On the Evolution of Altruistic Ethical Rules for Siblings

by

Ted Bergstrom
University of Michigan

Current version: July 27, 1993

On the Evolution of Altruistic Ethical Rules for Siblings

Ted Bergstrom

> *‘‘If you have remarked errors in me, your superior wisdom must pardon them. Who errs not while perambulating the domain of nature? Who can observe everything with accuracy? Correct me as a friend and I as a friend will requite with kindness.’’*
> *--Linnaeus*

1. Introduction

Economists normally assume that people act perfectly selfishly toward strangers, but exhibit some generosity toward their children and perhaps toward their spouses. This premiss is taken as a basic axiom of the theory rather than deduced from a more fundamental model of human nature. In recent years there have been a few interesting efforts to draw conclusions about the nature of human preferences from exploration of models of the evolutionary origins of preferences. Jack Hirshleifer’s engaging manifesto (1978) has been influential in inspiring further work. Evolutionary foundations for altruism in the family were explored by Gary Becker (1976), Robert Frank (1988) discussed the evolutionary origins of emotions, Ingemar Hansson and Charles Stuart (1990) discussed the evolution of attitudes toward saving and work effort, Arthur Robson (1992a,1992b) suggested a theory of the evolution of preferences toward risk, and H. Peyton Young (1993) developed an evolutionary theory of social conventions.[1]

Evolutionary biologists such as William Hamilton (1964a, 1964b), Richard Dawkins (1976), John Maynard Smith (1978), and Robert Trivers (1985) have suggested a theoretical and empirical basis for the evolution of altruism between close relatives. Dawkins’ expression of this view in *The Selfish Gene*, is that in evolution, the replicating agent is the *gene* rather than the animal. Since a gene carried by one animal is likely to appear in its relatives, it follows that a gene that makes an animal help its relatives, at least when it is cheap to do so, will prosper relative to genes for totally selfish behavior.

The work reported here is intended to be a contribution both to economics and to evolutionary biology, but the current paper is written with economist readers in mind. At the most direct level this is a contribution to the *economics of the family*, offering an evolutionary explanation for the degree of altruism to be expected between siblings. Apart from its direct applications to economics, the biological model of evolution with kin selection is likely to interest economists for its own sake. This is an elegant logical structure, sufficiently similar to models of economic equilibrium to strike chords of familiarity, yet sufficiently different to inspire fresh ways of thinking about economic and social problems. The synthesis of Darwinian evolution with the Mendelian model of diploid inheritance is a rich prototype for cultural evolution, where norms and culture arise as people copy the actions of others to whom they are related through social rather than biological

---

I am grateful to Carl Bergstrom for teaching me some rudiments of population genetics and suggesting useful lines of inquiry.

[1] (x, y)

structures.

Several recent papers by game theorists (e.g. Daniel Friedman (1991), Ken Binmore and Larry Samuelson (1992), Michihiro Kandori, George Mailath, and Rafael Rob (1993) and Young (1993)) have proposed evolutionary foundations for solution concepts in the theory of games.[2] In these models, individuals are ‘‘programmed’’ to play a fixed strategy in a game. They encounter randomly selected members of a population of other programmed strategists, play their own predetermined strategy against their opponent’s strategy and receive their payoffs. Individuals programmed to use a given strategy are replicated in the population at a rate that is positively related to the payoffs they get in these encounters. Equilibrium is a rest point of the dynamic process so described. The evolutionary approach taken in the current paper is similar, but here one’s opponents in the game are relatives rather than randomly selected individuals. Because encounters are between relatives, it is necessary to pay explicit attention to the Mendelian genetics of a sexually reproducing species in a way that has not been attempted in previous economics papers.

As a contribution to evolutionary theory, this paper extends Hamilton’s model of kin-selection from the special case of linear costs and benefits to a much richer class of games allowing nonlinear interaction of individual actions, costs and benefits. It also extends the work of Maynard Smith on evolutionary games to the case where the players are relatives. Unlike most previous efforts to apply the theory of evolutionary games to interaction among siblings, this paper makes explicit use of the Mendelian combinatorics of diploid sexual reproduction and addresses the possibility of invasion by recessive as well as mutations.

2. The Game, the Players, and the Method of Reproduction

Strategies and the Game Siblings Play

This paper concerns interactions between siblings, where each pair of siblings plays a symmetric, two-person game.[3] Individuals are assumed to be able to distinguish their siblings from others, and the strategies they play in games with siblings are determined independently of the strategies used in games they may play with others. There is a set S of possible strategies for each individual such that when one sibling plays strategy $x \in S$ and the other plays strategy $y \in S$, the payoff to the sibling who played x is given by $\Pi(x, y)$ and the payoff to the sibling who played y is $\Pi(y, x)$.

---

[2] A very readable introduction to the theory of evolutionary games is found in Binmore (1992)

[3] Undoubtedly there is much to be learned from the study of a more general class of n-person games, simultaneously involving several relatives and allowing the possibility of asymmetries. A striking example is the ‘‘milkshake model’’ proposed by David Haig (1992). In Haig’s milkshake model, unborn children ‘‘play a game’’ with their mother and with siblings who have not yet been conceived. The embryo comes into conflict with its mother, who acts in part as an agent of the interests of her future offspring, over the extent to which the mother should sacrifice her own health for the interests of the child in her womb. To quote Haig, ‘‘A simple analogy may be helpful. Suppose that a mother buys a milkshake to be shared among her children, but the milkshake has only a single straw. If the first child takes a drink and passes the remainder on to the second, and so on down the line, then the greater the consumption of each child, the fewer children receive a drink.’’

Individuals do not consciously ‘‘choose’’ strategies; their actions are programmed by their genetic structure. Natural selection acts on the distribution of strategies in the population. The probability that an individual survives to reproduce will be higher the greater the average payoff that it receives in the games it plays with its siblings. Offspring tend to be like their parents, according to the rules of Mendelian inheritance for sexual reproduction.

This paper is concerned with ‘‘large’’ populations for which it is assumed that expected proportions are always realized. This makes it possible to treat the dynamical system as a deterministic system of difference equations.[4] Mating is assumed to be monogamous, so that siblings have two parents in common, and mate selection is assumed to be random with respect to the genes that control behavior toward siblings. It is assumed that an individual either dies without having any offspring or survives to mate and have exactly n offspring.

Sexual Diploid Reproduction and Stable Monomorphic Equilibrium

In *sexual diploid* species like our own, every individual carries two genes at each genetic locus. One of these genes is inherited from each parent. This gene is drawn randomly from the two genes that the parent carries at the corresponding locus. Following common practice in the biological literature, we use a ‘‘single locus model’’, in which the trait of interest is determined by the two gene copies in one genetic locus. An individual’s *genotype* is specified by the contents of this locus. An individual with two identical gene copies is said to be a *homozygote*. An individual with two different kinds of genes in this locus is said to be a *heterozygote*. Let a and A denote two different genes that could appear in the locus that determines behavior toward siblings. Suppose that type aa homozygotes choose strategy x and type AA homozygotes choose strategy $y \neq x$. If heterozygotes of genotype Aa choose choose strategy y, the gene A is said to be *dominant* (over a) and the gene a is said to be *recessive* (to A).

