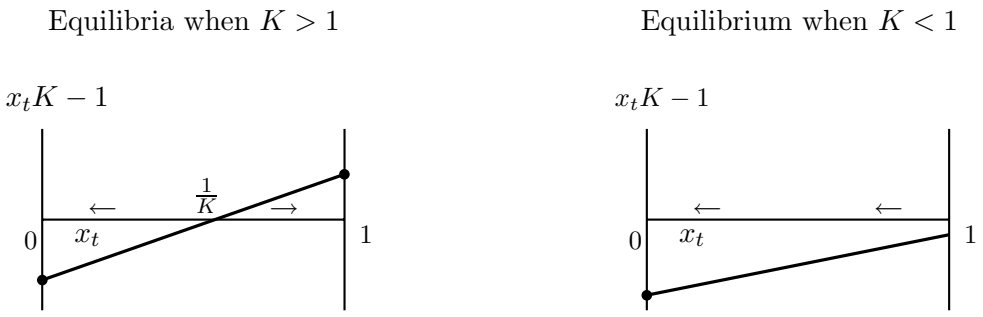
A bit of algebra applied to Equation 5 and the definition of ξ_t shows that

$$\frac{\xi_{t+1} - \xi_t}{\xi_t} = \frac{x_t K - 1}{1 + x} \tag{6}$$

Equation 6 gives us enough information to describe completely the qualitative dynamics of the mouse population. The fraction x of timid mice must

move in the same direction as the ratio ξ of timid mice to aggressive mice. It follows that when $0 < x_t < 1$, the sign of $x_{t+1} - x_t$ is the same as the sign of $x_t K - 1$. Figure 1 shows the dynamics of the haystack model. We see that when $K > 1$ there are two stable equilibria, one with timid mice only and one with aggressive mice only. There is also an unstable equilibrium where the fraction $1/K$ of mice are timid. We also see that if $K < 1$, then the population converges to a unique equilibrium where all mice are aggressive.

Figure 1: Dynamics of the Haystack Model



When $K = 1$, the number of mice produced in haystacks of timid mice is twice the number of mice in haystacks of aggressive mice. Keeping this in mind, the dynamics of the haystack model are summarized as follows.

Theorem 2 (Haystack Dynamics). *In the haystack model with random mating:*

- *If haystacks of timid mice produce more than twice as many mice as haystacks of aggressive mice, there will be two stable monomorphic equilibria, one in which $x = 0$ (all mice are aggressive) and one in which $x = 1$ (all mice are timid), as well as one unstable polymorphic equilibrium where $x = 1/K$. The basin of attraction of the equilibrium with $x = 0$ includes all initial points $x < 1/K$ and that of the equilibrium with $x = 1$ includes all initial points $x > 1/K$.*

- *If haystacks of timid mice produce fewer than twice as many mice as haystacks of aggressive mice, the only stable equilibrium is a monomorphic population of aggressive mice. Starting at any initial $x > 0$, the proportion of timid mice will decrease.*

3.2 Cohen and Eshel's Generalized Haystack Models

Dan Cohen and Ilan Eshel [16] generalize the haystack model in a very instructive way. Cohen and Eshel have two types of asexually reproducing individuals, "altruists" and "selfish." As in the haystack model, individuals group into distinct colonies where they live and breed. After some fixed length of time, all colonies are disbanded and new colonies are formed by individuals randomly selected from the population at large. In the Cohen-Eshel model, the number of individuals in the founding population is N. They assume that reproduction takes place continuously over time and that within any haystack, the reproduction rate of both types of individuals is an increasing function of the proportion who are altruists. However, the reproduction rate of altruists is lower than that of the selfish.

Cohen and Eshel focus on determining the stability of each the two possible monomorphic populations, all altruists and all selfish. This investigation is simplified by the following observation. With random group formation, when invaders are rare, almost all invaders will be selected into groups in which all other members are normal. Thus a monomorphic population of either type can be invaded by an initially small influx of the other type only if the reproduction rate of a single invader who joins $N-1$ normal individuals in founding a colony is larger than that of a normal individual among a group made up entirely of the normal type.

3.2.1 The Haldane Game

One of the models that Cohen and Eshel analyze is the Haldane public goods game. In the Cohen-Eshel formulation, if $x(t)$ is the fraction of a group that are altruists at time t, then the reproduction rate of selfish group members is $a+bx(t)$, while that of altruists in the same group is $a+bx(t)-c$. Cohen and Eshel find the ranges of parameter values in the Haldane game for which each kind of monomorphic equilibrium is stable.[4] The length of time T for which communities remain intact before dispersing is of critical importance.

Theorem 3 (Cohen-Eshel). *In the Cohen-Eshel haystack model, where reproduction rates are determined by the Haldane game and where T is the length of time for which groups remain intact:*

- *For small T, if $b/N < c$, the only stable equilibrium is a population*

[4] They are able to find closed-form solutions for the reproduction rates of a mutant cooperator in a population of defectors and of a mutant defector in a population of cooperators.

of selfish individuals and if $b/N > c$, the only stable equilibrium is a monomorphic population of cooperators.

- *If T is sufficiently large, and $b > c > 0$, there exist two distinct stable monomorphic equilibria; one with selfish players only and one with altruists only.*

The most surprising result is that if T is large enough, there exists a stable equilibrium with a population made up of altruists, even though groups are formed by an independent random matching process, and even though the Haldane game that determines instantaneous reproductive rates is an N-player prisoners' dilemma game. To see why this happens, recall that a population of altruists will be stable if the expected number of descendants of a single selfish individual who joins $N-1$ altruists in founding a community is higher than the expected number of descendants of an altruist who is among a founding group consisting entirely of altruists. The number of altruists in a group consisting entirely of altruists grows at the rate $a+b-c > a$. The descendants of the selfish invader will reproduce more rapidly than the *altruist members of the group which she joins.* But to invade the population, her descendants must reproduce more rapidly than *altruists who live exclusively among altruists.* As T is large, the descendants of a selfish invader will eventually comprise almost the entire group to which they belong. Hence the growth rate of the invader population will approach a. Thus when T is large enough, the growth rate of normal altruists is higher than that of the invading selfish. Moreover, this difference in growth rates does not diminish over time. It follows that there exists some survival period T such that if groups persist for longer than T, a monomorphic population of altruists is a stable equilibrium.

3.2.2 Nonlinearity and Polymorphic Equilibria

The Haldane model assumes that a community's growth rate depends linearly on its proportion of altruists. This implies constant returns to altruism in the sense that an additional altruist makes the same contribution to growth regardless of the number of other altruists. Cohen and Eshel show that without this linearity, monomorphic equilibrium do not always exist. They define a "generalized Haldane model" in which the reproduction rate of selfish individuals is $a+b\psi(x)$ when x is the proportion of altruists in their community; where $\psi(\cdot)$ is an increasing function such that $\psi(0)=0$ and $\psi(1)=1$. They show that if there is diminishing returns to the addition of altruists to the community, it can happen that the only equilibria

are polymorphic, with both types being present in equilibrium.

3.2.3 Congested Resources

Cohen and Eshel [16] also study a version of the haystack model, in which growth within each community is constrained by the amount of resources available. There are "selfish" individuals who reproduce more rapidly than "altruists," but consume more resources. At the end of a fixed period of time, T, the original communities are dispersed and new communities are founded by groups who are randomly selected from the entire population. In this model, a community whose founders are mainly altruists will produce more offspring because each uses less resources. On the other hand, the selfish members of a community produce more offspring than an altruistic member. For fixed growth rates and resource exhaustion parameters, if founding populations are small enough, there will be a stable equilibrium with altruists only, if populations are large enough, there will be a stable equilibrium with selfish only, and for intermediate sizes of population, there will be two distinct stable equilibria; one with altruists only and one with selfish only.

3.3 Migration and Stochastic Extinction

Haystack models are artificial in that they assume that groups persist in perfect isolation until they are simultaneously disbanded. More realistic models would allow some migration between groups and would have asynchronous extinctions and resettlement. Such models have been studied, with results that are qualitatively similar to those of the haystack models. Ilan Eshel [20], R. Levins [41], Bruce Levin and William Kilmer [40] and Scott Boorman and Paul Levitt [9] consider stochastic dynamic models of group selection, in which selfish individuals reproduce more rapidly than altruists within their own group, but where groups face a probability of extinction that increases with the proportion of their members who are selfish. Locations in which extinction has occurred are reoccupied by the descendants of a random selection from the population at large. In the Levins and Boorman-Levitt models, monomorphic populations of altruists are not stable, but polymorphism is favored if the difference in extinction rates between altruistic and selfish groups is large enough relative to the selective pressure within groups. Eshel adds random migration between groups to his model and finds that if the migration rate is sufficiently small, then with probability one, the population will fix at a monomorphic population of altruists, and for larger

migration rates the population will fix at a monomorphic selfish population. Levin and Kilmer [40] conducted Monte Carlo simulations of a model similar to that proposed by Eshel[5] and found that altruism emerged when founding populations wer no larger than 25 individuals and migration rates no larger than 5% per generation.

3.4 Haystacks and the Iron Rule

In the the haystack model and its generalizations, communities are formed by independent random draws and the game played by individuals within each community is an N-person prisoners' dilemma. Nevertheless, it is possible in these models for a population of altruists to survive evolutionary selection. How do these populations escape the reach of the "Iron Rule of Selfishness?"

There are two questions to ask. Is the matching process independent of strategy or is it assortative? Is the game that determines reproduction rates an N-player prisoners' dilemma game?

If we treat Maynard Smith's haystack model as a game played between the founders of colonies, then since the two founders of a haystack population are picked randomly from the population at large, the matching process is independent of type. The reproductive payoff to each of the founders is the number of descendants that she will have at the end of the season. Using Maynard Smith's assumptions, these payoffs are as follows. If two aggressive mice colonize a haystack, they will have a total of r descendants and thus each of them will have $r/2$ descendants. If an aggressive mouse and a timid mouse colonize a haystack, the timid mouse will have no descendants and the aggressive mouse will have r descendants. If two timid mice colonize a haystack, they will have a total of $r(1+K)$ descendants and thus each will have $r(1+K)/2$ descendants. This payoff matrix for the game played by founders in the haystack game is

[5] Eshel's model has asexual reproduction. The Levin-Kilmer model, like those of Levins and of Boorman-Levitt models has sexual diploid reproduction.

