UC Santa Barbara
Ted Bergstrom Papers

Title
The Algebra of Assortative Encounters and the Evolution of Cooperation

Permalink
https://escholarship.org/uc/item/03f6s9jt

Author
Bergstrom, Ted

Publication Date
2003-07-30

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

International Game Theory Review, Vol. 5, No. 3 (2003) 1–18
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

THE ALGEBRA OF ASSORTATIVE ENCOUNTERS AND THE EVOLUTION OF COOPERATION

THEODORE C. BERGSTROM
Aaron and Cherie Raznick Professor of Economics
University of California Santa Barbara, Santa Barbara, CA 93105 USA

This paper explores the way in which assortative matching can maintain cooperative behavior under evolutionary dynamics. If encounters are random, then in Prisoner's Dilemma games, defectors necessarily get higher payoffs than cooperators and thus will eventually prevail. But if matching is assortative, the cost of cooperating may be repaid by higher probabilities of playing against a cooperating opponent. This paper shows that a simple *index of assortativity* allows a unifying treatment of the evolutionary dynamics in a wide variety of models of social encounters.

Keywords:

1. Introduction

In Prisoner's Dilemma, everyone gets a higher payoff from playing *defect* than from playing *cooperate*, but everyone gets a higher payoff from playing against a cooperator than against a defector. If meetings between the two types are "random", then defectors will on average get higher payoffs than cooperators. But if matching is assortative, so that cooperators are more likely to meet cooperators than are defectors, then it may be that the cost of cooperating is repaid by a higher probability of playing against a cooperating opponent.

This paper explores the quantitative relation between non-random, assortative matching and the maintenance of cooperative behavior under evolutionary dynamics. We consider a population of individuals who are "hard-wired" to play either *cooperate* or *defect*. They meet other individuals according to some random process and play their programmed strategy in a game of Prisoner's Dilemma. The type that gets the higher expected payoff reproduces more rapidly. We define an *index of assortativity* of encounters and develop an "algebra of assortative encounters". In one set of applications, we calculate the index of assortativity for games between relatives with either cultural or genetic inheritance and we show the logical connection between the index of assortativity and Hamilton's theory of kin selection (Hamilton, 1964). We also apply the index of assortativity to determine the population dynamics when players select their partners, using partially informative cues about each others' types.

Table 1. A Prisoner's Dilemma Game.

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

2. Assortative Encounters in Prisoner's Dilemma

2.1. *The payoff matrix*

Consider a large population of players who meet other players according to a specified encounter rule. When two players meet, they play a game of Prisoner's Dilemma.[a] Each individual is programmed for one of two possible strategies, cooperate (C) or defect (D). The payoffs for the game are given in Table 1. This game is a Prisoner's Dilemma when the payoffs satisfy the inequalities, $T > R > P > S$.

2.2. *The algebra of encounters*

2.2.1. *The index of assortativity*

Let x denote the proportion of the population that are cooperators (C-strategists) and $1-x$ the proportion of defectors (D-strategists). Each member of the population randomly encounters one other member of the population. The probability that an individual encounters a cooperator depends, in general, both on that individual's own type and on the proportion of cooperators in the population. Let $p(x)$ be the conditional probability that one encounters a cooperator, *given that one is a cooperator* and let $q(x)$ be the conditional probability that one encounters a cooperator, *given that one is a defector*.

The fraction of all encounters between two individuals in which a cooperator meets a defector is $x(1-p(x))$. The fraction of all encounters in which a defector meets a cooperator is $(1-x)q(x)$. Since these are just two different ways of counting the same encounters, we have the following "parity equation"

$$x(1-p(x)) = (1-x)q(x)\,. \tag{2.1}$$

Let us define the *index of asssortativity* $a(x)$ to be the difference between the probability that a C-strategist meets a C-strategist and the probability that a D-strategist meets a C-strategist. Thus we have

$$a(x) = p(x) - q(x)\,. \tag{2.2}$$

Since $(1-q(x)) - (1-p(x)) = p(x) - q(x) = a(x)$, it follows that $a(x)$ is also the difference between the probability that a D-strategist meets a D-strategist and

[a]The algebra found here applies to any symmetric two-person, two-strategy game, including chicken, battle of the sexes, or the stag-hunt game. For concreteness, our discussion will focus on Prisoner's Dilemma games.

the probability that a C-strategist meets a D-strategist. Thus for either type, $a(x)$ is the difference between the probability that one meets one's own type and the probability that a member of the other type meets one's own type.

Rearranging terms in Eq. (2.1), and substituting from the definition in (2.2), we find that

$$q(x) = x[1 - (p(x) - q(x))] \tag{2.3}$$

$$= x(1 - a(x))\,. \tag{2.4}$$

From Eqs. (2.2) and (2.4), it follows that

$$p(x) = a(x) + x(1 - a(x))\,. \tag{2.5}$$

2.2.2. *A historical digression on measuring assortativity*

Sewall Wright (1921) defined the 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 (1981) interpret this 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 $p(x) = m + x(1 - m)$. The parameter m can be shown to be the coefficient of correlation between indicator random variables for possession of the trait for pairs of mates.[b] From Eq. (2.5), we see that in the special case where $a(x) = m$ is a constant, Wright's coefficient of correlation and Cavalli-Sforza's and Feldman's "fraction of the population that mates assortatively" are formally equivalent to the *index of assortativity* defined in this paper.