A population consisting of homozygotes, all of the same genotype, is called a *monomorphic population*. This paper will be mainly concerned with *stable monomorphic equilibrium*. A stable monomorphic equilibrium is a monomorphic population in which any mutant genes that might appear would reproduce less rapidly than the predominant type of genes.[5]

3. Kin Selection, Kantian Ethics, and Hamilton’s Rule

---

[4] In contrast, Foster and Young (1990), Young (1993), and Kandori et.al. (1993) study limiting probability distributions as time goes to infinity in models where evolution is described by stochastic differential equations. It would be interesting and useful to extend the models discussed here to a truly stochastic environment.

[5] One can also study *polymorphic equilibrium* in which more than one type of gene is present. In polymorphic equilibrium, the coexisting genes all reproduce at the same rate, while any mutant genes reproduce less rapidly than the types already present in equilibrium. Carl Bergstrom and Bergstrom (1993) examine polymorphic equilibria for the special case of games in which there are only two possible strategies.

Asexual Reproduction and the Kantian Golden Rule

In order to understand the logic of evolutionary kin selection in sexually reproducing species, it is helpful to look at the simpler case of asexual (parthenogenic) reproduction. Consider a population in which each individual who survives to reproductive age produces two daughters, who play a symmetric game with each other. The strategy that each daughter plays is controlled by a single gene, inherited from her mother. Therefore, except in the case of rare mutations, any two sisters will choose the same strategy as their mother did in the game she played with *her* sister. The probability that an individual survives to produce daughters of her own is higher, the higher the payoff she receives in the game played with her sister. Where S is the set of possible strategies, suppose that there is a strategy $\bar{x} \in S$ such that the function $\Pi(x, x)$ is strictly maximized at $x = \bar{x}$. Suppose that in a population where all individuals are programmed to take action $\bar{x}$, a mutant appears who takes action $x \neq \bar{x}$. Although the mutant may, herself, get a higher payoff than normal individuals, her progeny will not be able to invade the population. Each of the mutant's daughters will, like their mother, take action x, and therefore each gets a payoff of $\Pi(x, x) < \Pi(\bar{x}, \bar{x})$. Not only will the mutant's daughters get lower payoffs and have lower survival probabilities than the daughters of normal $\bar{x}$-strategists, but so will her granddaughters and of all their descendants. Therefore the line of x-strategists will reproduce less rapidly than the line of $\bar{x}$-strategists and will ultimately become arbitrarily small as a fraction of the population. This demonstrates that a monomorphic population of $\bar{x}$-strategists cannot be invaded by a mutants using a different strategy. Similar reasoning shows that a population in which individuals using strategies other than $\bar{x}$ will eventually be overrun by mutants who use strategy $\bar{x}$.

The argument just made shows that for asexually reproducing species, in equilibrium, an individual does not take her Nash equilibrium action, but rather takes the action that maximizes her survival probability given that her sister will choose the same action that she does. Sisters act toward each other as if they were following Immanuel Kant's *Categorical Imperative.*

The Kantian Golden Rule for Siblings. *"Choose the action that would be in your best interest if you were certain that your sibling's action would mimic your own."*

Sexual Reproduction and the Semi-Kantian Golden Rule

Careful observers of our own species will not be surprised to find that with sexual reproduction, equilibrium behavior for siblings is less perfectly cooperative. Consider a monomorphic population of aa genotypes, who play the strategy $\bar{x}$ with their siblings. Suppose that a mutant gene A appears and that individuals of genotype Aa play the strategy $x \neq \bar{x}$. Since mating is assumed to be random, individuals of genotype Aa who survive to reproduce will almost certainly mate with normal aa genotypes. If one parent is of genotype aa and the other of genotype Aa, each of their offspring will be of genotype Aa with probability 1/2 and of genotype aa with probability 1/2.

The proportion of the population carrying the mutant gene will increase or decrease, depending on whether an Aa genotype who is born to one Aa parent and one aa parent gets a larger or smaller expected payoff than normal offspring of normal parents. The payoff to an Aa genotype depends not only on its own action x, but also on the actions taken by its siblings. A sibling of an Aa offspring will be of genotype Aa with probability 1/2, and of genotype aa with probability 1/2. Therefore the expected payoff to an Aa genotype in the game played with each sibling is:

$$V(x,\bar{x}) = \frac{1}{2}\Pi(x,x) + \frac{1}{2}\Pi(x,\bar{x}). \tag{1}$$

The normal population consists of pairs of siblings of genotype aa, who take action $\bar{x}$. Their expected payoff is $\Pi(\bar{x},\bar{x}) = V(\bar{x},\bar{x})$. It follows that when A genes are rare, the Aa genotypes reproduce more rapidly than the normal population if $V(x,\bar{x}) > V(\bar{x},\bar{x})$. Therefore a necessary condition for a monomorphic population of $\bar{x}$-strategists to resist invasion by mutant heterozygotes is that $V(\bar{x},\bar{x}) \geq V(x,\bar{x})$ for all $x \in S$. This means that in equilibrium, everyone takes an action $\bar{x}$ is a symmetric Nash equilibrium strategy for the game with payoff function V. This remarkable fact allows us to predict that in equilibrium, individuals will act *as if* they were maximizing a utility function of the form $V(x,\bar{x})$. The payoff function V is ‘‘semi-Kantian’’, since it is an average of the ‘‘Kantian ’’ payoff $\Pi(x,x)$ and the ‘‘Nashian’’ payoff $\Pi(x,\bar{x})$. Stated in literary language, the precept by which siblings will abide is:

The Semi-Kantian Golden Rule for Siblings. *“Choose the action that would be in your best interest if with probability 1/2 your sibling would imitate you.”*

In more formal dress, this result is:

Proposition 1. *A necessary condition for a monomorphic population that uses strategy $\bar{x}$ to be stable against invasion by dominant mutant genes is that $\bar{x}$ be a symmetric Nash equilibrium for the two-player game with payoff function:*

$$V(x,\bar{x}) = \frac{1}{2}\Pi(x,x) + \frac{1}{2}\Pi(x,\bar{x}).$$

Evolutionary models offer an interesting perspective on the question of whether economic agents should be modelled as decision-makers with utility functions or as simple automatons, hard-wired for particular actions. In the formal model just discussed, individuals are programmed by their genes to perform certain actions. Evolution selects those genes that program ‘‘successful’’ actions. Although individuals are not *free to choose* their actions, natural selection produces a population of individuals that act as if they were consciously maximizing a utility function based on the Semi-Kantian Golden Rule. An alternative modelling strategy would have been to make natural selection act on utility functions rather than on actions. In the stationary environment that has been assumed in this paper, the equilibrium utility functions would have to be functions that rationalize the same action $\bar{x}$ that is an equilibrium for a model of hard-wired actions. Therefore there would be no operational way to distinguish the hypothesis of selection for preferences from the hypothesis of selection for hard-wired actions. But if, over time, the environment changes so as to alter the mapping from actions to survival probabilities, then individuals will be able to improve their survival probabilities by adapting their actions to the current environment. In this case, genes that shape preference functions but allow individuals to exercise choice are more likely to prosper than those that prescribe fixed actions for all times.

The intuitive rationalization for the emergence of an equilibrium in which actions towards siblings are rationalized by preferences that are not purely selfish is simple and appealing. Evolution is a blunt instrument which does not change the genetic structure of organisms in isolation. The Mendelian laws of inheritance imply that with probability 1/2, a gene that influences my behavior will have the same influence on my sibling's behavior. Those individuals who are programmed to treat their siblings relatively well are more likely than the average member of the population to have the good fortune to *receive* the same treatment from their siblings.