Table 1: The Haystack Game

	Timid	Aggressive
Timid	$R(1+K)/2$	0
Aggressive	R	$R/2$

Maynard Smith found that when $K > 1$, his haystack model has two stable equilibria, one of which is a monomorphic population of timid mice. The haystack game is a prisoners' dilemma game only if the aggressive strategy is the best response to either action by the other player. But when $K > 1$, *timid* is the best response to *timid* and *aggressive* is the best response to *aggressive*. Thus the haystack game has two pure strategy Nash equilibria, one in which both players play aggressive and one in which both play timid. A game with this payoff structure is known to economists as a a *stag hunt game*.

Similarly, the Cohen-Eshel model can be viewed as a game played between founders in each generation, where the payoff to each founder is the expected number of descendants she will have when groups are dispersed and new groups chosen at random. As Cohen and Eshel show, a monomorphic population of altruists will be stable if the expected number of descendants of a single selfish player among altruistic cofounders is lower than the expected number of descendants of an altruist in a group consisting entirely of altruists. This can happen only if the game played by founders is not an N-player prisoners' dilemma.

In the Cohen-Eshel model, if we measure payoffs of each individual by the number of her own offspring, then the game played between individuals is a multi-person prisoners' dilemma. But in their model, current rates of reproduction are not an adequate measure of reproductive success. The reason is that one's long run reproductive success depends not only on the number of one's own offspring, but on the rate at which these offspring, in turn, will be able to reproduce. In the Cohen-Eshel model, the long-term reproductive value of having an additional offspring depends on the proportion of altruists that are expected to be in one's group for the duration of survival of this group. In a population of altruists, an individual could

increase her current reproduction by switching to the selfish strategy. But over time, her selfish descendants will slow the rate of reproduction for each other and if groups are sufficiently long-lived, the number of her descendants at the time her group disperses will be lower than it would have been had she remained an altruist.

3.5 Dominant Strategy and Relative Payoff Within the Group

Some confusion in the debate on group selection has resulted from the fact that there exist games in which, paradoxically, *cooperate* is a dominant strategy, even though *defectors* always receive higher payoffs than cooperators. For example, consider N-player Haldane game in which x is the fraction of cooperators in the population, the payoffs are bx for each defector and $bx-c$ for each cooperator. Thus defectors always get higher payoffs than cooperators. But suppose that $b > \frac{b}{N} > c > 0$. If this is the case, then given the action of other players, any player will get a higher payoff if she cooperates than if she defects. To see this, notice that if K other players cooperate, a player will get $\Pi_C(K,N)) = b\frac{K+1}{N} - c$ if she cooperates and $\Pi_D(K,N) =$ if she defects. Thus we have $\Pi_C(K,N) - \Pi_D(K,N) = \frac{b}{N} - c > c$ and so *cooperate* is a dominant strategy.

Wilson [61] noticed this interesting case and suggests that it may be of great practical significance.

Wilson suggested that the someone who cooperates when $\frac{b}{n} > c$ but not when $\frac{b}{n} < c$ be called a weak altruist and someone who cooperates when $b > c > \frac{b}{n}$ be called a strong altruist.[6] Thus, in Wilson's terms, a weak altruist will cooperate if doing so increases his *absolute* payoff even if it lowers his payoff relative to that of all other players. A strong altruist will sometimes cooperate even when it reduces both his *absolute* and his *relative*payoff. altruism is likely to occur frequently and to be an important evolutionary force. He suggests that "many, perhaps most, group-advantageous traits such as population regulation, predation defense, and role differentiation" may be explained by weak altruism. Wilson argues that individual selection models will incorrectly predict that weak altruistic behavior will be selected *against*, while properly constructed group selection models will predict selection *for* such behavior.

Alan Grafen [27]does not dispute that there would be selection for cooperative behavior in the case that Wilson calls weak altruism, but thinks

[6] As we remarked in Section 2.2.2, Wilson formulates the game slightly differently. The conditions stated here are equivalent to his when Wilson's game is recast as an equivalent Haldane game.

that this use of language is misleading.

According to Grafen

> "Another source of misunderstanding arises from the use of the work 'altruism'. As we noted earlier, altruism will not evolve in simple one-generation groups that are formed at random from the population. ... (Wilson, Cohen and Eshel and others) ... redefined altruism to refer to relative success within the group rather than absolute success. Relative success is the individual's number of offspring divided by the average number of offspring of members of his group. Absolute success is number of children (or number of children relative to the whole population). Under the 'relative' definition, 'altruism' can spread. Wilson calls the acts that are altruistic under the relative definition, but not under the 'absolute' definition, 'weakly altruistic'. An alternative I prefer is 'a self-interested refusal to be spiteful'."

Apparently Grafen and Wilson would agree that models in which situations in which an action is "absolutely" but not "relatively" beneficial are of interest primarily when individuals interact in separate groups. Whether we call this behavior "weak altruism" or "self-interested non-spitefulness", it seems worthwhile to examine the formal dynamics of a haystack model in which cooperation increases an individual's absolute payoff, but reduces his payoff relative to that of other group members. Cohen and Eshel [16] have done exactly this and they found an interesting result that seems to have eluded the more rhetorical efforts of other participants in the debate. This result is reported above as part of Theorem 3. In a generalized haystack model in which the Haldane game is played within localities, if $\frac{b}{N} > c$ (which is the case of 'weak altruism') then if the length of time T between founding and dispersal groups is short, there will be a unique stable equilibrium and it is a population of cooperators only. If, however, T is sufficiently large, then there will be two distinct stable equilibria, one populated by cooperators only and one by defectors only.[7] Thus Cohen and Eshel's result as applied to "one generation groups formed at random from the population" is in full agreement with Grafen's statement. In equilibrium, for such groups, individuals will "cooperate" if and only the direct benefits that they get for

[7] Wilson [61] claims that theoreticians, including Cohen and Eshel, "tend to lump" the cases of weak altruism and strong altruism since neither is selected for in standard population models. In the case of Cohen and Eshel, I believe that Wilson is mistaken. As we see from Theorem 3, Cohen and Eshel's have sharply divergent results for the cases of "weak" and "strong" altruism.

themselves exceed the cost. More surprisingly, they also find that if groups have long persistence, even where cooperation is a dominant strategy in the single shot game, there will also exist an equilibrium in which all defect.

3.6 "Too stringent to be realistic?"

There now seems to be broad agreement with Maynard Smith [44] states that the argument about the significance of group selection for altruism is "not quantitative, but qualitative." At least for some time, there also appeared to be agreement that conditions under which group selection could work were not plausible.

"David S. Wilson [60] said that

> "recent models ... make it plausible that (group selection) can occur–the main question is to what extent the conditions for its operation (small group size, high isolation, high extinction rates) are met in nature. The current consensus is that the proper conditions are infrequent or at least limited to special circumstances..."

In his survey of the theory of group selection and kin selection, Alan Grafen [27] asserted that

> "the final consensus on these models was that the conditions for (them) to be successful were too stringent to be realistic."

Even the beleagured V.C. Wynne-Edwards called it quits (But only temporarily. In a book written in 1986 [68], Wynne-Edwards attempts to reestablish his group-selectionist arguments on firmer theoretical ground.) Grafen's survey quotes a 1978 article by Wynne-Edwards as follows:

> "but in the last 15 years, many theoreticians have wrestled with it and in particular with the specific problem of the evolution of altruism. The general consensus of theoretical biologists at present is that credible models cannot be devised by which the slow march of group selection could overtake the much faster spread of selfish genes that bring gains in individual fitness. I therefore accept their opinion. [67]

Levin and Kilmer [40] seem to have been the first to explore the plausibility of the parameter values under which models of group selection with random matching can lead to altruism. They conducted Monte Carlo simulations of a model similar to Eshel's extinction model [20] and report that

> "Interdemic selection favoring an allele was able to overide the effects of Mendelian selection operating against it and led to maintenance of this allele in either fixed or polymorphic states. However, with potentially realistic deme survival functions and significant levels of Mendelian selection, restrictive conditions were necessary for this result to be obtained. In our simulated population, genetically effective deme sizes of less than 25 and usually closer to 10 were required, and the rate of gene exchange, through flow migration could not be much greater than 5% per generation."

Wilson [62] ran Monte Carlo simulations of a model similar to Maynard Smith's haystack model, with founding populations of two individuals, and with dispersal and rematching of the population at the end of a fixed length of time. Wilson drops Maynard Smith 's assumption that in populations with some genes for aggressive behavior, all carriers of the gene for timidity are eliminated before the haystack population is dispersed. In Wilson's simulation, in each generation, an altruist reduces its own reproduction rate by c and contributes $b > c$ to the reproduction rate of a randomly selected other member of the group. As in the Eshel-Cohen model, a group stays together for a fixed, finite number of periods before dispersing and mating at random. But while reproduction is asexual in the Eshel-Cohen model, Wilson has sexual diploid reproduction. Wilson points out that if communities disperse after a single period, then the model is the same as Hamilton's model of kin-selection [29] , and Hamilton's rule applies; there will be a unique stable equilibrium, which will be populated by altruists if $b > 2c$ and by selfish individuals if $b < 2c$. When the number of generations is 5, his simulation results that indicate that with $b/c = 2.2$, there are two distinct equilibria, a stable polymorphic equilibrium with a population of 80% altruists and a stable monomorphic equilibrium where the entire population is selfish.

Theoretical considerations may help us to recognize circumstances under which Maynard Smith's haystack model and its generalizations would plausibly support a population of altruists. In the Maynard Smith model where each haystack population gets genetic material from just individuals, we find that a monomorphic population of altruists will be a stable equilibrium if at season's end, the number of mice produced in haystacks of altruists is more than twice the number of mice produced in haystacks of selfish mice. In the Cohen-Eshel extension, with N co-founders, in order for a monomorphic population of altruists to be a stable equilibrium, it must

be that a single selfish individual in a community of altruists will have fewer descendants within that community at the time of dispersal than the *per capita* number of descendants of a community consisting entirely of altruists. Thus, if at the time the group disperses, the descendants of the selfish individual constitute the fraction s of its community, then it must be that groups consisting entirely of altruists have more than sN times as many inhabitants as groups that included a selfish individual among their founders. If, much as in Maynard Smith's model, descendants of a selfish individual dominate the population of their community quickly and thoroughly, then the purely altruistic groups would have to produce more than N times as many descendants as groups that included a selfish cofounder.