2.3. *Evolutionary dynamics and comparing payoffs*

We assume that at any time, the growth rate of the proportion of cooperators in the population is positive, zero, or negative, depending on whether the expected payoff of cooperators is higher than, equal to, or lower than the expected payoff of defectors. Weibull (1995) calls this the assumption of *payoff monotonicity*.[c] To study payoff monotone dynamics, we simply compare the expected payoffs of cooperators and defectors.

[b]Let I_i be a random variable that takes on value 1 if individual i is of the given type and 0 otherwise. For two partners, 1 and 2, the correlation coefficient between I_1 and I_2 is defined to be $(E(I_1I_2) - E(I_1)E(I_2))/(\sigma_1\sigma_2)$ where σ_i is the standard deviation of I_i. Now $E(I_1I_2) = xp(x)$, and for $i = 1, 2$, $E(I_i) = x$ and $\sigma_i = \sqrt{x(1-x)}$. Making the appropriate substitutions, we find that the correlation coefficient is $(xp(x) - x^2)/x(1 - x) = m$.

[c]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 difference between the average payoff to that strategy and the average payoff in the entire population. (Weibull, 1995). The results found in this paper do not require the special structure of replicator dynamics.

With probability $p(x)$ a cooperator meets another cooperator and gets payoff R and with probability $1-p(x)$ a cooperator meets a defector and gets payoff S. Therefore the expected payoff to a cooperator is:

$$p(x)R + (1-p(x))S = S + p(x)(R-S)$$
$$= S + a(x)(R-S) + x(1-a(x))(R-S)\,, \qquad (2.6)$$

where the last expression is obtained by substituting for $p(x)$ from Eq. (2.5). Similar reasoning shows that the expected payoff to a defector is

$$q(x)T + (1-q(x))P = P + x(1-a(x))(T-P)\,. \qquad (2.7)$$

Let us define $\delta(x)$ to be the difference between the expected payoff of a cooperator and that of a defector when the proportion of cooperators in the population is x. Subtracting Eq. (2.7) from Eq. (2.6), we have

$$\delta(x) = S - P + a(x)(R-S) + x(1-a(x))[(R+P)-(S+T)]\,. \qquad (2.8)$$

The function $\delta(\cdot)$ plays the starring role in our study of payoff monotonic dynamics, since for all x between 0 and 1 the sign of the growth rate of the proportion of cooperators is the same as the sign of $\delta(x)$.

2.4. *Prisoner's Dilemma games with additive payoffs*

We can define a special class of Prisoner's Dilemma games in which the benefits of being helped and the costs of helping the other player are "additive". The restrictive assumption in the additive case is that the cost to one player of helping the other is independent of whether the help is reciprocated, and the benefit that a player gets from being helped is independent of whether he is also helping. In this class of games, each player has the option of helping the other player (the C-strategy) or not helping (the D-strategy). A player who helps bears a cost of c and confers a benefit of b on the other player, where $0 < c < b$. If each player helps the other, then both get a benefit of b and both bear costs of c. Thus the payoff to mutual cooperation is $R = b - c$. If one player helps and the other does not, the player who helps bears a cost of c and gets no benefit and the player who doesn't help gets a benefit of b and bears no cost. Thus $S = -c$ and $T = b$. Finally if neither helps, both get payoffs $P = 0$. These payoffs satisfy the inequalities necessary for the game to be a Prisoner's Dilemma, since $S = -c < P = 0 < R = b - c < T = b$.

For additive Prisoner's Dilemma games, we have $R + P = S + T = b - c$ and therefore

$$(R+P) - (S+T) = 0\,. \qquad (2.9)$$

Therefore for additive Prisoner's Dilemma games, Eq. (2.8) for the difference between the expected payoffs of cooperators and defectors simplifies to

$$\delta(x) = S - P + a(x)(R-S) = a(x)b - c\,. \qquad (2.10)$$

According to Eq. (2.10), with additive payoffs, cooperators will do better or worse than defectors depending on whether the product of the index of assortativity times the benefits conferred by help is greater than or less than the cost of helping.

2.5. *Prisoner's Dilemma games with non-additive payoffs*

It is important to understand that not all Prisoner's Dilemma games have additive payoffs and that the evolutionary dynamics can be qualitatively different in the non-additive cases. We will show that every Prisoner's Dilemma game is equivalent to a game in which the two players each contribute costly effort to produce a joint output that is shared equally between them. With this parameterization, the class of additive Prisoner's Dilemma games corresponds to those in which the joint "production function" exhibits constant returns to scale.