Hamilton's Inclusive Fitness Rule

Evolutionary biologists have long been aware that natural selection can favor costly behavior that increases the survival probability of one's siblings and kin. William Hamilton (1964a,b) was the first to formalize this observation in an equilibrium model. Hamilton's main proposition, which has come to be known as ''Hamilton's Rule'' is as follows:

> *'' The social behavior of a species evolves in such a way that in each distinct behavior-evoking situation the individual will seem to value his neighbors' fitness against his own according to the coefficients of relationship appropriate to that situation.'' (1964b, p 19.)*

Since the coefficient of relationship between siblings is 1/2,[6] Hamilton's rule states that in equilibrium, an individual will take an action x that maximizes its *inclusive fitness*, where the inclusive fitness of an individual whose sibling takes action $\bar{x}$ is defined to be:[7]

$$H(x, \bar{x}) = \Pi(x, \bar{x}) + \frac{1}{2}\Pi(\bar{x}, x). \tag{2}$$

Stated more poetically, Hamilton proposed:

Hamilton's Rule for Siblings. *''Love thy sibling half as well as thyself''.*

Hamilton proved his proposition only for the case where benefits and costs interact ''additively''. As applied to the interaction between siblings, this model is as follows:

Hamilton's additive model---The strategy set S for each individual i consists of pairs of numbers (b_i, c_i), such that individual i can add b_i units to its sibling's payoff at a cost of c_i units to itself. In the absence of any helping activity, each sibling has a fitness of a. If the two siblings choose strategies $y = (b_1, c_1)$ and $x = (b_2, c_2)$, then the payoff to Sibling 1 will be $\Pi(y, x) = a + b_2 - c_1$ and the payoff to Sibling 2 will be $\Pi(x, y) = a + b_1 - c_2$.

---

[6] The coefficient of relationship between two individuals is defined by biologists to be the probability that a randomly selected gene in one of these individuals will have an exact copy located in the other individual by descent from a common ancestor. (Trivers (1985)) For parent and child the coefficient or relationship is 1/2, for half-siblings it is 1/4, and for cousins it is 1/8.

[7] Hamilton's paper treats the more general case of simultaneous interaction among an entire network of relatives. The description of Hamilton's theory in this paper is confined to its application to pairs of sexual diploid siblings.

In Hamilton's model, the gains or losses conferred by a relative's actions are independent of one's own actions. As we will see, additivity of benefits and costs is a strong restriction that rules out many interesting games and economic interactions.

Let me Count the Ways

Proper accounting for the gains and losses to genes for ''altruistic'' behavior becomes a subtle matter when these genes determine ones actions toward relatives. The payoff function V registers the fact that those who carry a mutant gene are more likely are than normal individuals to have relatives who carry the mutant gene. The argument for using the payoff function H is that a mutant gene will be more likely to proliferate if it makes an individual help relatives since in doing so it is likely to increase the number of surviving copies of itself. For symmetric games between siblings, the notions embodied in these functions are similar but not identical. In general, a strategy that is a symmetric Nash equilibrium for the payoff function V need not be a symmetric Nash equilibrium for H, nor *vice versa*. Only in special cases such as Hamilton's additive form, do the two different formulations lead to the same predictions of equilibrium.

Some formal ''bookkeeping'' is helpful for understanding the difference between Nash equilibrium for V and Nash equilibrium for H. With these two functions, the gains or losses to a gene that specifies action x when all other genes specify action $\bar{x}$, are respectively:

$$\begin{aligned} V(x,\bar{x}) - V(\bar{x},\bar{x}) &= \frac{1}{2}\Pi(x,\bar{x}) + \frac{1}{2}\Pi(x,x) - \frac{1}{2}\Pi(\bar{x},\bar{x}) - \frac{1}{2}\Pi(\bar{x},\bar{x}) \\ &= \Pi(x,\bar{x}) - \Pi(\bar{x},\bar{x}) + \frac{1}{2}(\Pi(x,x) - \Pi(x,\bar{x})). \qquad (3) \\ H(x,\bar{x}) - H(\bar{x},\bar{x}) &= \Pi(x,\bar{x}) - \Pi(\bar{x},\bar{x}) + \frac{1}{2}(\Pi(\bar{x},x) - \Pi(\bar{x},\bar{x})). \qquad (4) \end{aligned}$$

The first term of each expression is $\Pi(x,\bar{x}) - \Pi(\bar{x},\bar{x})$, which measures the direct increase or decrease in the individual's payoff from taking the deviant action x rather than the normal action $\bar{x}$. The second term in Equation 3 is half of the gain or loss to one's sibling from changing one's own action from the normal action $\bar{x}$ to the deviant action x, *conditional on the sibling also taking the deviant action x.* The second term in Equation 4 is half of the gain or loss to one's sibling from changing one's own action from $\bar{x}$ to x, *conditional on the sibling taking the normal action $\bar{x}$.* The two rules will be in agreement if the effect of one's own action on a sibling's payoff is independent of the action that the sibling takes.

The difference between the rewards to a deviant action under the two payoff functions V and H is closely tied to whether the payoff functions tend to reward players for acting similarly or for acting differently from each other. This consideration arises frequently in other economic contexts, where the distinction is made between strategic complementarity and strategic substitutability in games.

Strategic Complementarity and Substitutability

Stated heuristically, there is strategic complementarity or substitutability, depending on whether the total payoff is higher if both players act similarly or if they act differently. For a two-player game and for any two strategies x and y, one determines whether there is strategic complementarity or substitutability by comparing the expected *total* payoff to the two players in two situations: (i) one of the two strategies is selected at random and both players are made to play the selected strategy (ii) one player is made to play strategy x and the other player is made to play strategy y. There is strategic complementarity if for all strategies x and y, the expected total payoff in situation (i) is at least as great as that in situation (ii); there is strategic substitutability if this inequality is reversed.

Definition. *A payoff function* $\Pi(x, y)$ *is said to exhibit strategic complementarity*[8] *if for all* $x \in S$ *and* $y \in S$*,*

$$\Pi(x, x) + \Pi(y, y) \geq \Pi(x, y) + \Pi(y, x)$$

and to exhibit strategic substitutability if for all $x \in S$ *and* $y \in S$*,*

$$\Pi(x, x) + \Pi(y, y) \leq \Pi(x, y) + \Pi(y, x).$$

When the individual payoff function Π displays strategic complementarity or substitutability, there is a crisply-drawn relationship between the set of Nash equilibria for the semi-Kantian payoff function V and the set of Nash equilibria for the ‘‘Hamiltonian’’ payoff function H. In the Appendix, the following is proved:

Proposition 2. *If the payoff function* Π *exhibits strategic complementarity, then every symmetric Nash equilibrium for the game with payoff function* V *is also a symmetric Nash equilibrium for the game with payoff function* H*. If* Π *displays strategic substitutability, then every symmetric Nash equilibrium for the game with payoff function* H *is also a symmetric Nash equilibrium for the game with payoff function* V*.*

In Hamilton’s additive model, the payoff function Π has the property that for all strategies x and y in S, $\Pi(x, x) - \Pi(y, x) = \Pi(x, y) - \Pi(y, y)$ and therefore $\Pi(x, x) + \Pi(y, y) = \Pi(x, y) + \Pi(y, x)$. In this special case, the payoff function exhibits both strategic complementarity and strategic substitutability. It follows from Proposition 2 that in Hamilton’s additive model, a strategy is a symmetric Nash equilibrium for the payoff function H if and only if it is also a symmetric Nash equilibrium for the function V. This means that Hamilton’s theorem (as applied to siblings) follows as a corollary to Proposition 2.