In haystack models, with durable groups, we have seent that where there is a stable equilibrium of cooperators, there is typically also another equilibrium comprised entirely of defectors. We need to be concerned about whether and how the system could move into the basin of attraction of an equilibrium of cooperators. One possibility is that payoffs to particular actions are likely to shift across time and space. As Wilson [61] suggested, actions that are "strongly altruistic" in the current environment may have emerged as equilibrium actions in an environment where costs were smaller or benefits were greater so that these actions were once individually rational in single shot games. These equilibria might survive changes in costs or benefits such that self-interested play in single shot games no longer supports cooperation.

It is useful to notice at this point that the conditions of group stability that favor cooperation in haystack models are also favorable to the evolution of reciprocal agreements that support cooperation.

4 Assortative Matching

In prisoners' dilemma games, everyone gets a higher payoff from playing with a cooperator than with a defector, but in any encounter, playing *defect* yields a higher payoff than playing *cooperate*. In a population where both types are equally likely to play with cooperators, defectors will enjoy higher expected payoffs. But if matching is assortative, so that cooperators have better chances of meeting cooperators than do defectors, the cost of cooperation may be repaid by a higher probability of playing a cooperative opponent.

4.1 Measures of Assortativity

Suppose that a population is made up of two types of individuals and each of these individuals is matched with a partner. Let $x = (x_1, x_2)$ where x_1 is fraction of the population that is of type 1 and x_2 the fraction that is of type 2. Let $p_{ij}(x)$ be the conditional probability that an individual is matched with a type j, given that she, herself, is of type i. Since an individual is matched either to its own type or to the other type, it must be that $p_{11}(x) + p_{12}(x) = 1$ and $p_{21}(x) + p_{22}(x) = 1$. These equations imply that $p_{22}(x) - p_{12}(x) = p_{11}(x) - p_{21}(x)$. This equality allows us to define a useful measure of assortativity.

Definition 4 ((Pairwise) Index of Assortativity). *Let there be two types of individuals i and j and let $x = (x_1, x_2)$ where x_i is the proportion of type i's in the population. If individuals are matched in pairs, the index of assortativity $a(x)$ is the difference between the probability that an individual of type i is matched with its own type and the probability that an individual of type j is matched with a type i. That is, $a(x) = p_{11}(x) - p_{21}(x) = p_{22}(x) - p_{12}(x)$.*

Sewall Wright [63] defined assortativeness of mating with respect to a given trait as "the coefficient of correlation m between the two mates with respect to their possession of the trait." Cavalli-Sforza and Feldman [15] interpret Wright's correlation as follows. "The population is conceived of as containing a fraction $(1 - m)$ that mates at random and a complementary fraction m which mates assortatively." With this interpretation, if the population frequency of a type is x, then the probability that an individual of that type mates an individual of its own type is $p(x) = m + x(1 - m)$. Wright's definition and that of Cavalli-Sforza and Feldman are seen to be equivalent where we take Wright to mean that m is the coefficient of correlation between indicator random variables for possession of the trait by mates.[8] It turns out that the definition of assortativeness proposed by Wright and by Cavalli-Sforza and Feldman is equivalent to the special case of our definition where $a(x)$ is constant.

[8]Let I_i be an indicator variable that takes on value 1 if mate i has the trait and 0 otherwise. Wright's definition of the correlation coefficient between mates 1 and 2, is the correlation between the random variables I_1 and I_2. Thus we have $m = (E(I_1 I_2) - E(I_1)E(I_2))/(\sigma_1 \sigma_2)$ where σ_i is the standard deviation of I_i. Now $E(I_1 I_2) = xp(x)$, and for $i = 1, 2$, $E(I_i) = x$ and $\sigma_i = \sqrt{x(1-x)}$. Therefore $m = (xp(x) - x^2)/x(1-x)$. Rearranging terms, we find that this expression is equivalent to $p(x) = m + x(1 - m)$.

Remark 1. *Where there are two types of individuals and $a(x)$ is the index of assortativity,*

- $p_{ii}(x) = a(x) + (1 - a(x))\, x_i$ *for each* i.
- $p_{ji} = a(x)(1 - x_i)$.

Proof. The fraction of all partnerships in which a type i is matched to a j is $x_i p_{ij}(x)$ and the fraction of all partnerships in which a type j is matched to a type i is $x_j p_{ji}(x)$. Since these are just two different ways of counting the same encounters it must be that $x_i p_{ij}(x) = x_j p_{ji}(x)$. From the definition of $a(x)$, we have $p_{ii}(x) = a(x) - p_{ji}(x)$. We also have $p_{ij}(x) = 1 - p_{ii}(x)$ and $x_1 + x_2 = 1$. Simple manipulations of these equations yields the claimed results. □

The simplest, and perhaps most useful, way to generalize the index of assortativity from groups of two members to groups of arbitrary size is to simply restate the pairwise definition in terms of expected proportions. Thus for an individual of type i, let p_{ij} to be the expected proportion of other group members who are of type j. Where group size is two, this expected proportion is simply the conditional probability that one is matched with a type j, given that one's own type is i. As in the case of pairwise matching, it follows that $p_{11}(x) - p_{21}(x) = p_{22}(x) - p_{12}(x)$, and once again we can define this difference as a measure of assortativity.

Definition 5 ('Generalized' Index of Assortativity). *Where there are two types of individuals and and groups are of size N, for an individual of type i, let $p_{ij}(x)$ be the expected proportion of the $N - 1$ other group members who are of type j. The index of assortativity is defined as $a(x) = p_{11}(x) - p_{21}(x) = p_{22}(x) - p_{12}(x)$.*

4.2 Assortative Matching and Kin Selection

4.3 Hamilton's Kin Selection Theory

Families are among the most conspicuous examples of non-randomly formed groups. William G. Hamilton [29] developed a theory that predicts the strength of benevolent interactions between relatives based on their degree of relatedness. Maynard Smith [43] conferred the name *kin selection theory* on this theory, while Dawkins [19] gave it the evocative name *theory of the selfish gene*.

Biologists define the coefficient of relatedness between two individuals to be the probability that the alleles found in a randomly selected genetic locus

in the two individuals are inherited from the same ancestor. In a population without inbreeding, the coefficient of relatedness is one half for full siblings, one fourth for half siblings, and one eighth for first cousins. According to Hamilton's theory, evolutionary dynamics selects for individuals who are willing to help a genetic relative if (and only if) if the increase in reproductive value, b gained by the beneficiary, multiplied by the *coefficient of relatedness* r between the two relatives exceeds the cost in reproductive value c to the helper. The following "maxim" has come to be known as *Hamilton's rule*.

Definition 6 (Hamilton's Rule). *Help someone whose coefficient of relatedness to you is r if and only if $br > c$.*

Hamilton's work on kin selection came almost 10 years before G.R. Price and John Maynard Smith [45] introduced formal game theory to biologists. Therefore he did not think of the interaction between relatives as a game, but it is instructive to model Hamilton's interactions as a two-person game. In Hamilton's model, each player can choose whether "cooperate" by helping the other or to "defect" by not helping. A player who helps the other player reduces her own reproductive success by an amount $c > 0$, but increases that of the other player by $b > c$. The payoff matrix for this game is as follows.

Table 2: Hamilton's Help Game

		Player 2	
		C	D
Player 1	C	$b - c$	$-c$
	D	b	0

When $b > c > 0$, we see that Hamilton's help game satisfies the conditions of Definition 1 for a two-person prisoner's dilemma. Since $b - c > 0$, both players are better off when both cooperate than when both defect. Given the other player's action, a player is always better off defecting than cooperating, since $b > b - c$ and $0 > -c$. As we will demonstrate in later discussion, Hamilton's help games are *special cases* of a prisoners' dilemma. There is a large class of prisoners' dilemma games that for the purposes of evolutionary dynamics are not equivalent to games of this type.

A two person Haldane public goods game differs from a Hamilton's help game in the following way. In a Haldane game a cooperator incurs a cost to produce "public benefits that help the other player *and also benefit herself*, while in the Hamilton game the other player is the only beneficiary of a helper's efforts. But simple linear transformations of payoffs allow one to express every Hamilton game as a Haldane game and *vice versa*. For example, consider a Hamilton game with benefit b conferred on the other player at cost c to the helper. This game has the same payoff matrix as a Haldane public goods game in which *each* the payoff to a defector is $b'x$ and the payoff to a cooperator is $b'x - c'$, where x is the fraction of contributors and where $b' = 2b$, $c' = b + c$. Similarly a two-player Haldane game, with parameters b' and c' will will be equivalent to a Hamilton helping game with parameters $b = b'/2$ and $c = c' - (b'/2)$. This game will be a prisoners' dilemma if and only if $b > c > 0$, which is equivalent to $b' > c' > b'/2$.

4.3.1 Index of Assortativity for Relatives

In later work, Hamilton [30] recognized that his theory of kin selection could usefully be understood as a special case of assortative matching of partners in social interactions. It is helpful to see just how this is done by calculating the index of assortativity between prisoners' dilemma playing siblings who inherit their type by copying one of their parents.

We follow Hamilton [29] in considering a simplified version of genetics, known to biologists as *sexual haploidy*. Most animals, including humans, are sexual diploids. A sexual diploid carries two alleles in each genetic locus, one of which is inherited from its mother and one from its father. These two alleles jointly determine those individual characteristics governed by this locus. A sexual haploid has only one allele at each locus. This allele is a copy of the allele in the corresponding locus of one of its parents, chosen at random. Sexual haploidy occurs as a genetic process among some living organisms, but is of special interest in the theory of cultural transmission since it is formally identical to a theory in which for a specified behavior, a child randomly selects one of its parents to copy.[9]

Suppose that individuals can adopt one of two possible strategies, cooperate or defect, in games played with their siblings. Each child is able to observe the type of its father and of its mother and copies one or the other with probability 1/2; independently of the choice made by its siblings. Sup-

[9]Similar techniques can be applied and similar results obtained in the study of monomorphic equilibria in kin selection models with diploid sexual reproduction. For details, see Bergstrom [3] or Boorman and Levitt [9].

pose further that parents mate monogamously and independently of their strategy in games with siblings.