2.5.1. *Working and shirking in a game of shared output*

Let us describe the family of *PD games with shared output* as follows. Each player, i, can contribute either *work*, in which case $s_i = 1$ or *shirk*, in which case, $s_i = 0$. Total output is divided equally between the players, regardless of their work effort. We will assume that the amount of output produced is given by a production function Φ that takes the following form:

$$\Phi(s_1, s_2) = 2b(s_1 + s_2) + 2ks_1s_2\,, \tag{2.11}$$

where k is a parameter such that $-\infty < k < c$. The cost of effort for each i is assumed to be $(c + b)s_i$. With this production function, total output when both players work will be greater than, equal to, or less than twice total output when one works and one shirks, depending on whether $k > 0$, $k = 0$, or $k < 0$. These three cases correspond respectively to production processes that are superadditive (increasing returns to scale), additive, (constant returns to scale), and subadditive (decreasing returns to scale).

The payoff to player 1 for a game of shared outputs is given by the function

$$\begin{aligned}\Pi(s_1, s_2) &= \frac{1}{2}\Phi(s_1, s_2) - (c + b)s_1 \\ &= b(s_1 + s_2) + ks_1s_2 - (c + b)s_1\,. \end{aligned}\tag{2.12}$$

For this game we see that $T = \Pi(0, 1) = b$, $R = \Pi(1, 1) = 2b+k-(c+b) = b+k-c$, $P = \Pi(0, 0) = 0$, and $S = \Pi(1, 0) = b - (b + c) = -c$. For all $k < c$, these payoffs satisfy the inequalities $S < P < R < T$. A simple calculation shows that $(R+P)-(S+T) = k$. The game has additive payoffs only where $k = 0$ and in this case, $R + P = S + T$.

To see that every Prisoner's Dilemma game can be described by this parameterization of a *PD game with shared output*, notice that without changing the dynamics,

we can normalize the game so that $P = 0$.[d] For any Prisoner's Dilemma game with $S < P = 0 < R < T$, we can show that there is exactly one set of parameters for a *PD game with shared output* for which these are the payoffs. In particular, a Prisoner's Dilemma game for which the parameters are S, $P = 0$, R, and T will be equivalent to a PD game with shared output if and only if the parameters b, k, and c of the production function and cost function satisfy $b = T$, $c = -S$, and $k = R - (S + T)$.

2.6. *Dynamics with a constant index of assortativity*

As we will show, there are interesting applications in which the index of assortativity, $a(x) = a$ is constant with respect to x. When $a(x)$ is constant, the dynamics are especially simple, since the expression for $\delta(x)$ Eq. (2.8) is seen to be linear in x.

The dynamics are further simplified if payoffs are additive and $a(x)$ is constant. Then, $\delta(x) = ab - c$ is constant for all x. When this is the case, equilibrium must be unique and stable. If $ab > c$, cooperators always have higher expected payoffs than defectors and the only equilibrium is $x = 1$. If $ab < c$, defectors always have higher payoffs and the only stable equilibrium is $x = 0$.

For non-additive Prisoner's Dilemma games when a is constant, the qualitative dynamics are determined by the signs of

$$\delta(0) = [aR + (1 - a)S] - P \text{ and} \tag{2.13}$$

$$\delta(1) = R - [aP + (1 - a)T]\,. \tag{2.14}$$

There are four distinct possibilities.

- The case $\delta(0) > 0$ and $\delta(1) > 0$ occurs when $aR + (1 - a)S > P$ and $R > aP + (1 - a)T$. In this case, $\delta(x) > 0$ for all x between 0 and 1. Therefore cooperators always have higher expected payoffs than defectors and there is a unique stable equilibrium with $x = 1$.
- The case $\delta(0) < 0$ and $\delta(1) < 0$ occurs when $aR + (1 - a)S < P$ and $R < aP + (1 - a)T$. In this case, $\delta(x) < 0$ for all x between 0 and 1. Therefore defectors always have a higher expected payoff than cooperators and there is a unique stable equilibrium with $x = 0$.
- Figure 1 shows the graph of $\delta(x)$ when $\delta(0) > 0$ and $\delta(1) < 0$. If these inequalities are satisfied, cooperators have a higher expected payoff than defectors when cooperators are rare and defectors have a higher expected payoff than cooperators when defectors are rare. We see from Fig. 1 that in this case, the unique equilibrium is a polymorphic population that includes some cooperators and some defectors. This case occurs when $aR + (1 - a)S < P$ and $R < aP + (1 - a)T$.

[d]Since we assume that the population dynamics of Prisoner's Dilemma game depends on expected payoffs, these dynamics will be unchanged if we subtract a constant amount from each payoff so as to make $P = 0$.

- Figure 2 shows the graph of $\delta(x)$ when $\delta(0) < 0$ and $\delta(1) > 0$. If these inequalities are satisfied, cooperators have a higher expected payoff than defectors when $x = 1$ and defectors have a higher expected payoff than cooperators when $x = 0$. In this case, there are two monomorphic equilibria, one with cooperators only and one with defectors only. This case occurs when $aR + (1 - a)S < P$ and $R > aP + (1 - a)T$.