Corollary--Hamilton’s Theorem. *In Hamilton’s additive model, a necessary condition for a monomorphic population with individuals using strategy* $\bar{x}$ *to be stable against invasion*

---

[8] Game theorists will notice that strategic complementarity (substitutability) is implied by the property of supermodularity (submodularity). If the strategy space is an interval in R^1, these properties are equivalent. (See Paul Milgrom and John Roberts (1990) or Xavier Vives (1990).)

by dominant mutant genes is that $\bar{x}$ is a symmetric Nash equilibrium for the two-player game with payoff function:

$$H(x, \bar{x}) = \Pi(x, \bar{x}) + \frac{1}{2}\Pi(\bar{x}, x).$$

The coincidence of Nash equilibria for V and H that is found in Hamilton's additive model does not extend to payoff functions for general symmetric games. The following examples show that the predicted equilibrium states are quite different under the competing hypotheses that the outcome is a Nash equilibrium with payoffs given by the semi-Kantian rule or by Hamilton's rule.

Example 1---Rousseau's Stag Hunt

Drew Fudenberg and Jean Tirole (1991) use an example called the Stag Hunt to illustrate the way that game theory approaches coordination problems. The name of this game and the story that goes with it is suggested by a passage from Jean Jacques Rousseau's *Discourse on the Origin and Basis of Inequality among Men,* written in 1754.[9] . Two hunters set out to kill a stag. One of them has agreed to drive the stag through the forest, and the other to post at a place where the stag must pass. If each performs his assigned task, they will surely kill the stag and each will get a share of the prey worth $R > 1$. During the course of the hunt, each hunter is tempted to chase a hare which runs by him. If either hunter pursues his hare, he will catch it, but in so doing, he will ensure that the stag is not caught. The value of a hare to the individual who catches it is 1. The payoff matrix for this game is given below, where S denotes the strategy, *persevere in the stag hunt* and where H denotes the strategy, *desert the stag hunt to pursue the hare.*

The Stag Hunt

		Sibling 2	
		S	H
Sibling 1	S	R, R	$0, 1$
	H	$1, 0$	$1, 1$

The coordination problem in this game arises from the fact that there are two Nash equilibria, one of which is better for both players than the other. There is a Nash equilibrium where both

---

[9] The original passage as quoted by Fudenberg and Tirole is *‘‘If a group of hunters set out to take a stag, they are fully aware that they would all have to remain faithfully at their posts in order to succeed; but if a hare happens to pass near one of them, there can be no doubt that he pursued it without qualm, and that once he had caught his prey, he cared very little whether or not he had made his companions miss theirs.’’*

hunters faithfully play their assigned roles and where each gets a payoff of $R > 1$. But there is another Nash equilibrium in which each deserts the hunt and each gets a payoff of 1. Although both players prefer the first of these equilibria, if each believes that the other will desert the stag hunt and pursue a hare, than both will choose to desert. As Fudenberg and Tirole remark, ‘‘without more information about the context of the game and the hunters’ expectations it is difficult to know which outcome to predict.’’

Where the game is played between siblings, the set of possible equilibria is not the same as the set of Nash equilibria for the game with the original selfish payoffs, $\Pi(x, y)$. A monomorphic equilibrium is resistant to invasion by dominant mutants only if the strategy played is a Nash equilibria for the game with payoffs given by the semi-Kantian function, $V(x, y) = \Pi(x, x)/2 + \Pi(x, y)/2$. The payoffs for the function V are displayed in the game matrix below.[10]

Semi-Kantian Payoff Matrix for the Stag Hunt

		Sibling 2	
		S	H
Sibling 1	S	R, R	$R/2, 1$
	H	$1, R/2$	$1, 1$

If $R > 2$, this game has only one Nash equilibrium; the efficient outcome in which both faithfully pursue the stag. It follows from Proposition 1 that for $R > 2$, the only equilibrium that resists invasion by dominant mutants is the efficient outcome for this game, so in this case, kin selection ‘‘solves’’ the coordination problem.

Interestingly, this is a conclusion that would have eluded us if we had attempted to predict the equilibria by studying Nash equilibrium for the game in which payoffs are given by Hamilton’s inclusive fitness function, $H(x, y) = \Pi(x, y) + \frac{1}{2}\Pi(y, x)$ whose payoffs are displayed in the game matrix below. [11]

---

[10] The payoffs $V(x, y)$ are calculated as follows: $V(S, S) = (S, S)/2 + (S, S)/2 = R$, $V(S, H) = (S, S)/2 + (S, H)/2 = R/2 + 0 = R/2$, $V(H, S) = (H, H)/2 + (H, S)/2 = 1/2 + 1/2 = 1$, $V(H, H) = (H, H)/2 + (H, H)/2 = 1/2 + 1/2 = 1$.

[11] The payoffs $H(x, y)$ are calculated as follows: $H(S, S) = (S, S) + (S, S)/2 = 3R/2$, $H(S, H) = (S, H) + (H, S)/2 = 0 + 1/2 = 1/2$,$H(H, S) = (H, S) + (S, H)/2 = 1 + 0 = 1$, $H(H, H) = (H, H) + (H, H)/2 = 3/2$.

The Stag Hunt Under Hamilton's Rule

		Sibling 2	
		S	H
Sibling 1	S	$3R/2, 3R/2$	$1/2, 1$
	H	$1, 1/2$	$3/2, 3/2$

Examination of this matrix reveals that for all $R > 1$, this game has exactly the same equilibria as the Stag Hunt with selfish payoffs; one equilibrium in which both faithfully hunt the stag and one in which each chases the fleeting hare. But when $R > 2$, a population in which siblings abandon the stag hunt to pursue the hare would be invaded by mutants programmed to persevere in hunting the stag.

This example is an instance of the general result from Proposition 2, which informs us that when there is strategic complementarity, every symmetric Nash equilibrium for game with payoff function V is also a symmetric Nash equilibrium for the game with payoff function H. The Stag Hunt exhibits (strict) strategic complementarity, since $\Pi(s, s) + \Pi(h, h) > \Pi(s, h) + \Pi(h, s)$ and as we have seen, for all values of R, the Nash equilibria for the game with payoff V are also Nash equilibria for the game with payoff function H. On the other hand, when $R > 2$, the outcome where each chases a hare is a Nash equilibrium for the game with payoff function H, but not for the game with payoff function V.

Example 2---Prisoners' Dilemma

Consider a two-person prisoners' dilemma game with two possible strategies, *Cooperate* and *Defect*. Payoffs are given by the game matrix:

Prisoners' Dilemma

		Sib 2	
		Cooperate	Defect
Sib 1	Cooperate	R,R	$0, 1$
	Defect	$1, 0$	P,P

For the game to be a prisoners' dilemma, it must be that $0 < P < R < 1$ and $R > 1/2$. In prisoners' dilemma, there is a unique Nash equilibrium; the outcome where both players defect.

But other possibilities for equilibrium emerge if the selfish payoff matrix is replaced either by the payoffs determined according to Hamilton's Rule, or by the payoff matrix determined by the function V. Depending on parameter values, there can be a unique Nash equilibrium where both cooperate, a unique Nash equilibrium where both defect, two Nash equilibria or no Nash equilibrium in pure strategies. As in the case of the Stag Hunt, the set of Nash equilibria under the function H need not be the same as the set of Nash equilibria under the function V.