Let x_c be the proportion of cooperators in the entire population. If a child is a cooperator, then with probability 1/2 its sibling will have copied the same parent. In this case, the sibling must be a cooperator. With probability 1/2, the sibling role will have copied the other parent. Since parents are assumed to mate independently of their strategies, the probability that the other parent is a cooperator is x_c. Therefore the probability that a randomly chosen sibling of a cooperator is also a cooperator is

$$p_{cc}(x) = \frac{1}{2} + \frac{1}{2}x_c. \tag{7}$$

If a child is a defector, then its sibling will be a cooperator only if the sibling's role model is different from the defector's. With probability 1/2, the two siblings will have different role models, and given that they have different role models, the probability that the other parent is a cooperator is x_c. Therefore the probability that a randomly chosen sibling of a defector is a cooperator is

$$p_{dc}(x) = \frac{1}{2}x_c. \tag{8}$$

Notice that in a family of N siblings, $p_{cc}(x)$ and $p_{dc}(x)$ are equal to the expected proportion of an individual's siblings who are cooperators, conditional on that individual being a cooperator or a defector, respectively. Therefore the index of assortativity between full siblings is

$$a(x) = p_{cc}(x) - p_{dc}(x) = \frac{1}{2} \tag{9}$$

Thus we see find that with non-assortative monogamous mating, the index of assortativity between siblings is constant and equal to their coefficient of relatedness, $r = 1/2$.

Similar calculations show that the index of assortativity between other related individuals is equal to their degree of relatedness. For example, the index of assortativity between half-siblings is 1/4 and the index of assortativity between first cousins is 1/8. Bergstrom [4] calculates the index of assortativity for siblings under a variety of more general assumptions. For example, if parents mate assortatively, with an index of assortativity of mating m, then the index of assortativity between full siblings is $(1 + m)/2$. If with some probability v a child copies neither of its parents, but a randomly chosen stranger, the index of assortativity is $v(1 + m)/2$. That paper also calculates indexes of assortativity for children of polygamous marriages, and for cases where children preferentially copy the mother or the father.

4.4 Evolutionary Dynamics with Assortative Mating

4.4.1 The N-player Haldane Game

We can now investigate the evolutionary dynamics of populations of prisoners' dilemma players under assortative mating. The effect of assortative mating on expected payoffs is particularly easy to calculate when payoffs depend linearly on the proportion of cooperators in the group as in an N-player Haldane game. Let $x = (x_c, x_d)$ where x_c is the proportion of cooperators and x_d the proportion of defectors in the entire population. Define $p_{cc}(x)$ to be the expected proportion of cooperators that a cooperator finds among other members of her group and and $p_{dc}(x)$, the expected proportion of cooperators that a defector finds in her group. Recalling Equations 1 and 2, the expected payoff of a cooperator is $p_{cc}(x)b - c'$ and the expected payoff of a defector is $p_{dc}(x)b$. Therefore the difference beween the expected payoff of cooperators and that of defectors is just

$$p_{cc}(x)b - c' - p_{dc}(x)b = a(x)b - c' \tag{10}$$

where $a(x)$ is the index of assortativity.

Equation 10 generalizes Hamilton's rule from linear pairwise interactions to an N player Haldane game with voluntary provision of public goods. In this generalization, the index of assortativity plays the same formal role that the coefficient of relatedness plays in kin selection theory. In the case of kin selection theory $a(x) = r$ is constant with respect to x.

If $a(x) = a$ is constant, then except for the knife-edge case where $ab = c$, there will be a unique stable equilibrium. If $a > b/c$, then so long as both types are present, the proportion of cooperators will grow relative to that of defectors. If $a < b/c$, the reverse is true. Thus the unique stable equilibrium is a population made up entirely of cooperators if $a > c/b$ and a population made up entirely of defectors if $a < b/c$.

If $a(x)$ is variable, then it is possible that there may be more than one equilibrium, or there may be a polymorphic equilibrium with some individuals of each type. In Section 4.6 we analyze an interesting example in which $a(x)$ is variable and where there is a stable polymorphic equilibrium.

4.5 Nonlinear Payoff Functions

Alan Grafen [26] and Gordon Hines and Maynard Smith [36] show that Hamilton's rule is not correct in general for the wider class of games in which the costs of helping and the benefits of being helped may depend on the actions taken by both players. Bergstrom [3] classifies two-player non-linear

games according to whether there is complementarity or substitutability between actions and shows the way that equilibrium is altered from the Hamilton's rule predictions in each of these cases.

We follow Rappaport and Chammah [50], in denoting the payoffs in a general prisoners' dilemma game by R (reward) for mutual cooperation, P (punishment) for mutual defection, T (temptation) to a defector whose opponent cooperates, and S (sucker's payoff) to a cooperator whose opponent defects.

Table 3: Payoff Matrix

		Player 2 C	Player 2 D
Player 1	C	R	S
	D	T	P

This game is a prisoners' dilemma whenever $T > R > P > S$.[10] In the case of Hamilton's help game, described by Table 2 in Section 4.3, we have $T = b$, $R = b - c$, $P = 0$, $S = -c$. It follows that for Hamilton's game, $R+P = T+S = b-c$. Not every prisoners' dilemma game has this property. There are prisoners' dilemma games in which $R + P > T + S$ and some in which $R+P < T+S$. The evolutionary dynamics of each of these prisoners' dilemma games are qualitatively different from those of Hamilton's helper game.

Let $x = (x_c, x_d)$ where x_c is the proportion of cooperators and $x_d = 1 - x_c$ the proportion of defectors in the population. Define $p_{cc}(x)$ to be the probability that a cooperator is matched with a cooperator and $p_{dc}(x)$, the probability that a defector is matched with a cooperator. Then the expected payoff to a cooperator is:

$$\begin{aligned} p_{cc}(x)R + (1 - p_{cc}(x))\, S &= S + p_{cc}(x_c)(R - S) \qquad (11) \\ &= S + a(x)(R - S) + x\,(1 - a(x))\,(R - S) \end{aligned}$$

[10] Some writers use a definition that adds the addtional restriction that $2R > T + P$ which ensures that mutual cooperation yields a higher *total* payoff than the outcome where one player cooperates and the other defects.

where the latter equation follows from Remark 1.

The expected payoff to a defector is:

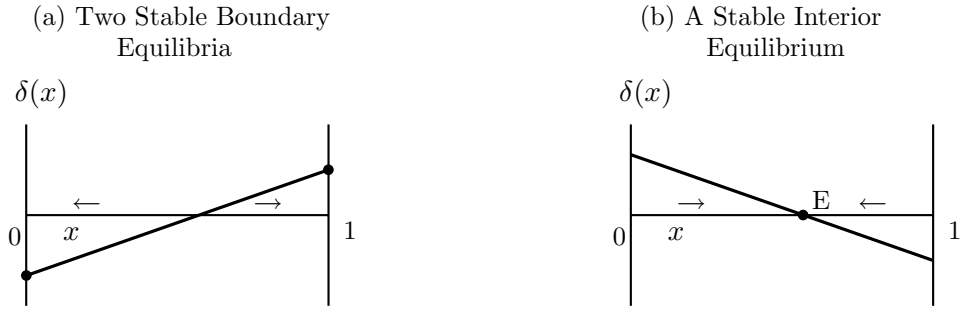
$$\begin{aligned} p_{dc}(x)T + (1 - p_{dc})P &= P + p_{dc}(T - P) \\ &= P + x_c\,(1 - a(x))\,(T - P) \end{aligned} \tag{12}$$

where again the latter equation follows from Remark 1. If we subtract the expression in Equation 12 from that in Equation 11, we can express the difference between the expected payoff to a cooperator and that to a defector as a function of x_c:

$$\delta(x_c) = S - P + a(x)(R - S) + x_c\,(1 - a(x))\,(R + P - (S + T)) \tag{13}$$

Equation 13 can be used to characterize the equilibria, under the assumption of monotone dynamics (see Section 2.3.1), of any symmetric two-player, two-strategy games with assortative matching.[11]

Figure 2: Dynamics of Prisoners' Dilemma



Where $a(x) = a$ is constant, we see from Equation 13 that the difference between the payoffs to the two strategies is linear in the proportion x_c of cooperators in the population. In this case, we see that $\delta(0) = aR + (1 - a)S - P$ and $\delta(1) = R - (aP + (1 - a)T)$. A simple calculation shows that $\delta(1) - \delta(0) = (1 - a)(R + P - S - T)$. Thus the graph of $\delta(x)$ slopes upward if $R + P > S + T$, downward if $R + P < S + T$, and is horizontal if $R + P = S + T$. It could happen that $\delta(0)$ and $\delta(1)$ are both positive, in which case there is a unique stable equilibrium populated entirely of cooperators or both negative, in which case there is a unique stable equilibrium populated

[11] Though most of our discussion focusses on prisoners' dilemma, this formula applies as well to games without a dominant strategy, such as *chicken*, and the *stag hunt*.

entirely by defectors. But there are also two other interesting cases. In figure 2(a), where $\delta(0) < 0$ and $\delta(1) > 0$, there are two distinct "monomorphic" equilibria, one consisting of cooperators only and one consisting of defectors only. In figure 2(b) where $\delta(0) > 0$ and $\delta(1) < 0$, neither monomorphic population is stable and there is a unique stable "polymorphic" equilibrium at the point E.

4.6 Assortative Matching with Partner Choice

We can expect to see assortative matching if individuals have some evidence of each others' types and some choice about with whom they match. In a multiplayer prisoners' dilemma game, everyone would rather be matched with cooperators than with defectors. If players' types were perfectly observable and if groups are able to restrict entry, then groups of cooperators would not admit defectors, and so the two types would be strictly segregated. But suppose that detection is less than perfectly accurate.