2.6.1. *Interpretation for games of shared outputs*

If we parameterize Prisoner's Dilemma games as games of shared outputs with production function $\Phi(s_1, s_2) = 2b(s_1 + s_2) + 2ks_1s_2$ and cost of effort $(c + b)s_i$, then we have

$$\delta(0) = [aR + (1 - a)S] - P = ab + ak - c \text{ and} \tag{2.15}$$

$$\delta(1) = R - [aP + (1 - a)T] = ab + k - c\,. \tag{2.16}$$

In the additive case, where $k = 0$, $\delta(1) = \delta(0) = ab - c$. Where $k \neq 0$, we have $\delta(1) - \delta(0) = k(1 - a)$. Therefore $\delta(1) - \delta(0)$ is positive in when the production function is superadditive and negative where the production function is subadditive. We see from Figs. 1 and 2 that if $a(x)$ is constant, a stable polymorphic equilibrium is possible only in the case of subadditive production and two distinct stable monomorphic equilibria are possible only in the case of superadditive production.

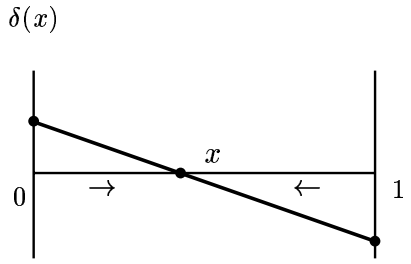

Fig. 1. Unique polymorphic equilibrium.

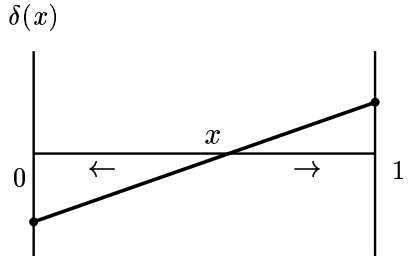

Fig. 2. Two stable equilibria; $x = 0$ and $x = 1$.

3. Kin-selection and Assortative Matching

Biologists are familiar with Eq. (2.10) under the name *Hamilton's rule*. According to W. G. Hamilton's (1964) theory of *kin selection*, natural selection will favor individuals who are willing to help a genetic relative if and only if the product of the benefits from helping times the *coefficient of relatedness* between the two relatives exceeds the cost of helping. Hamilton defines the coefficient of relatedness between two individuals as 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.

Hamilton's coefficient of relatedness plays the same formal role in his theory as our "index of assortativeness" for the case of additive Prisoner's Dilemma games. We will see that this is no accident. The examples below show how the index of assortativeness can be calculated for siblings under a variety of assumptions about mechanisms of cultural or genetic inheritance. It is interesting to notice that in all of these examples, the index of assortativeness, $a(x)$, turns out to be constant with respect to x.

3.1. *Sexual haploid siblings*

Let us consider a population in which children play a Prisoner's Dilemma game with their siblings. Each child has two parents and inherits its strategy for this game from one of its parents, chosen at random. Since for most animals, including humans, genetic inheritance is sexual diploid rather than sexual haploid, the sexual haploid model is more suitable as a model of culturally transmitted rather than genetically transmitted characteristics.[e]

3.1.1. *A symmetric case with monogamy and random mating*

Suppose that parents mate monogamously and randomly with respect to the strategy that they play with siblings. Assume that each child is equally likely to inherit its strategy from its mother or its father and that the parent copied by one child is statistically independent of that copied by its siblings.

Where x is the proportion of cooperators in the entire population, let us calculate the probability that a randomly chosen sibling of a cooperator is also a cooperator. With probability 1/2, the sibling will have the same role model as the cooperator, in which case the sibling will surely be a cooperator. With probability 1/2, the sibling will have a different role model. Since the parents are assumed to be mated randomly with respect to their strategies toward siblings, the probability that the

[e]A sexual diploid individual carries two alleles, one inherited from its mother and one inherited from its father, in each genetic locus. The two alleles found at a genetic locus then determine the effect of this locus on the individual's characteristics.

other parent is a cooperator will be x. Therefore the probability that a cooperator has a sibling who is also a cooperator is

$$p(x) = \frac{1}{2} + \frac{1}{2}x\,. \tag{3.1}$$

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. Therefore we have

$$q(x) = \frac{1}{2}x\,, \tag{3.2}$$

and

$$a(x) = p(x) - q(x) = \frac{1}{2}\,. \tag{3.3}$$

3.1.2. *Assortative mating and extra-familial influence*

Suppose that all else is as in the previous section, but that there is assortative mating between parents. The degree of assortativeness in mating can be defined with the same formalism that we used to define the degree of assortativeness in matching for game-playing encounters. In particular, the degree of assortativeness in mating is a number m between 0 and 1, such that when the proportion of C-strategists in the population is x, the probability that a C-strategist mates with another C-strategist is $x + m(1 - x)$.