This prisoners' dilemma game exhibits strategic complementarity if $R + P \geq 1$ and strategic substitutability if $R + P \leq 1$. Only in the special case where $R + P = 1$ does the game take Hamilton's additive form. The payoff matrix for Prisoners' Dilemma when the payoff functions for both players are given by Hamilton's Rule is:

Prisoners' Dilemma under Hamilton's Rule

		Sib 2	
		Cooperate	Defect
Sib 1	Cooperate	$3R/2, 3R/2$	$1/2, 1$
	Defect	$1, 1/2$	3P/2, $3P/2$

The payoff matrix for Prisoners' Dilemma when the payoffs are given by the function $V(x, y) = \frac{1}{2}\Pi(x, x) + \frac{1}{2}\Pi(x, y)$ is:

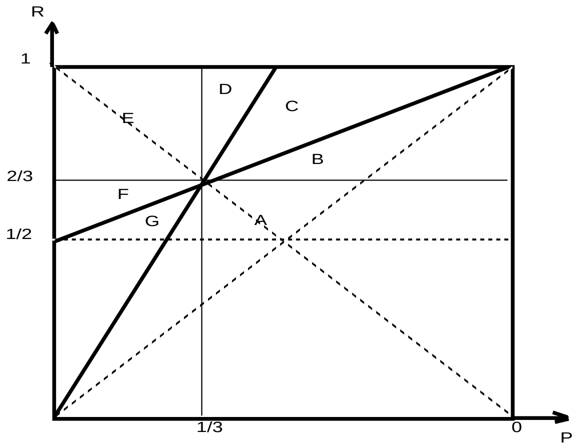
Prisoners' Dilemma for Diploid Siblings

		Sib 2	
		Cooperate	Defect
Sib 1	Cooperate	R, R	$R/2, (P + 1)/2$
	Defect	$(P + 1)/2, R/2$	P, P

In the game with payoffs determined by Hamilton's Rule, there is a Nash equilibrium in which both siblings cooperate if and only if $R \geq 2/3$, and there will be a Nash equilibrium in which both defect if and only if $P \geq 1/3$. In the game with payoffs given by the function V, there will be a Nash equilibrium in which both players cooperate if $R \geq P/2 + 1/2$ and there will be a Nash equilibrium in which both players defect if $R \geq 2P$.

As in the example of the Stag Hunt, there are parameter values for which the Hamiltonian function H and the semi-Kantian function V predict different outcomes. But here the possibilities are more elaborate. Figure 1 demarcates the possibilities for alternative parameter values. All points that are above the dashed horizontal line and above the upward-sloping diagonal represent parameter values for which the game with the selfish payoff function Π is a prisoners' dilemma. For parameter values in the region A, with either of the payoff functions H and V, there is a unique Nash equilibrium in which both siblings defect. In the game with payoff function H, for all parameter values in regions B, C, and D, there are two Nash equilibria--one in which both siblings cooperate and one in which they both defect. But with the payoff function V, for parameter values in region B there is only one equilibrium--the outcome where both siblings defect. With payoff function V and parameter values in Region C, there are two equilibria--one where both siblings defect, and one where both siblings cooperate. With payoff function V, in region D, there is a unique equilibrium, in which both siblings cooperate. In region E, for either of the payoff functions H and V, the only Nash equilibrium has both siblings cooperating.

Figure 1. Equilibrium Regimes for Prisoners' Dilemma



In the original prisoners' dilemma game there is strategic complementarity when $R + P \geq 1$. This is the parameter region above the downward-sloping diagonal of the box in Figure 1. As Proposition 2 predicts, in this region every Nash equilibrium for the payoff function V is a Nash equilibrium for the payoff function H. But the converse is not true. In the interior of Region B, the strategy *Defect* is a Nash equilibrium for H but not for V and in Region D, the strategy *Cooperate* is a Nash equilibrium for H but not for V.

In the original prisoners' dilemma game there is strategic substitutability when $R + P \leq 1$. This is the parameter region below the downward-sloping diagonal of the box in Figure 1. When the payoff function is H; in Region G, there is no Nash equilibrium in pure strategies, in region F there is a unique Nash equilibrium where both siblings cooperate, and in region H there is a unique Nash equilibrium where both siblings defect. These facts are in accordance with Proposition 2, since every Nash equilibrium for H is a Nash equilibrium for V. The example shows that the converse is not true, since in regions F and H, there is a Nash equilibrium for payoff function V that is not a Nash equilibrium for payoff function H.

4. When Mutants Mate with Mutants--Second-Order Invasion

Where $\bar{x}$ is a symmetric Nash equilibrium strategy for a game with the semi-Kantian payoff function V, a population of aa genotypes who take action $\bar{x}$ can still be invaded by a mutant gene if that gene gets the *same* expected payoff as normal genes when it appears in the offspring of an Aa and an aa genotype, but gets a *higher* expected payoff than normal genes if its parents carry two or more mutant genes. Such a mutant gene is said to stage a *second-order invasion* of the initial population.

Populations that Resist Invasion by Recessive Mutants

Invasion by recessive mutant genes is an especially interesting example of second-order invasion. Recessive genes have no effect on the behavior of an individual who only carries only one such gene. Therefore if A is a recessive mutant gene, when an individual of genotype Aa mates with a normal aa genotype, their children will be either of type Aa or of type aa. Since Aa genotypes behave in the same way as aa genotypes, all offspring of this match take the same actions, get the same payoffs, and reproduce at the same rate as normal offspring of normal individuals. Similarly, if an AA genotype mates with an aa genotype, all of their offspring will be of genotype Aa and hence will take the same actions and get the same payoffs as normal aa genotypes. Only if both parents have at least one A gene will there be some AA genotypes among the offspring and only then can the payoffs to the recessive mutant gene differ from the payoffs to normal genes.

The task of analyzing recessive invasion is simplified by the fact that when mating is random and when mutant A genes are rare, almost all of the A genes in the adult population are carried by Aa genotypes rather than by AA genotypes. Formally:

Lemma 1. *Suppose that in the adult population the proportion of mutant A genes is ϵ and the proportion of normal a genes is $1-\epsilon$. If mating is random and if the probability that the offspring of normal parents survive to reproduce is at least $\bar{p} > 0$, then in the next generation among those who survive to adulthood, the ratio of the number of AA genotypes to the number of Aa genotypes approaches zero as ϵ approaches zero.*

If mating is random, it follows from Lemma 1 that as the proportion of A genes in the population approaches zero, almost all AA genotypes are born to families in which both parents are of genotype Aa. Therefore a necessary condition for a population of aa genotypes to be resistant to invasion by recessive mutants is that the expected rate of reproduction of A genes in families where both parents are of genotype Aa is no greater than the expected rate of reproduction of normal genes in the population.

In a population of aa genotypes that use strategy $\bar{x}$, let A be a recessive mutant gene such that double recessive AA genotypes use strategy x. The expected payoff to A genes born to two parents of genotype Aa turns out to be a linear combination of the payoff that an x-strategist would get if its sibling also took action x, the payoff that it would get if its sibling took action $\bar{x}$, and the payoff that it's sibling would get if the sibling took action $\bar{x}$. This last term appears because since A is recessive, it will occasionally happen that a sibling of an AA genotype who takes the normal action $\bar{x}$ nevertheless carries a single copy of the mutant gene A. The combinatorics are worked out explicitly in the Appendix, yielding the following explicit result.