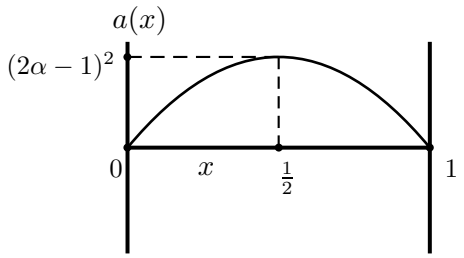
Bergstrom [4] presents a model in which players are labelled with an imperfect indicator of their type. (e.g. reputation based on partial information, behavioral cues, or a psychological test) Assume that with probability $\alpha > 1/2$, a cooperator is correctly labelled as a cooperator and with probability $1 - \alpha$ is mislabelled as a defector with probability. Assume that with probability $\beta > 1/2$, a defector is correctly labelled and with probability $1 - \beta$ is mislabelled as a cooperator.

Everyone sees the same labels, so that at the time when players choose partners there are only two distinguishable types: players who appear to be cooperators and players who appear to be defectors. Although everyone realizes that the indicators are not entirely accurate, everyone prefers to match with an apparent cooperator rather than an apparent defector. Therefore, with voluntary matching, there will be two kinds of groups, those made up entirely of apparent cooperators and those made up entirely of apparent defectors.

In this model, in contrast to the case if kin selection, the index of assortativity varies with the proportion of cooperators in the population. If we graph $a(x_c, 1 - x_c)$ as a function of x_c, the graph looks qualitatively like Figure 3.

There is a simple intuitive explanation for the fact that $a(0) = a(1) = 0$. In general, a cooperator is more likely to be matched with a cooperator than is a defector because a cooperator is more likely to be labelled a cooperator than is a defector. But if x is small, so that actual cooperators are rare, the advantage of being matched with an apparent cooperator is small because

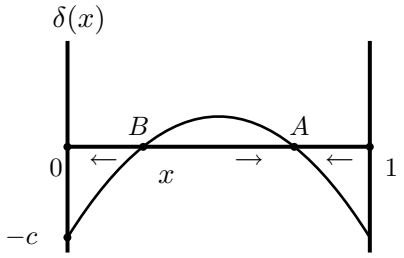
Figure 3: **Graph of $a(x)$ where $\alpha = \beta$**



almost all apparent cooperators are actually defectors who have been mislabelled. Similarly, when x is close to one, defectors are rare, so that most apparent defectors are actually cooperators who have been mislabelled. In the latter case, even if a defector is labelled a defector, his chance of getting matched with a cooperator are good. Thus in the two extreme cases, where x approaches zero and where x approaches one, the chances of being matched with a cooperator are nearly the same for a defector as for a cooperator.

Recall from Equation 10, that in the Haldane linear multiperson prisoners' dilemma game, the difference between the expected payoff of cooperators and that of defectors is simply $\delta(x) = a(x)b - c$ where x is the fraction of cooperators in the population and $a(x)$ is the index of assortativity. Figure 4 shows the graph of $\delta(x)$ for a case in which $\delta(x)$ takes some positive values. As we see from the graph, under monotone dynamics there are two locally stable equilibria. One of these equilibria occurs where $x = 0$ and the other is at the point marked A. For any level of x to the left of the point B or to the right of the point A, $\delta(x) < 0$ and so x, the proportion of cooperators in the population, would decline. For any level of x between the points A and B, $\delta(x) > 0$ and so in this region x would increase.

Figure 4: **Graph of $\delta(x)$ for additive Prisoner's Dilemma**



For Prisoners' Dilemma games with additive payoffs, $\delta(x) = a(x)b - c$. We have shown that that $a(0) = a(1) = 0$, $a'(0) > 0$, $a'(1) < 0$, and $a''(x) < 0$ for all x between 0 and 1. It follows that $\delta(0) = \delta(1) < 0$, $\delta'(0) > 0$, and $\delta'(1) < 0$, and $\delta''(x) < 0$ for all x between 0 and 1. The fact that $\delta''(x) < 0$ on the interval $[0, 1]$ implies that the graph of $\delta(x)$ is "single-peaked" as in in Figure 4. Where this is the case, and if $\delta(x) > 0$ for some x, there must be exactly one stable polymorphic equilibrium and one stable monomorphic equilibrium with defectors only.

An earlier model by Robert Frank [23] also explores the evolutionary dynamics in a population of cooperators and defectors.[12] In Frank's model, each member of each type projects a "signal of trustworthiness" that is a random draw from a continuous signal distribution. The two types draw from different distributions, whose supports overlap, but where the higher one's signal is the more likely it is that one is a cooperator. Each individual has the option of matching with a partner or of having no partner. Partners play a game of prisoners' dilemma. Those who choose to have no partner are assumed to receive the same payoff as that received by a defector matched with a defector. Players understand the game, including the payoff values an statistical distributions of payoffs and can rationally calculate their own optimal responses. Since each player prefers those who project higher signals, every individual will be matched with someone who projects approximately the same signal. In equilibrium, cooperators who project a signal lower than some critical value realize that the partners that they can attract are so likely to be defectors that it is better to stay unmatched. Frank shows that for this model there is a unique stable equilibrium and it occurs with a polymorphic population that includes both cooperators and defectors.

Skyrms and Pemantle [55] explicitly model the dynamic formation of group structure by reinforcement learning. Individuals begin to interact at random to play a game. The game payoffs determine which interactions are reinforced and a social network emerges. They report that social interaction groups that tend to form in their model consist of small interaction groups within which there is partial coordination of strategies.

4.7 Assortative Matching Induced by Spatial Structure

The reason that evolution selects for individuals who value their siblings' well-being is that two siblings have a high probability of carrying the same genetic program. Hence an individual who is programmed to be kind to

[12] Frank calls them "honest" and dishonest" types.

his brother is likely to be the beneficiary of a kind brother. Similarly, if neighbors have a significant probability of sharing the same role model, those who cooperate with neighbors may enjoy a higher likelihood of benefiting from neighborly cooperation than those who act selfishly.

Evolutionary biologists have stressed the importance of spatial structure on the spread of mutations, genetic variation and the formation of species. Wright [64] studied the degree of inbreeding in a model in which a population is distributed uniformly over a large area, but individuals are more likely to find mates who live nearby. Kimura and Weiss [38] studied genetic correlations in a one dimensional "stepping stone model" in which there is an array of colonies along a line and where "in each generation an individual can migrate at most 'one step' in either direction " and extended this model to colonies located on two and three dimensional lattices.

More recent authors have explored the dynamics of a population of agents located on a spatial grid, who repeatedly play a game with their neighbors and who may switch their strategies either deterministically or stochastically in response to their observations of the payoffs realized by themselves and their neighbors. Nowak and May [48] ran computer simulations with a deterministic model of prisoners' dilemma playing agents located on a two-dimensional grid. The grid is initially populated with some assortment of cooperators and defectors. In each round, each individual uses its preassigned strategy in a game of prisoners' dilemma with each of its immediate neighbors. After this round, each site is occupied by its original owner or by one of its neighbors, depending on who had the highest score in the previous round. Their simulations show that this process can generate chaotically changing spatial patterns in which the proportions of cooperators and defectors fluctuate about long-term averages.

Bergstrom and Oded Stark [2] model a population of farmers located on a road that loops around a lake. Each farmer plays prisoners' dilemma with his two adjacent neighbors, using one of the two strategies cooperate or defect. The farmers' sons observe the strategies and payoffs of their fathers and their immediate neighbors and imitate the most successful of these individuals. For this setup, it turns out that any arrangement of cooperators and defectors will be stable if cooperators appear in clusters of three or more and defectors in clusters of two or more. Bergstrom and Stark also show that if the sons do not pay attention to their fathers, but copy the more successful of their father's neighbors, then it is possible to find patterns that "move in a circle" around the lake. If there are at least eight farmers on the road, a pattern of the form $CDCCC$ can form an this pattern would move clockwise by one farm in each generation. Thus

a long-lived chronicler, who observed behavior at a single farm would see "cyclic behavior" in which spells of cooperation are interrupted by defection according to a regular temporal pattern.

Eshel, Larry Samuelson, and Avner Shaked [21] present a thorough analysis of the circular setup considered by Bergstrom and Stark. For the non-stochastic case, they show that in addition to an equilibrium with defectors only, there are stable equilibria in which some cooperators and some defectors survive and that in all such equilibria, at least 60 per cent of the population must be cooperators. They also show that if the initial distribution of cooperators and defectors is determined by independent random draws, then as the size of the population gets large, the probability that the initial distribution is in the basin of attraction of one of the equilibria that includes cooperators approaches unity.

Surprisingly, Eshel *et al* were able to show that when there is a positive probability of mutations, in the limit as the mutation rate becomes small, the only stationary states that have positive probability are the ones in which at least 60 percent of the population are cooperators. As the authors explain:

> One's initial impression might be that mutation 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 whither and die when introduced into a collection of Egoists... 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 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 ont the number of Egoists that can appear.

Although the structure of equilibrium sets in the Bergstrom-Stark model and in the Eshel-Samuelson-Shaked model seem too complicated and diverse for a simple measure of assortativity to be of any use, Eshel, Shaked, and Emilia Sansone [22] constructed a model of prisoners' dilemma players on a line in which, quite remarkably, the dynamics depend on the index of assortativity for a specific critical configuration of cooperators and defectors. The model starts with an initial configuration of cooperators and defectors. In each period, each individual plays a prisoners' dilemma game with each of her k nearest neighbors. A player will not change strategies from one period to the next if her two nearest neighbors use the same strategy that she uses. But one of these neighbors uses a different strategy, she will observe

the average realized payoffs of cooperators and of defectors who are within n positions of herself. She will randomly adopt a strategy for the next period, where the probability that a strategy is adopted is proportional to the average payoff of those whom she observes using that strategy. The authors show that the long run fate of this system depends entirely on what happens at a frontier between long strings of individuals of each type. From this configuration, one can calculate the probability that a defector situated at the boundary will switch to cooperation and the probability that a cooperator situated at the boundary will switch to defection. These two probabilities depend on comparisons of the average payoffs of cooperators and of defectors who are located within n positions of the boundary of between a long strong of cooperators and defectors. The dynamics is a simple random walk in which the limiting outcome is a population of cooperators or of defectors, depending on whether defectors are more likely to switch than cooperators or *vice versa*.