Now we can calculate the probablility $p(x)$ that a random sibling of a cooperator child is also a cooperator. A cooperator child has at least one cooperator parent, whose strategy the child imitates. With probability 1/2, the child's sibling imitates the same parent and is also a cooperator. With probability 1/2, the sibling imitates the other parent. Given the degree m of assortativeness in mating, the other parent will be a cooperator with probability $x + m(1 - x)$. Therefore the probability that a random sib of a cooperator child is also a cooperator is

$$p(x) = \frac{1}{2} + \frac{x + m(1-x)}{2} = \frac{1+m}{2} + \frac{(1-m)x}{2}\,. \tag{3.4}$$

A sibling of a defector child is a cooperator only if the two siblings have different role models and if the sibling's defector parent is mated with a cooperator. The probability of this event is

$$q(x) = \frac{(1-m)x}{2}\,. \tag{3.5}$$

Therefore the of assortativity between children is simply

$$a(x) = p(x) - q(x) = \frac{1+m}{2}\,. \tag{3.6}$$

Another wrinkle that can be added to this calculation is to suppose that with some probability v, a child copies neither of its parents but rather chooses a random member of the population to copy. In this case it is not hard to show that

$$p(x) = v\left(\frac{1}{2} + \frac{x + m(1-x)}{2}\right) + (1-v)x\,, \tag{3.7}$$

$$q(x) = v\frac{(1-m)x}{2} + (1-v)x\,, \tag{3.8}$$

and hence

$$a(x) = p(x) - q(x) = \frac{v(1+m)}{2}\,. \tag{3.9}$$

3.1.3. *Some asymmetry and some polygamy*

Consider a partially polygamous population where the probability that two children of the same mother have the same father is μ. Suppose that children copy their strategy from their mother with probability λ and their father with probability $1-\lambda$.

We first calculate the probability that a random sibling of a cooperator is a cooperator. This can happen in any of the following ways: both siblings inherit their strategies from their mother, both siblings inherit their strategies from the same father, or the two siblings inherit their strategies from different parents. Adding the relevant probabilities, we find that

$$p(x) = \lambda^2 + (1-\lambda)^2\mu + 2\lambda(1-\lambda)x + (1-\lambda)^2(1-\mu)x\,. \tag{3.10}$$

A defector's sibling will be a cooperator only if the two siblings inherit their strategies from different parents and the defector's sibling inherits its strategy from a cooperator. The probability of this happening is

$$q(x) = 2\lambda(1-\lambda)x + (1-\lambda)^2(1-\mu)x\,. \tag{3.11}$$

Thus we have

$$a(x) = p(x) - q(x) = \lambda^2 + (1-\lambda)^2\mu\,. \tag{3.12}$$

3.2. *Sexual diploid siblings*

Most sexually-reproducing creatures reproduce by diploid rather than haploid genetics. In each genetic locus, an individual carries two alleles, one inherited from its father and one inherited from its mother. The allele inherited from each parent is randomly selected from that parent's allele pair. An individual's *genotype* is a specification of the two genes that it carries. If there are two allele types A and a, then there are three possible genotypes, AA, Aa, and aa. An individual who carries two alleles of the same type is said to be a *homozygote* and one who carries two different types of alleles is said to be a *heterozygote*. The allele A is said to be *dominant* over allele a if heterozygotes of genotype Aa use the same strategy as homozygotes of type AA.

The full dynamics of polymorphic equilibria in diploid populations is complex and in general seems not to be amenable to the simple methods discussed in this paper. However, our methods work well to characterize necessary conditions for a monomorphic equilibrium to be stable against *dominant* mutant alleles A monomorphic population is one in which all individuals, except for rare mutants, are homozygotes of the same type. A monomorphic equilibrium is stable against dominant mutant alleles if, so long as the number of mutant alleles is small, the expected payoff to players who carry the mutant allele is smaller than that of players who are homozygotes with the normal allele.

Suppose that there are two types of alleles C and D, and that individuals of genotype CC play cooperate, while individuals of genotype DD play defect in games with their siblings. Let x denote the fraction of C genes in the population, let $p(x)$ be the probability that a random sibling of a child who plays cooperate also plays cooperate and let $q(x)$ be the probability that a random sibling of a child who plays defect will play cooperate. We can calculate an index of assortativity, $a(x)$ for the limiting values of $x = 1$ and $x = 0$. With these values in hand, we can determine conditions under which a monomorphic equilibrium population of cooperators or of defectors is stable.

Consider a monomorphic population of CC genotypes. This population is stable only if individuals who carry a dominant mutant D allele receive a lower expected payoff than the normal CC genotypes. Let us assume that mating is sufficiently random with respect to genotype so that when the fraction of carriers of mutant genes is small, the probability that an individual who carries the mutant allele mates with another carrier of the mutant allele is nearly zero. Then so long as the mutant D allele is rare, almost every child that is born with the mutant allele has one heterozygote CD parent and one homozygote CC parent.

When x is nearly one, almost every cooperator child has two homozygote CC parents, and almost every defector child has one homozygote CC parent and one heterozygote CD parent. A cooperator child is therefore almost certain to have a cooperator sibling. A defector child's sibling is sure to inherit a C allele from its CC parent, but the probability is only $1/2$ that it will receive a C allele from the CD parent. Thus in the limit for x close to 1, we have $p(x) = 1$ and $q(x) = 1/2$ and hence the limiting value of $a(x) = p(x) - q(x)$ is $1/2$.

Since $\lim_{x\to 1} a(x) = 1/2$, it follows from Eq. (2.8) that the limiting value of the difference in the expected payoffs of cooperators and of defectors is

$$\lim_{x\to 1} \delta(x) = R - \frac{1}{2}(P+T)\,. \tag{3.13}$$

From Eq. (3.13), it follows that a monomorphic population of C alleles will be stable against invasion by dominant mutant D alleles if and only if $R > (P+T)/2$.