Proposition 3. *A necessary condition for a monomorphic population that uses strategy $\bar{x}$ to be stable against invasion by recessive mutant genes is that $\bar{x}$ be a symmetric Nash equilibrium for the two-player game with payoff function:*

$$W(x,\bar{x}) = \frac{1}{5}\Pi(x,x) + \frac{3}{5}\Pi(x,\bar{x}) + \frac{1}{5}\Pi(\bar{x},x).$$

Second-Order Invasion by Dominant Mutants

It is also possible that *dominant* mutant genes might stage a second-order invasion of a population of individuals of genotype aa who play a symmetric Nash equilibrium strategy $\bar{x}$. Consider a dominant mutant gene A such that all Aa and AA genotypes take action x where $V(x,\bar{x}) = V(\bar{x},\bar{x}) = \Pi(\bar{x},\bar{x})$. Since $V(x,\bar{x}) = V(\bar{x},\bar{x})$, individuals with the mutant gene have the same survival probability as normals when they have one parent of genotype Aa and one of genotype aa. Mutant A genes will therefore reproduce faster than normal genes if in families where the parents have a total of two or more A genes, the expected payoff to carriers of the A gene exceeds that of normal genes. In such families, dominant mutant genes that make an individual take action x will do better than normal genes if $V(x,x) > V(\bar{x},\bar{x})$. A formal statement of this result is the following.

Proposition 4. *A necessary condition for stability against invasion by dominant mutant genes in a monomorphic population where all individuals use strategy $\bar{x}$ is that if $V(x,\bar{x}) = V(\bar{x},\bar{x})$, then $V(x,x) \le V(\bar{x},\bar{x})$.*[12]

5. On the Existence of Stable Monomorphic Equilibrium

Three Related Payoff Functions

Our discussion has produced three payoff functions that are candidates for measuring evolutionary ‘‘fitness’’ of strategies to be played in games between siblings. These functions are:

- (The Semi-Kantian Payoff)

$$V(x,\bar{x}) = \frac{1}{2}\Pi(x,x) + \frac{1}{2}\Pi(x,\bar{x}).$$

---

[12] Readers familiar with Maynard Smith's definition of ESS may be surprised to see that this condition differs from Maynard Smith's ‘‘second-order condition’’ for ESS. As we will explain later, Maynard Smith's condition is appropriate in the case where creatures encounter randomly chosen strangers, but not for the case of games between siblings.

- (Hamilton's Inclusive Fitness)

$$H(x,\bar{x}) = \Pi(x,\bar{x}) + \frac{1}{2}\Pi(\bar{x},x).$$

- (Fitness Against Recessives)

$$W(x,\bar{x}) = \frac{1}{5}\Pi(x,x) + \frac{3}{5}\Pi(x,\bar{x}) + \frac{1}{5}\Pi(\bar{x},x).$$

We have argued that in order for a monomorphic population to resist invasion both by dominant mutants and by recessive mutants, this population must use a strategy $\bar{x}$ that is a symmetric Nash equilibrium for *both* the semi-Kantian payoff function and for Hamilton's inclusive fitness function.

Two questions come to mind:

(I) When, if ever, does there exist a strategy $\bar{x}$ that is *simultaneously* a symmetric Nash equilibrium for the payoff functions V and W?

(II) Is Hamilton's inclusive fitness function related in an interesting way to equilibrium for either or both of the other two payoff functions?

Answers to these questions emerge as consequences of the fact that the function W is a positive linear combination of the functions V and H, and the fact that when an individual and its sibling are choosing the same action, the three functions V, W, and H all have the same gradients with respect to an individuals own actions. These results are reported as Lemma 2, which follows immediately from the definitions, and Lemma 3 which is proved in the Appendix.

Lemma 2. *For all $x \in S$ and $y \in S$:*

$$W(x,y) = \frac{2}{5}H(x,y) + \frac{2}{5}V(x,y)$$

Lemma 3. *If the function $\Pi(x,y)$ is differentiable, and if $V_1(x,y)$, $W_1(x,y)$, and $H_1(x,y)$ denote respectively the gradients with respect to x of the functions $V(x,y)$, $W(x,y)$, and $H(x,y)$, then for all $x \in S$, $V_1(x,x) = H_1(x,x)$ and $W_1(x,x)$ is proportional to $V_1(x,x)$ and $H_1(x,x)$.*

From Lemma 2 it follows that if a strategy $\bar{x}$ is a symmetric Nash equilibrium for the semi-Kantian function V and also a symmetric Nash equilibrium for the inclusive fitness function H, then it will be also be a symmetric Nash equilibrium for W. As will be spelled out in later in this paper, Lemma 2, used together with Proposition 2, will allow us to show that when there is complementarity or substitutability, conditions that guarantee the existence of a symmetric Nash equilibrium for one of the two functions V and W will also guarantees the existence of a symmetric Nash equilibrium for the other.

If payoff functions are differentiable, then a necessary condition for a strategy $\bar{x}$ to be an *interior* Nash equilibrium for both of the functions H and V is that the two gradients $V_1(\bar{x},\bar{x})$ and $H_1(\bar{x},\bar{x})$, both be zero. Since H and V are quite different functions, it is reasonable to ask whether they can ever have equal gradients at the same point. According to Lemma 3, the answer is that when $y = x$, the gradients of $H(x,y)$ and $V(x,y)$ with respect to x are equal. Therefore the calculus first-order conditions for a Nash equilibrium for either of the payoff functions H and V will be satisfied whenever the first order conditions for the other are satisfied.

Some Useful Facts from Game Theory

There are simple, two person symmetric games that have no symmetric Nash equilibrium in pure strategies. A famous example of such a game is Maynard Smith's *Hawk-Dove game*, Maynard Smith (1982).[13] Therefore it is important to know that there exists a symmetric Nash equilibrium for a large, readily-identifiable class of symmetric two person games. Moreover, for a symmetric game that has a finite number of possible pure strategies, there will always exist a symmetric Nash equilibrium in mixed strategies. Both of these results can be found in any standard reference on game theory. (e.g., Rasmusen (1989, pp 123-127). Formal statements of these are designated as Facts 1 and 2.

Fact 1. *There exists a symmetric equilibrium in pure strategies for any symmetric game in which S is a compact, convex set and where the payoff functions are continuous functions, concave in each player's own strategy*[14]

Fact 2. *There exists a symmetric equilibrium in mixed strategies for any symmetric game in which there is a finite number of pure strategies.*

Equilibrium in Pure Strategies

We now have sufficient apparatus to prove the existence of strategies that are simultaneously symmetric Nash equilibria for the games with payoff functions V and H. The first of these propositions is ''calculus free'' in the sense that it depends on concavity of functions but not on differentiability. It is an immediate consequence of Fact 1 and of the results on strategic complements and substitutes found in Proposition 2.

Proposition 5. *If the set S is a closed, bounded, convex subset of R^n. (i) If the payoff function Π displays strategic complementarity and if $V(x,y)$ is a concave function of x, then there exists a strategy $\bar{x}$ that is a symmetric Nash equilibrium for all three functions H, V, and W. (ii) If the payoff function Π displays strategic substitutability and if $H(x,y)$ is a concave function of x, then there exists a strategy $\bar{x}$ that is a symmetric Nash equilibrium for all three functions H, V, and W.*

For differentiable functions, the following is implied by the equality of the gradients of H and V (Lemma 3).(See the Appendix for a proof.)