In the Eshel, Shaked, Sansone model the critical observers on the frontier see their own payoffs and the payoffs to their n neighbors. Each observed individual plays prisoners' dilemma with her k nearest neighbors. Since the observed defectors are located somewhere in a string of defectors and the observed cooperators are located somewhere in a string of cooperators, the cooperators enjoy the advantage of a larger proportion of encounters with cooperators than that experienced by defectors. If k, the number of opponents played in each direction is large and if n the distance over which the boundary individuals observe payoffs, this advantage will be slight since both the cooperators and defectors that are observed will be near the boundary and will play nearly equal numbers of cooperators and defectors. On the other hand, if n is large relative to k, then the average payoff of the observed cooperators will be close to the payoff in a community of cooperators only and the average payoff of the observed defectors will be close to the payoff in a community of defectors only.

The defectors would of course get higher payoffs if they played against the same number of cooperators as did the cooperators, but in this spatial setup, the defectors will be matched with more defectors than the cooperators and individuals living further from the frontier will have larger proportions of their neighbors being of their own type. The larger n is relative to k, the greater the proportion of observed neighbors who play their own type. The authors find expressions for the proportions of cooperators and of defectors encountered by those members of each type who can be observed by the frontier individual. From these calculations they produce an explicit function $r(k, n)$ that corresponds exactly to the *index of assortativity* as we have

defined it. In the special case where the prisoners' dilemma game has the linear payoffs that we have earlier described as the Haldane game, they observe that the outcome is exactly as would be predicted by Hamilton's rule where the coefficient of relatedness is $r(k, n)$. That is to say, cooperation will prevail if $r(k, n)b > c$ and defection will prevail if $r(k, n)b < c$.

5 Repeated Games and Group Selection

"Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizeable Groups" by Robert Boyd and Peter Richerson [12] is one of the rare scholarly titles that sharply nudges the mind toward a productive line of thought.[13] In an earlier paper[11] notice that where games with more than one Nash equilibrium are played within relatively distinct subpopulations, selection for group selection becomes a plausible mechanism for equilibrium selection. They suggest that this force is likely to be strong "if processes increasing the frequency of successful strategies *within* groups are strong compared to rate of migration among groups" and if "individuals drawn from a single group make up a sufficiently large fraction of newly formed groups." In [11], Boyd and Richerson succinctly explain the essence of group selection among alternative local Nash equilibria in the following words. "Viewed from the within-group perspective, behavior will seem egoistic, but the egoistically enforced equilibria with the greatest group benefit will prevail." In [12], they strengthen the case for group selection by noting that within stable groups where individuals encounter each other repeatedly, and can punish defections from a group norm, an extremely diverse range of results can be sustained as Nash equilibria.

Ken Binmore [7] observed that "If our Game of Life were the one-shot Prisoners' Dilemma, we should never have evolved as social animals." Binmore argues that the "Game of Life" is best modeled as an indefinitely repeated game in which reciprocal rewards and punishments can be practiced. As Binmore points out, this idea is not new. In the seventh century before Christ, Hesiod [35] stated the maxim "Give to him who gives, and do not give to him who does not." David Hume [37] says in language that is suggestive of modern game theory:

> "I learn to do service to another, without bearing him any real kindness, because I foresee, that he will return my service in expectation of another of the same kind, and in order to maintain

[13] Dawkins' *The Selfish Gene* is another member of this class.

> the same correspondence of good offices with me and others. And accordingly, after I have serv'd him ... he is induc'd to do his part, as foreseeing the consequences of his refusal." (page 521)

Several game theorists in the 1950's nearly simultaneously discovered *folk theorem*, which informs us that in indefinitely repeated games, almost all possible patterns of individual behavior can be sustained as Nash equilibria. For example, in the simple case of repeated prisoners' dilemma between two players, almost any intertemporal pattern of cooperation and defection on the part of each players can be maintained as a Nash equilibrium. The logic of the folk theorem is that in repeated games, almost any behavior can be induced as a Nash equilibrium by the threat of punishment for deviant actions. Individuals can be coordinated on a configuration of strategies by a stable, self-policing norm. Such a norm prescribes a course of action to each player conditional on the actions of the others and it includes instructions on how to punish any deviant player who violates his prescribed course of action. The punishments for each deviation must be sufficient to ensure that each individual takes the prescribed action.

Where the game is single-shot prisoners' dilemma, the theory of individual selection almost inevitably predicts universal defection, but in repeated games, even repeated prisoners' dilemma, individual selection theory leaves us an embarrassment of Nash equilibria and essentially no predictive power. As Boyd and Richerson [12], [13], Binmore [5], [6], [7], and Sober and Wilson [56] suggest, the stage is set for group selection to play a mighty role. Consider a population in which individuals are clustered into semi-isolated groups within which most of their social interactions occur. Within groups, these individuals play a repeated game that has many equilibria, some of which are better for all members than others. Binmore [7] suggests that we can expect groups using Pareto-superior equilibria to grow in size and number relative to the rest of the population and that eventually the groups that coordinate on Pareto-inferior equilibria will disappear. The transmission process may be speeded either by migrants who move to more successful groups and adopt local ways or by imitation. Boyd and Richerson [13] propose that in geographically structured populations, imitation of behavior in successful neighboring groups is likely to greatly speed the spread of Pareto-superior equilibria.

While the *folk theorem* goes a long way toward explaining the power of norms and punishment threats for maintaining a great variety of possible outcomes as Nash equilibria within communities, there remain some troubling details to be resolved in determining whether plausible evolutionary

processes will sustain the punishment strategies needed to support all of the outcomes that folk theorem postulates. As Henrich and Boyd [33] put it

> Many students of human behavior believe that large-scale human cooperation is maintained by threat of punishment. ... However, explaining cooperation in this way leads to a new problem: why do people punish noncooperators? ... Individuals who punish defectors provide a public good, and thus can be exploited by non-punishing cooperators if punishment is costly."

The standard game theoretic answer to this conundrum is that part equilibrium strategies include instructions to punish others if they are "supposed to punish" and fail to do so. These instructions for punishing include instructions to punish those who won't punish others when they are supposed to do so. In equilibrium, if you fail to perform your obligation to punish someone who doesn't do his task, you will be punished by someone else who in turn would be punished if he did not punish you, and so on *ad infinitum.* From an evolutionary point of view, this resolution seems unsatisfactory. Can we really expect that people or animals will keep track of their obligations to do nth order punishment for n greater than one or two? Moreover if the society is really in an equilibrium, deviations that require punishment will be rare and usually the result of a "mistake." Selection is likely to be very weak in such circumstances.

As Rajiv Sethi and R. Somanathan [53], point out in their survey paper "Understanding Reciprocity," "(The) problem of reciprocity being undermined by the gradual encroachment of unconditional cooperation is pervasive in the literature." Not only is it likely that punishment is costly in terms of direct payoffs. A strategy that involves unused punishments is by, any reasonable measure, more complex than a strategy that dictates the same actions in a world of cooperators but omits the punishment branch. Binmore and Samuelson [8] present a model in which strategies are modeled as finite-state automata and complexity is measured by the number of states. They postulate that a mutant that gets the same payoff as the incumbents but is less complex will invade a population. This assumption eliminates the possibility that 'nice' strategies, such as tit-for-tat will be stable monomorphic equilibria.

Nowak and Sigmund [49] introduce an evolutionary model in which individuals accumulate reputations. In each generation, a large number of pairs of individuals are selected randomly. One member of each pair is given a chance to play donor and the other is the potential recipient. Those who

choose to donate reduce their wealth by c, while the recipient's wealth increases by $b > c$. Each player has an *image score* that starts out at 0 at the beginning of life and is incremented by one unit every time that she makes a donation. A strategies for any individual i takes the form of a threshold k_i, such that if given a chance to donate to a recipient with image score s, i will do so if and only if $s \geq k_i$. After the pair interactions for the current generation have taken place, members of this generation are replaced by their offspring, who inherit the strategies of their parents (but not their image scores). The number of offspring that a parent has is proportional to the wealth that she accumulates during the course of her life. Nowak and Sigmund run computer simulations of this model. They find that when the model is run for about 150 generations, almost all population members adopt a strategy of donating to everyone with an image score of 0 or higher. When these strategies are played out, this means that almost everyone donates at every opportunity. When Nowak and Sigmund add a very small rate of mutation to new strategies, the results are very different. According to Nowak and Sigmund,

> "with mutation the population, long term simulations with mutation ... show endless cycles. ... defectors are invaded by discriminators, who only help players whose score exceeds some threshold. Next discriminators are undermined by unconditional cooperators. The prevalence of these indiscriminate altruists subsequently allows the return of defectors."

The Nowak-Sigmund model does not follow the course suggested by game theoretic constructions of punishment strategies. In their model, one's reputation improves whenever one makes a donation, regardless of whether the potential recipient has been generous or not. The kind of punishment strategy that a follower of the folk theorem would devise would have to go something like this. Initially, everyone is in *good standing*. After each play, a person is in good standing if and only if whenever she has had a chance to donate to a person in good standing she donated and whenever she had a chance to donate to a person in not in good standing, she did not donate.

Bowles and Gintis [10] build an evolutionary model of a population that includes some *shirkers* and some *reciprocators* who don't shirk and who, despite the fact that it is costly, will punish shirkers when they catch them shirking. Equilibrium in their model has a mixed population of workers and shirkers and shirkers. However, they evade the problem of the evolutionary stability of by not allowing the possibility of an invader who neither shirks nor punishes. "The Viability of Vengeance" by Dan Friedman and Nirvikar

Singh [24] has a nice discussion of the issue of the evolutionary stability of costly punishment. Friedman and Singh distinguish between punishment of group members and of outsiders. They suggest that within groups, ones actions are observed and remembered. A reputation for being willing to avenge actions harmful to oneself may be sufficient compensation for the costs of retribution. They propose that in dealing with outsiders, one is remembered not as an individual but as a representative of one's group. Accordingly, a willingness to avenge harm done by outsiders is a *public good* for one's own group since deters outsiders from uncooperative behavior to group members. They propose that a failure to avenge wrongs from outsiders is punished (costlessly) by one's own group, through loss of status.