Now let us explore conditions under which a monomorphic population of D alleles is stable against invasion by dominant mutant C alleles. When x is nearly zero, almost all defector offspring will have two homozygote DD parents, and thus their siblings will almost certainly be defectors. When x is nearly zero, a cooperator

child will almost certainly have one heterozygote CD parent and one homozygote DD parent. Its sibling will inherit a D allele from the homozygote parent, and will inherit either a C or D allele from the heterozygote parent with equal probability. Since the C allele is assumed to be dominant, the sibling will cooperate with probability close to 1/2. Therefore the limiting value of $a(x) = p(x) - q(x)$ as x approaches zero is again equal to 1/2. Applying Eq. (2.8) once again, we have

$$\lim_{x\to 0} \delta(x) = \frac{1}{2}(R+S) - P\,. \tag{3.14}$$

From Eq. (3.14), it follows that a monomorphic population of D alleles will be stable against invasion by dominant mutant C alleles if and only if $(R+S)/2 < P$.

For Prisoner's Dilemma games with additive payoffs, we have $R = b-c$, $P = 0$, and $T = b$. In this case, the condition for a monomorphic population of cooperators to be an equilibrium is that $b/2 > c$ and the condition for a monomorphic population of defectors to be an equilibrium is $b/2 < c$. Thus, in the additive case where $b/2 \neq c$, we have either a monomorphic equilibrium of cooperators or a monomorphic population of defectors but not both. Which type of behavior prevails in equilibrium is determined by Hamilton's Rule.

In an earlier paper (Bergstrom, 1995), I worked out necessary conditions for a monomorphic equilibrium to be stable against *recessive* mutant alleles. In this case, it can be shown that any carrier of a rare recessive allele is almost certainly the offspring of two heterozygote parents. Some of the rare recessive genes in the population will appear in heterozygote children who behave like members of the normal population and some will appear in homozygotes who behave differently. Calculations show that a recessive mutant defector gene can invade a monomorphic population of cooperators if $\frac{1}{5}P + \frac{3}{5}T + \frac{1}{5}S > R$. A recessive mutant cooperator gene can invade a monomorphic population of defectors if $\frac{1}{5}R + \frac{3}{5}S + \frac{1}{5}T > P$. In the special case of a Prisoner's Dilemma game with additive payoffs, these conditions reduce to $b/2 < c$ and $b/2 > c$ respectively. Thus for additive Prisoner's Dilemma games a monomorphic population will repel invasions of both recessive and dominant mutants according to the dictates of Hamilton's Rule.

4. Assortative Matching with Partner Choice

4.1. *A model of labelling*

Interesting possibilities for assortative matching arise when players have some choice about their partners. In a game of Prisoner's Dilemma, everyone would prefer to be matched with a cooperator rather than with a defector. Let us assume that each player can make a fixed number of matches, that search costs are negligible, and that a match requires mutual consent of the two matched players. If players' types were observable with perfect accuracy, then the only equilibrium outcome would have cooperators matched only with cooperators and defectors with defectors. In this case, $a(x) = 1$ for all x and therefore cooperators receive a payoff R while

defectors receive $P < R$. Thus a population consisting only of cooperators would be the only stable equilbrium.

But suppose that detection is less than perfectly accurate. 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.[f]

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, apparent cooperators will all be matched with apparent cooperators and apparent defectors with apparent defectors.

4.2. *Calculating the index of assortativity*

The proportion of actual cooperators among apparent *cooperators* is

$$C_c(x) = \frac{\alpha x}{\alpha x + (1-\beta)(1-x)}\,, \tag{4.1}$$

and the proportion of actual cooperators among apparent *defectors* is

$$C_d(x) = \frac{(1-\alpha)x}{(1-\alpha)x + \beta(1-x)}\,. \tag{4.2}$$

There are two possible ways in which a cooperator can be matched with another cooperator. One way is that the cooperator is correctly labelled as a cooperator and the apparent cooperator that it is matched with is an actual cooperator. The other way is that the cooperator is mislabelled as a defector, but has the good fortune to be matched with a cooperator that has been mislabelled as a defector. Adding the probabilities of these two outcomes, we find that

$$p(x) = \alpha C_c(x) + (1-\alpha)C_d(x)\,. \tag{4.3}$$

Similarly, a defector can be matched with a cooperator in either of two ways. The defector can be mislabelled as a cooperator and be matched with a correctly labelled cooperator, or the defector can be correctly labelled as a defector and matched with a mislabelled cooperator. Thus we have

$$q(x) = (1-\beta)C_c(x) + \beta C_d(x)\,. \tag{4.4}$$

[f]It would be interesting to pursue an expanded model in which ability to detect and ability to deceive were also subject to evolutionary pressures.