Proposition 6. *If $H(x,y)$, and $V(x,y)$ are both differentiable concave functions of x for fixed y, then there exists an action $\bar{x} \in S$ such that $\bar{x}$ is a symmetric Nash equilibrium for each of the payoff functions, H, V, and W. Moreover, if $\bar{x}$ is a symmetric Nash equilibrium*

---

[13] Another example, familiar to poultry lovers and game theorists, is the game of *chicken*.

[14] A payoff function (x, y) is defined to be concave in a player's own strategy if for all x', x'', and y in S, and for all $\lambda \in [0,1]$, $(\lambda x' + (1-\lambda)x'', y) \geq \lambda(x', y) + (1-\lambda)(x'', y)$.

for any one of these three functions, it is also a symmetric Nash equilibrium for the other two.

The next corollary is found by articulating the hypothesis of Proposition 6 in terms of the Jacobean matrices of the payoff function $\Pi(x, y)$. (A proof is in the Appendix.)

Corollary. *Let $\Pi_{11}(x, y)$ be the Jacobean matrix of $\Pi(x, y)$ with respect to x, holding y constant, let $\Pi_{22}(x, y)$ be the Jacobean matrix of $\Pi(x, y)$ with respect to y, holding x constant, and let $\Pi_{12}(x, y)$ be the matrix whose ijth element is the $\frac{\partial \Pi(x,y)}{\partial x_i \partial y_j}$. If these three matrices are all negative semi-definite,*[15] *then there exists $\bar{x}$ such that $\bar{x}$ is simultaneously a symmetric Nash equilibrium for H, V, and W.*

Equilibrium in Mixed Strategies

If there is a finite number n of pure strategies and if randomized strategies are allowed, then the set S of possible strategies is an $n - 1$ dimensional simplex, where an element of S is a vector whose ith component is the probability weight that the player uses the ith pure strategy. Let A be the $n \times n$ matrix whose ijth element is the payoff to an individual who chooses strategy i and whose sibling chooses strategy j. If an individual uses strategy mixed strategy x and its sibling uses mixed strategy y, the expected payoff to the individual who played x is given by $\Pi(x, y) = x^t A y = \sum_i \sum_j A_{ij} x_i y_j$. This function is linear in x for fixed y (and linear in y for fixed x.) Then $H(x, y) = x^t A y + y^t A x/2 = x^t(A + A^t/2)y$ is also linear in x for fixed y and linear in y for fixed x. The function V, on the other hand is quadratic rather than linearin x, since $V(x, y) = x^t A x/2 + x^t A y/2$. Accordingly, the function V will be concave in A if and only if the matrix A is negative semi-definite.

Since H is guaranteed to be a concave function on S, application of Proposition 5 gives us a strong result on the existence of equilibria that resist both .

Proposition 7. *If there are a finite number of pure strategies and if Π displays strategic substitutability, then there exists a strategy $\bar{x}$ that is a symmetric Nash equilibrium for each of the payoff functions V, H, and W.*

6. Sex, Games, and Biologists

The ideas in this paper are offspring of William Hamilton's theory of kin-selection and John Maynard Smith's work on applications of game theory to evolution. So far as I know, its closest relatives sharing descent from early work by Hamilton and Maynard Smith are discussions of games played between relatives by A. Grafen (1979) and Maynard Smith and W. G. S. Hines (1979). The following discussion attempts to relate the current results to this earlier work.

---

[15] The $n \times n$ matrix A is defined to be *negative semi-definite* if $x^t A x \leq 0$ for all n-vectors x.

Games and Evolutionary Stable Strategies

Maynard Smith and G. R. Price (1973) and Maynard Smith (1974) introduced explicit game theoretic models of social interaction to the theory of evolution. These papers developed the notion of an ‘‘evolutionary stable strategy’’ (ESS) which characterizes evolutionary equilibrium in a population of individuals who play symmetric two-person games with randomly selected members of their own species. While these early papers are not explicit about the asexual nature of reproduction, in a later paper Maynard Smith (1978, p. 37, p. 51), remarked that ESS was formulated under the *implicit* assumption that inheritance is asexual and that surviving offspring all play the same strategy as their parent.

If the payoff to a player who plays x with an opponent who plays y is given by $\Pi(x, y)$, then an ESS is a strategy $\bar{x}$ with the following two properties:

(a) *First-order condition for ESS*

$\Pi(x, \bar{x}) \leq \Phi(\bar{x}, \bar{x})$ for all $x \in S$.

(b) *Second order condition for ESS* If $\Pi(x, \bar{x}) = \Pi(\bar{x}, \bar{x})$, then $\Pi(x, x) < \Pi(\bar{x}, x)$.

The first order condition requires that if a strategy $\bar{x}$ is an ESS, then no other strategy x gets a higher payoff than $\bar{x}$ when played against x; that is to say, $\bar{x}$ is a symmetric Nash equilibrium strategy. The second order condition must be satisfied because should an invading strategy x get the *same* payoff against the equilibrium strategy $\bar{x}$, then the eventual success or failure of the invader will be determined by whether the invader does better or less well than the normal strategy when it encounters another invader.

Maynard Smith establishes his second order condition with an argument that is similar to that used in this paper to establish Proposition 4. Despite the similarity of approach, it is interesting that the necessary condition in Proposition 4 is different from the second order condition of an ESS. The ESS second order condition is that the mutant strategy gets a lower payoff when it plays itself than does the equilibrium strategy when it plays a mutant. Proposition 4 has it that a necessary condition for stability is that a mutant will get a lower payoff when it plays itself than the equilibrium strategy will get when it plays itself. The difference between these two second order conditions arises from the fact that in our model individuals are playing against siblings rather than against randomly chosen members of the population.

Kin Selection and ESS

In a later paper, Maynard Smith (1978) proposed that Hamilton’s theory of kin selection could be extended to games with general payoff functions by studying ESS for the game in which the payoff function is Hamilton’s inclusive fitness function. Grafen (1979) criticized this approach, arguing that Hamilton’s inclusive fitness does not account for the fact that ‘‘*an individual is more likely to play against his own strategy than he would if he played the population at random.*’’ Grafen proposed therefore that in games between siblings, evolutionary forces will select for individuals that maximize a function equivalent to the function V discussed in this paper. Grafen used the example of a two-strategy Hawk-Dove game to show that the set of Nash equilibria under the payoff function V is not the same as the set of Nash equilibria for the payoff function H. Subsequently, Maynard Smith and Hines (1979) accepted Grafen’s argument and named the function V the ‘‘personal fitness function’’. Maynard Smith and Hines explain the difference between personal fitness and Hamilton’s inclusive fitness as follows:

> ‘‘ *Personal fitness modifies classical fitness by allowing for the effects that an individual’s relatives will have on the number of his own offspring to survive, whereas inclusive fitness modifies classical fitness by allowing for the effects that an individual will have on the number of his relative’s offspring that survive.*’’ *(1979, p 20.)*

Mendelian Genetics and Sexual Diploids

Like Maynard Smith, Grafen avoided an explicit model of the genetics of a sexual diploid species. Indeed, Maynard Smith and Hines interpret the population in the Grafen model as

> ‘‘ *a set of parthenogenetic clones, each adopting a particular strategy. . . The pattern of dispersal is such that with probability r an individual plays against a member of its own clone and with probability $1 - r$ against a randomly chosen opponent.*’’ *(1979, p. 21)*

A major difference between this paper and the papers by Grafen and by Smith and Hines, is the use of an explicit model of sexual diploid reproduction. A consequence of this difference is that the state variable for our dynamical system is the distribution of *genes* rather than the distribution of *strategies* in the population. If reproduction is asexual, it is sufficient to describe the population at any time by the distribution of strategies. But with sexual diploid reproduction, there is not in general a one-to-one correspondence between distributions of genes in the population and the distribution of strategies taken by individuals. For example, in the case where there are behavior is controlled by dominant and recessive genes, more than one genotype can dictate the same strategy. Even if there is a one-to-one mapping between genotypes and strategies, strategies do not necessarily beget their own kind. For example, among the offspring of two heterozygote parents, there can be three distinct genotypes.