The model of enforcement of norms that seems to me most persuasive is that of Henrich and Boyd [33] in their paper "Why Punish Defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas"[14] The authors propose that "the evolution of cooperation and punishment are plausibly a side effect of a tendency to adopt common behaviors during enculturation." They suggest that since it is simply not possible to analyze and "solve" the complex social games that we play, imitation plays a large role in decision-making. Moreover, since observation of the realized payoffs of others is not always possible, much of this imitation takes the form of 'copy-the-majority' rather than 'copy-the-most-successful.'

Henrich and Boyd test the workings of this idea on a multi-stage game. The first stage of this game is a "Haldane" game in which each individual can choose whether to make a contribution to the group at a cost of c to himself and with a total benefit of b divided equally among all group members. Those who don't contribute share the benefits but don't pay the cost. With a small probability, individuals who intend to contribute mistakenly do not. The game has a second stage in which each individual decides whether or not to punish those who defected in the first stage. Punishing costs ϕ to the punisher and ρ to the punished, where $\phi < \rho < c$. There is a second punishing stage in which individuals decide, with the same cost structure, whether to punish those who have not punished the malefactors of the first stage. And a finite number of additional stages is constructed recursively. At each stage the authors suppose that there is some small probability of mistakes.

At each stage of the game, there are two possible strategies, cooperate or defect. In the first stage, cooperate means to contribute. In later stages,

[14]This paper is a contender with the earlier cited Boyd-Richerson paper for an "informative title award".

cooperate means to punish those who defected in the previous stage. The population evolves according to "replicator dynamics" applied separately to the strategy used in each stage. In particular the difference between the growth rate of cooperators and the growth rate of defectors for this stage is a weighted average of two differences: the difference between the average payoffs of cooperators and defectors in that stage and the difference between the fraction of the population who are cooperators and the fraction who are defectors. The latter difference reflects the force of conformism.

If the weight placed on conformism is sufficiently large, then of course any strategy, including cooperate and don't punish can be maintained, simply because an invader's payoff advantage would be overwhelmed by the conformist advantage of the incumbent strategy. But while placing some weight on copying the majority seems is plausibly adaptive, placing such a large weight does not seem likely to be so. The authors stress that

> "... stabilization of punishment is from the gene's point of view a maladaptive side-effect of conformist transmission. If there were genetic variability in the strength of conformist transmission and cooperative dilemmas were the *only* problem humans faced, then conformist transmission might never evolve."

The key to Henrich and Boyd's result is that it takes only a very small weight on conformity to maintain an equilibrium that supports punishment strategies. To see why, let us look at a version of the Henrich-Boyd model with only one punishment stage. Suppose that the population is initially one in which everyone tries to cooperate at the first stage and also in the punishment stage. Then the only defections observed will be mistakes (or possibly actions of a few mutants). Individuals who defect in the first stage will get lower payoffs than those who cooperate in the first stage because almost everyone is cooperating in the punishment stage by punishing first-stage defectors. Individuals who defect in the punishment stage by not punishing first stage defectors *will* get higher payoffs than those who cooperate by punishing first stage defectors, but only slightly higher since there are very few defections in the first stage. Since almost everyone is observed to cooperate in the second stage, even a very small coefficient on conformism will be sufficient to overcome this small payoff difference. Henrich and Boyd show that when higher levels of punishment are accounted for, an even smaller coefficient on conformism is sufficient to maintain cooperation at all stages.

The Henrich-Boyd argument leaves some room for skepticism. If defections on the first round are rare, isn't it likely that in realistic models few individuals would observe a defection? But if that is the case, then

conformists who observe a defection might not be able to determine that first-order punishment is the social norm. Perhaps a polymorphic equilibrium that has just enough defectors to make the prevalence of punishment observable to conformists could be obtained in this setting.

There is room to question whether the visceral, seemingly irrational anger that people feel when they are cheated or otherwise violated can really be explained as a result of cultural transmission rather than as genetically hard-wired. Leda Cosmides, a psychologist and John Tooby [17] [18], an anthropologist, offer experimental evidence indicating that people are much better at solving logical problems that are framed as "cheater-detection" problems than at solving equivalent problems in other frameworks. In their view, this is evidence that individuals have evolved special modules in their brains for solving such problems.

There is interesting evidence that the desire for cultural transmission plays an important role in determining when people get angry. Richard Nisbett and Dov Cohen [47] have conducted experiments in which male college students are subjected to rude and insulting behavior in the laboratory. Using questionaires, behavioral responses, and checks of testosterone levels, they find that students who were raised in the American South become much angrier and more ready to fight than those who were raised in the North. The authors attribute this difference to the existence of a "culture of honor" in the South that is not present in the North.

Economists and anthropologists have recently conducted a remarkable series of experimental studies of how people in different cultures play the *ultimatum game*. In an ultimatum game, two players are matched and there is a fixed sum of money to be allocated. The first player, "the proposer" offers a portion of the total to the second player, "the responder." The responder can either accept or reject the offer. If the responder accepts, the division is that proposed. If the responder rejects, both players receive nothing. If this game is played by rational players who care only about their money payoff, then equilibrium in this behavior is for the proposer to offer the responder a very small share, which the responder will accept. In actual experiments with laboratory subjects in the United States, it was discovered that typically proposers offered a share of nearly one half, and this was accepted. When proposers attempted to capture a significantly larger share, responders would usually reject the proposal, thus acting as if they were willing to forego the small share that they were offered in order to "punish" a greedy proposer. In 1991, Alvin Roth and his coworkers [51] did a "cross-cultural" conducted in which they compared the results from running the experiment in the U.S., and in Israel, Japan, and in Slove-

nia. They found very similar results in all four countries. In 2000, Joe Henrich [32], published a study of an ultimatum game performed with the Machiguenga of Peru. The Machiguenga live in mobile, single-family units and small extended-family hamlets scattered throughout the tropical forests of the Amazon, where they practice hunting, fishing, gathering, and some horticulture. According to Henrich, among the Machiguenga, "cooperation above the family level is almost unknown." Henrich found that in sharp contrast to the results in the Western countries, where the modal offer was usual fifty percent, the modal share offered by the Machiguenga was only fifteen percent. Moreover, although the Machiguenga responders were offered a much smaller share than the their counterparts in the developed world, they accepted these offer about about 95 percent of the time–a higher acceptance rate than the average in the developed world. A recent study [34] reports on game experiments that have been conducted in a total of 15 "small-scale societies", including hunter-gathers, pastoralists, and farmers, and villagers. The studies found a great deal of divergence among these societies. In some of them results strongly suggested an equal-split norm and in others most proposers made offers much less generous than equal splitting and were not punished for doing so.

6 Conclusion

6.1 Further Reading

The literature on social evolution is large, diverse, and multi-disciplinary. There is a great deal of good work that I have failed to discuss. Some of the omissions are simply due to my ignorance. Some work that I admire and intended to include, didn't find its way into the survey because I had to narrow my focus to limit its length. Fortunately, the seriousness of these omissions is diminished by the fact that much of the omitted work is beautifully presented in other sources.

For a survey article that partially overlaps this material, but also examines a lot of good work not covered here, I recommend Rajiv Sethi and R. Somanathan's [53] lucid and insightful article, "Understanding Reciprocity."

There are several books that I strongly recommend to anyone interested in the subject of social evolution. These books tell their stories better than I could, so I confine my remarks to brief descriptions and hope that readers will find and enjoy them in undilluted form.

Cavalli-Sforza and Feldman's book, *Cultural Transmission and Evolution* [15] pioneered formal modeling of this subject. Their introductory

chapter is richly endowed with examples and presents a clearheaded formulation of the way that the implications of mutation, transmission, and natural selection can be extended from the study of genetically transmitted characteristics to that of culturally transmitted characteristics. There formulation of the contrasting effects of *vertical transmission*, (from parent to child) and of *oblique* and *horizontal* transmission are is a provocative ideas and one that they support with fascinating examples such as he spread of linguistic patterns, the introduction of birth control methods, tthe spread of the kuru virus in the Fore tribe of New Guinea, which is contracted by "ceremonial contact with dead relatives, " There is also a very interesting empirical study of the transmission from parents to children of such cultural behavior as religious beliefs, political affiliation, listening to classical music, reading horoscopes, and high salt usage.

Robert Trivers' book, *Social Evolution* [57] is a stimulating and attractive treatise on the evolution of social behavior of animals (including humans) and plants. It is full of interesting examples from the natural world, thought-provoking bits of theory, and delightful photographs and drawings.

Brian Skyrms' short book, *The Evolution of the Social Contract* [54], is a beautifully written and highly accessible application of the methods of evolutionary dynamics to behavior in bargaining games and the evolution of notions of fairness and "the social contract".

My own thinking about matters related to the evolutionary foundations of social behavior has been strongly influenced by Ken Binmore's two volume work, *Game Theory and the Social Contract* [6], [7]. This book combines social philosopy, political theory, evolutionary theory, anthropology, and modern game theory with great depth and subtlety.

Sober and Wilson's book *Unto Others*, is written in advocacy of a modern version of the group selectionist view. It contains an extensive and interesting history of theoretical controversies between group selectionists and individual selectionists. There are also reports on interesting empirical work with group selection as well as a useful survey of group norms in a sample of twenty-five cultures that they selected *randomly* from the anthropological literature.

H. Peyton Young's *Individual Strategy and Social Structure: An Evolutionary Theory of Social Institutions* [69] contains a remarkably accessible introduction to the mathematical theory of stochastic dynamics and to its applications in the study of the evolution of social institutions. Almost all of the work discussed in the present review uses deterministic dynamics to approximate the outcomes in a stochastic model. Heuristically, the justification for doing so is that if an equilibrium that is locally stable under

deterministic dynamics receives a small, one-time stochastic shock, then as the shock wears off, equilibrium will be restored.[15] Young observes that the difficulty with this argument is that occasionally, *albeit* extremely rarely, the system may receive a sufficiently large number of shocks to knock it out of the basin of attraction of any locally stable equilibrium that is not globally stable. Thus, Young argues, a proper treatment of the very long run must directly incorporate the stochastic process into the laws of motion. He shows that in models with multiple equilibria, "long run average behavior can be predicted much more sharply than that of the corresponding determinate dynamics."