We can calculate the index of assortativity, which is

$$a(x) = p(x) - q(x) = (\alpha + \beta - 1)(C_c(x) - C_d(x))\,. \tag{4.5}$$

It is straightforward to verify that $C_c(0) = C_d(0) = 0$ and $C_c(1) = C_d(1) = 1$ and therefore, according to Eq. (4.4), $a(0) = 0$ and $a(1) = 0$. For this model, in contrast to the examples that we have looked at so far, the index of assortativity is not constant with respect to x. Applying simple calculus to Eq. (4.4), we find that $a'(0) > 0$, $a'(1) < 0$. We also find that $a''(x) < 0$ for all x between 0 and 1, which implies that $a(x)$ is a concave function, that slopes upwards at $x = 0$, reaches a maximum somewhere between 0 and 1, and then slopes downward until it reaches $x = 1$.

In the special case where $\alpha = \beta$, $a(x)$ attains its maximum value of $(2\alpha - 1)^2$ at $x = 1/2$. Figure 3 shows the qualitative nature of the graph of $a(x)$ when $\alpha = \beta$.

There is a simple intuitive explanation of 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 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.

In the case where $\alpha > \beta$, so that a cooperator is more likely to be correctly labelled than is a defector, the graph of $a(x)$ remains convex downward but the peak occurs at $x > 1/2$, as in Fig. 4. Simple calculations show that in the limiting case as α approaches 1, the peak of the graph occurs arbitrarily close to the point where $x = 1$ and $a(x) = \beta$.

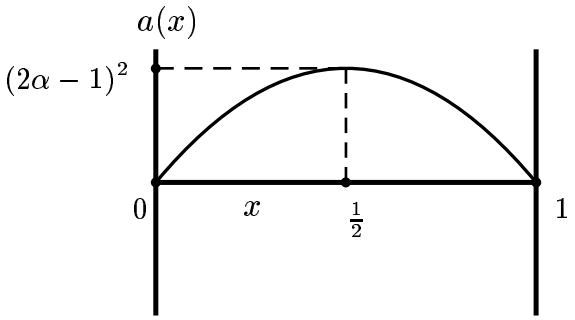

Fig. 3. Graph of $a(x)$ where $\alpha = \beta$.

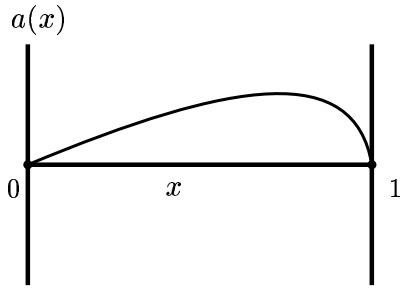

Fig. 4. Graph of $a(x)$ with $\alpha > \beta$.

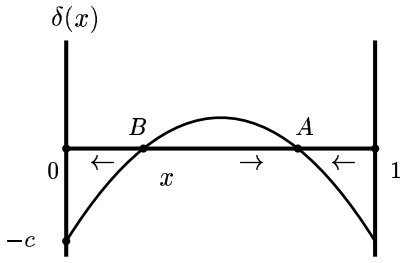

Fig. 5. Graph of $\delta(x)$ for additive Prisoner's Dilemma.

4.3. *Population dynamics*

We can use what we know about the function $a(x)$ to analyze the function $\delta(x)$ that expresses the difference between the average payoff to cooperators. Since $a(0) = a(1) = 0$, it must be that $\delta(0) = \delta(1) = -c < 0$. Thus we know that cooperators get lower payoffs than defectors in monomorphic populations, consisting either entirely of defectors or of cooperators. This implies that a population consisting entirely of defectors is locally stable and that a population consisting entirely of cooperators is locally unstable. There remains however, the possibility for a stable polymorphic equilibrium with some cooperators and some defectors.

Figure 5 graphs $\delta(x)$ for a Prisoner's Dilemma with additive payoffs. We have drawn this figure so that $\delta(x)$ is positive for some values of x between zero and 1. The small arrows show directions of movements strarting from any intitial position. There are two locally stable equilibria. One stable equilibrium occurs at the point where $x = 0$. 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.

For Prisoner's 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 Fig. 5. 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.

For Prisoner's Dilemma games with non-additive payoffs, it remains true that $\delta(0) = \delta(1) < 0$, $\delta'(0) > 0$, and $\delta'(1) < 0$. It is not necessarily the case however that $\delta''(x) < 0$ on the interval $[0, 1]$, so the graph of $\delta(x)$ need not be single-peaked. In this case, there may be more than one polymorphic equilibrium, but it still must be that if $\delta(x) > 0$ for some x, then there is at least one locally stable polymorphic equilibrium and one locally stable monomorphic equilibrium with $x = 0$.

5. Related Literature

W. D. Hamilton's classic paper (1964) on the evolution of social behavior among relatives showed that for some kinds of interactions between relatives, natural selection will lead to a degree of cooperativeness that can be quantified in terms of genetic relatedness. In later work, Hamilton (1971; 1975) observed that the theory of kin selection can be viewed as a special case of assortative matching between individuals who are not necessarily related genetically.