Those looking for clear, mathematical presentations of the major technical issues in evolutionary game theory will do well to look at Jörgen Weibull's *Evolutionary Game Theory* [58] and Larry Samuelson's *Evolutionary Games and Equilibrium Selection* [52].

[15] Michel Benaim and Jörgen Weibull [1] have developed a careful formal treatment of the circumstances in which deterministic approximation of stochastic dynamic evolutionary processes is justified.

References

[1] Michel Benaim and Jörgen Weibull. Deterministic approximation of stochastic evolution in games. Technical Report 534, IUI Working Paper Series, Stockholm, September 2000.

[2] Theodore Bergstrom and Oded Stark. How altruism can prevail in an evolutionary environment. *American Economic Review*, 83(2):149–155, 1993.

[3] Theodore C. Bergstrom. On the evolution of altruistic ethical rules for siblings. *American Economic Review*, 85(1):58–81, 1995.

[4] Theodore C. Bergstrom. The algebra of assortative encounters and the evolution of cooperation. *International Game Theory Review*, to appear, 2001.

[5] Ken Binmore. *Fun and Games*. D.C. Heath, Lexington, MA, 1992.

[6] Ken Binmore. *Game Theory and the Social Contract I Playing Fair*. MIT, Cambridge, MA, 1994.

[7] Ken Binmore. *Game Theory and the Social Contract II Just Playing*. MIT Press, Cambridge, Ma, 1994.

[8] Ken Binmore and Larry Samuelson. Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory*, 57:278–305, 1992.

[9] S.A. Boorman and P.R. Levitt. *The Genetics of Altruism*. Academic Press, New York, 1980.

[10] Samuel Bowles and Herbert Gintis. The evolution of reciprocal preferences. Technical report, Santa Fe Institute, Santa Fe, N.M., 2000.

[11] Robert Boyd and Peter Richerson. Group selection among alternative evolutionarily stable strategies. *Journal of Theoretical Biology*, 145:331–342, 1990.

[12] Robert Boyd and Peter Richerson. Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology*, 113:171–195, 1992.

[13] Robert Boyd and Peter Richerson. Group beneficial norms can spread rapidly in structured populations. Technical report, UCLA anthropology department, Los Angeles, CA, April 2001.

[14] A.M. Carr-Saunders. *The population problem: a study in human evolution*. Clarendon Press, Oxford, 1922.

[15] L.L. Cavalli-Sforza and M.W. Feldman. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, Princeton, NJ, 1981.

[16] D. Cohen and E. Eshel. On the founder effect and the evolution of altruistic traits. *Theoretical Population Biology*, 10:276–302, 1976.

[17] Leda Cosmides. The logic of social exchange: has natural selection shaped how humans reason? *Cognition*, 31:187–276, 1989.

[18] Leda Cosmides and John Tooby. Evolutionary psychology and the generation of culture ii: A computational theory of exchange. *Ethology and Sociobiology*, 10:51–97, 1989.

[19] Richard Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 1976.

[20] Ilan Eshel. On the neighbor effect and the evolution of altruistic traits. *Theoretical Population Biology*, 3:258–277, 1972.

[21] Ilan Eshel, Larry Samuelson, and Avner Shaked. Altruists, egoists, and hooligans in a local interaction structure. *American Economic Review*, 88:157–179, 1998.

[22] Ilan Eshel, Emilia Sansone, and Avner Shaked. The emergence of kinship behavior in structured populations of unrelated individuals. *International Journal of Game Theory*, 28:447–463, 1999.

[23] Robert H. Frank. If *homo economicus* could choose his own utility function, would he want one with a conscience? *American Economic Review*, 77(4):593–604, September 1987.

[24] Daniel Friedman and Nirvikar Singh. The viability of vengeance. Technical report, U.C. Santa Cruz, Santa Cruz, CA, 1999.

[25] Michael Ghiselin. *The Economy of Nature and the Evolution of Sex*. University of California Press, Berkeley, CA, 1974.

[26] Alan Grafen. The hawk-dove game played between relatives. *Animal Behaviour*, 27(3):905–907, 1979.

[27] Alan Grafen. Natural selection, group selection, and kin selection. In J.R. Kreb and N.B. Davies, editors, *Behavioural Ecology*, chapter 3, pages 62–80. Blackwell, London, 2 edition, 1984.

[28] J.B.S. Haldane. *The Causes of Evolution*. Harper & Brothers, New York and London, 1932.

[29] W.D. Hamilton. The genetical evolution of social behavior, Parts i and ii. *Journal of Theoretical Biology*, 7:1–52, 1964.

[30] William D. Hamilton. Innate social aptitudes in man: an approach from evolutionary genetics. In Robin Fox, editor, *Biosocial Anthropology*. Malaby Press, London, 1975.

[31] Garrett Hardin. *The Limits of Altruism*. Indiana University Press, Bloomington, Indiana, 1977.

[32] Joseph Henrich. Does culture matter in economic behavior? ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *American Economic Review*, 90(4):9730979, September 2000.

[33] Joseph Henrich and Robert Boyd. Why people punish defectors. *Journal of Theoretical Biology*, 208:79–89, 2001.

[34] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, and McElreath Richard. In search of homoeconomicus: behavioral experiments in 15 small-scale societies. *American Economic Review*, 91(2):73–78, May 2001.

[35] Hesiod. *Hesiod: The Homeric Hymns and Homerica*. Heineman, London, 1929. (Edited by H. Evelyn Waugh).

[36] W.G.S. Hines and John Maynard Smith. Games between relatives. *Journal of Theoretical Biology*, 79:19–30, 1979.

[37] David Hume. *A treatise of human nature*. Clarendon Press, Oxford, second edition, 1978. (Edited by L.A. Selby-Bigge. Revised by P. Nidditch. First published 1739).

[38] Motoo Kimura and George H. Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49:561–576, 1964.

[39] John O. Ledyard. Public goods: A survey of experimental research. In John Kagel and Alvin Roth, editors, *The Handbook of Experimental Economics*, chapter 2, pages 111–181. Princeton University Press, Princeton, N.J., 1995.

[40] B.R Levin and W.L. Kilmer. Interdemic selection and the evolution of altruism. *Evolution*, 28(4):527–545, December 1974.

[41] R. Levins. Extinction. In M. Gerstenhaber, editor, *Some Mathematical Problems in Biology*, pages 77–107. American Mathematical Society, Providence, 1970.

[42] C. Matessi and S.D. Jayakar. Conditions for the evolution of altruism under darwinian selection. *Theoretical Population Biology*, 9, 1976.

[43] John Maynard Smith. Group selection and kin selection. *Nature*, 201:1145–1147, 1964.

[44] John Maynard Smith. Group selection. *Quarterly Review of Biology*, 51:277–283, June 1976.

[45] John Maynard Smith and G.R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

[46] John Nachbar. Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, 19:59–89, 1990.

[47] Richard E. NIsbett and Dov Cohen. *Culture of Honor: the psychology of violence in the South*. Westview Press, Boulder, Co, 1996.

[48] M.A. Nowak and R.M. May. Evolutionary games and spatial chaos. *Nature*, 359:826–829, 1993.

[49] Martin A. Nowak and Robert M. May. Evolution of indirect reciprocity by image scoring. *Nature*, 393:573–577, 1998.

[50] Anatol Rappaport and Albert M. Chammah. *Prisoner's Dilemma*. University of Michigan Press, Ann Arbor, Mi, 1965.

[51] Alvin E. Roth, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo, an experimental study. *American Economic Review*, 81(5):1068–95, December 1991.

[52] Larry Samuelson. *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, Ma, 1997.

[53] Rajiv Sethi and E. Somanathan. Understanding reciprocity. *Journal of Economic Behavior and Organization*, to appear.

[54] Brian Skyrms. *Evolution of the Social Contract*. Cambridge University Press, Cambridge, 1996.

[55] Brian Skyrms and Robin Pemantle. A dynamic model of social network formation. *Proceedings of the National Academy of Science*, 97(16):9340–9346, August 2000.

[56] Eliot Sober and David Sloan Wilson. *Unto Others*. Harvard University Press, Cambridge, MA, 1999.

[57] Robert Trivers. *Social Evolution*. Benjamin Cummings, Menlo Park, CA, 1985.

[58] Jörgen Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.

[59] George C. Williams. *Adaptation and Natural Selection A Critique of Some Current Evolutionary Thought*. Princeton University Press, Princeton, N.J., 1966.

[60] David Sloan Wilson. A theory of group selection. *Proceedings of the National Academy of Sciences*, 72(1):143–146, January 1975.

[61] David Sloan Wilson. Structured demes and trait-group variation. *American Naturalist*, 113:157–185, 1979.

[62] David Sloan Wilson. Altruism in mendelian populations derived from sibling groups. *Evolution*, 41(5):1059–1070, 1987.

[63] Sewall Wright. Systems of mating. *Genetics*, 6(2):111–178, March 1921.

[64] Sewall Wright. Isolation by distance. *Genetics*, 28:114–138, 1943.

[65] Sewall Wright. Tempo and modes in evolution: A critical review. *Ecology*, 26:415–419, 1945.

[66] V.C. Wynne Edwards. *Animal Dispersion in Relation to Social Behaviour*. Oliver and Boyd, Edinburgh and London, 1962.

[67] V.C. Wynne-Edwards. Intrinsic population control: an introduction. In F.J. Ebling and D.M. Stoddart, editors, *Population Control by Social Behaviour*, volume Population Control by Social Behaviour, pages 1–22. Institute of Biology, London, 1978.

[68] V.C. Wynne-Edwards. *Evolution through group selection.* Alden Press, Oxford, 1986.

[69] H. Peyton Young. *Individual Strategy and Social Structure.* Princeton University Press, Princeton, N.J., 1998.