Wilson and Dugatkin (Wilson, 1997) consider assortative matching in N-player groups where players are programmed with possible strategies that can be described by a real number x from some specied interval. An individual with strategy x bears a personal cost cx and confers benefits of bx on *every* member of the group, where it is assumed that $Nb > c > b$.[g] The authors point out that everyone will prefer to belong to a group consisting of others with higher x. Therefore if group membership requires mutual consent and if individual x's are common knowledge, groups would segregate exactly according to their values of x and those with the highest value of x would receive the highest payoffs. They show through simulations that where individual types are imperfectly known so that assortativity is partial, there can be selection pressure for high individuals with high x.

Myerson, Pollock, and Swinkels (1991) introduce a *viscosity parameter* that is equivalent to a constant index of assortativity as defined in this paper. For a population with a finite number of pure strategies, they define a *δ-viscous population equilibrium* to be an assignment of proportions of the population playing each strategy such that for each strategy played by a positive proportion, this strategy is a best response to the current strategy mix under the assumption that an individual will meet its own type with probability δ and will meet a randomly selected member of the overall population with probability $1 - \delta$. Thus a δ-viscous equilibrium is a symmetric Nash equilibrium for a game in which the payoff to a player is his expected payoff if with probability δ he encounters a someone playing the same

[g]This game is equivalent to the "voluntary provision of public goods game" that is much studied in experimental economics (Ledyard, 1995).

strategy as his own and with probability $1 - \delta$ he encounters a randomly chosen member of the entire population.[h]

Bergstrom (1995) shows that for symmetric games between relatives in a population of sexual diploids, a monomorphic equilibrium will be stable against invasion by dominant mutants and by recessive mutants if each player acts so as to maximize a "semi-Kantian" expected utility function that assigns a probability weight k to the event that one's opponent mimics one's own behavior and $1 - k$ to the event that one's opponent is a random draw from the population, where k is defined by the genetic coefficient of relatedness between the two relatives.

The published papers that seem to be most closely related to this one are Eshel and Cavalli-Sforza[i] (1982) and by Bergstrom and Stark (1993). Both papers study models with assortative matching of players in Prisoner's Dilemma games and consider examples in which sexual haploid and sexual diploid players recognize genetic kin. Eshel and Cavalli-Sforza also study a model in which players are able to actively choose partners who are likely to treat them favorably, though their model of partner selection differs from ours in that they emphasize search costs rather than inaccuracy of determining others' types.

The current paper suggests a unifying method for treating a great variety of models of assortative encounters and I think offers a simpler, more transparent method of deriving both known and new results.

Acknowledgment

This paper has benefited from comments by participants in the Evolutionary Game Theory workshop held at the University of Odense, Denmark in September 2000 and a workshop on Groups, Multi-level Selection and Economic Dynamics, held at the Santa Fe Institute in January of 2001. Special thanks for helpful remarks are due to Marc Feldman, Hillard Kaplan,Thorbjørn Knudsen, Chris Proulx, and David Sloan Wilson.

References

Bergstrom, T. C. and O. Stark (1993). "How altruism can prevail in an evolutionary environment," *American Economic Review*, Vol. 83, No. 2, 149–155.

Bergstrom, T. C. (1995). "On the evolution of altruistic ethical rules for siblings," *American Economic Review*, Vol. 85, No. 1, 58–81.

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

Eshel, I. and L. L. Cavalli-Sforza (1982). "Assortment of Encounters and Evolution of Cooperativeness," *Proceedings of the National Academy of Sciences*, Vol. 79, 1331–1335.

[h]The authors also propose an interesting Nash equilibrium refinement which selects only those Nash equilibria that are the limit of δ-viscous equilibria as δ approaches zero.

[i]Toro and Silio (Toro and Silio, 1986) offer interesting extensions of Eshel and Cavalli-Sforza's paper.

Hamilton, W. D. (1964). "The genetical evolution of social behavior, Parts I and II," *Journal of Theoretical Biology*, Vol. 7, 1–52.

Hamilton, W. D. (1971). "Selection of Selfish and Altruistic Behavior in Some Extreme Models," in J. F. Eisenberg and W. S. Dillon (eds.), *Man and Beast*: *Comparative Social Behavior*, Washington, D.C.: Smithsonian Press, 57–91.

Hamilton, W. D. (1975). "Innate Social Aptitudes of Man: An Approach from Evolutionary Genetics," in R. Fox (ed.), *Biosocial Anthropology*, London: Malaby Press, 133–155.

Ledyard, J. (1995). "Public Goods: A Survey of Experimental Research," in J. H. Kagel and A. E. Roth (eds.), *The Handbook of Experimental Economics*, Princeton, N.J.: Princeton University Press, Ch. 2, 111–181.

Myerson, R. B., G. B. Pollock and J. M. Swinkels (1991). "Viscous population equilibrium," *Games and Economic Behavior*, Vol. 3, 101–109.

Toro, M. and L. Silio (1986). "Assortment of encounters in the two-strategy game," *Journal of Theoretical Biology*, Vol. 123, 193–204.

Weibull, J. (1995). *Evolutionary Game Theory*, Cambridge, MA: MIT Press.

Wilson, D. S. and L. A. Dugatkin (1997). "Group selection and assortative interactions," *The American Naturalist*, Vol. 149, No. 2, 336–351.

Wright, S. (1921). "Systems of mating," *Genetics*, Vol. 6, No. 2, 111–178.