In the introduction to *Evolution and the Theory of Games* (1982), Maynard Smith defends his use of ‘‘a game theory model when a classical population genetics model more precisely represents biological reality’’ on the grounds that ‘‘the basic assumption of evolutionary game theory--that like begets like--corresponds to what we actually know about heredity in most cases.’’ Maynard Smith also acknowledges that there are ‘‘contests in which population genetic models become necessary’’ and later in the book, he briefly explores models that combine game theoretic payoffs and the population genetics of sexual diploidy. It is hoped that this paper has demonstrated that behavior towards relatives belongs to the class of contests in which population models are necessary and, more generally, that a marriage of the methods of game theory and population genetics will engender a rich supply of interesting theorems and insights.

7. Things Left to Be Done

Several of the restrictive assumptions that have been made here could fruitfully be relaxed. It would be interesting to examine various patterns of nonrandom mating. Of particular interest would be a model with more attention to inbreeding. It seems likely that the environments which shaped our genes consisted of small, rather isolated populations in which the amount of inbreeding was significant. If relatives mate fairly frequently, there can be a high probability that carriers of mutant genes mate with other carriers of mutant genes, even if the mutant gene would be extinguished if mutants always mated with normal individuals. It would be very interesting to develop the Mendelian combinatorics of equilibrium in such an environment.[16]

Relaxing the assumption of monogamous mating also raises some fascinating issues. If females mate with more than one male, individuals who share the same mother may have different fathers. Therefore if an individual inherits a rare mutant gene from its father, the probability that this gene is shared by the mother's other children is less than 1/2. But if the mutant gene is inherited from the mother, the probability that the mother's other siblings have the same gene is 1/2. If genes cannot determine whether they are inherited paternally or maternally, then the results of this paper can be applied with appropriate adjustment in the probabilities. However, biologists have established that some genes are able to act differently, depending on whether they are maternally or paternally inherited. This phenomenon, known as genomic imprinting is discussed by Haig (1992). If mating is promiscuous and genomic imprinting is present, then the theory has to be fairly drastically expanded to incorporate the conflict of interest between maternally inherited and paternally inherited genes residing in the same individual.

Although this paper has dealt explicitly with genetic inheritance, a strikingly similar formalism applies to ‘‘cultural evolution’’, in which individual beliefs, skills, and knowledge are learned or copied from others who may or may not be one's biological parents. Although the relation between individuals and their ‘‘cultural parents’’ may not be as highly structured as genetic relationships, there are still likely to be patterns of inheritance that lend themselves well to analysis. Marcus Feldman and Luigi Cavelli-Sforza (1981) and Robert Boyd and Peter Richerson (1985) have laid promising foundations for this endeavor. Bergstrom and Oded Stark (1993) also outline some simple models of cultural inheritance. In one of these models, an individual may choose a strategy which is a copy of that used by a randomly selected parent, or with some probability he may ignore his parents and copy a randomly selected outside individual. They also present a model in which individuals are located along a road and play prisoners' dilemma with their neighbors. Over time, these individuals are replaced by their offspring, who choose the strategy used by the most successful individual in some contiguous geographic region that is visible to them. Despite the great diversity of possible models of cultural inheritance, such models have much in common with the special genetic model studied here. Typically, an individual who ‘‘inherits’’ a particular behavior is more likely than a randomly selected member of the population to interact with others whose behavior is copied from the same source. When this is the case, evolution does not come to rest at a Nash equilibrium where all individuals take actions that maximize their selfish payoffs. Instead, selection takes account of the benefits and costs of interaction with ‘‘cultural relatives’’ who share the same role models. While this effect is simple and dramatic for sibling interaction, similar forces will be present where there is cultural transmission within extended families and other social networks.

---

[16] The genetic consequences of various patterns of inbreeding have been extensively studied because of practical applications for the breeding of livestock. A useful reference to models of nonrandom mating is Karlin (1969).

Appendix

Proof of Proposition 2:

From equations 3 and 4 it follows that for all $x \in S$:

$$(H(x,\bar{x}) - H(\bar{x},\bar{x})) - (V(x,\bar{x}) - V(\bar{x},\bar{x})) = \frac{1}{2}\left[(\Pi(x,\bar{x}) + \Pi(\bar{x},x)) - (\Pi(\bar{x},\bar{x}) - \Pi(x,x))\right]. \tag{1}$$

The right side of this equation is nonpositive for all x if Π exhibits strategic complementarity and nonnegative if Π exhibits strategic substitutability. Therefore if Π exhibits strategic complementarity,

$$H(x,\bar{x}) - H(\bar{x},\bar{x}) \leq V(x,\bar{x}) - V(\bar{x},\bar{x}) \tag{2}$$

for all $x \in S$ and if Π exhibits strategic substitutability,

$$V(x,\bar{x}) - V(\bar{x},\bar{x}) \leq H(x,\bar{x}) - H(\bar{x},\bar{x}) \tag{3}$$

for all $x \in S$. If $\bar{x}$ is a symmetric Nash equilibrium for the game with payoff function V, then $V(x,\bar{x}) - V(\bar{x},\bar{x}) \leq 0$ for all $x \in S$. Then if Π exhibits strategic complementarity, it follows from (2) that $H(x,\bar{x}) - H(\bar{x},\bar{x}) \leq 0$ for all $x \in S$ and hence $\bar{x}$ is a symmetric Nash equilibrium for the game with payoff function H. If $\bar{x}$ is a symmetric Nash equilibrium for the game with payoff function H, then $H(x,\bar{x}) - H(\bar{x},\bar{x}) \leq 0$ for all $x \in S$. Then if Π exhibits strategic substitutability, it follows from (3) that $V(x,\bar{x}) - V(\bar{x},\bar{x}) \leq 0$ and hence $\bar{x}$ is a symmetric Nash equilibrium for the game with payoff function V.

Proof of Lemma 1:

If the proportions of A genes and a genes in the adult population are respectively ϵ and $(1-\epsilon)$, a simple combinatorial argument shows tht the expected proportions of genotypes aa, Aa, and AA found in the offspring born to this generation are respectively, $(1-\epsilon)^2$, $2\epsilon(1-\epsilon)$, and ϵ^2. (These proportions are known to geneticists as the *Hardy-Weinberg ratios.*) At birth, therefore, the ratio of the number of AA genotypes to the number of Aa genotypes is $\frac{\epsilon}{2(1-\epsilon)}$. The ratio of individuals of the two genotypes that survive to adulthood will therefore be $\frac{\epsilon}{2(1-\epsilon)}\rho$ where ρ is the ratio of the survival probability of AA genotypes to that of Aa genotypes. The survival probability of AA genotypes can be no larger than 1. The survival probability of an Aa genotype will be $\bar{p}$ if its siblings are either of type aa or of type Aa, since in this case both siblings take the same actions as normal individuals and therefore have the same survival probability. As ϵ approaches zero, the probability an individual's siblings are all of type AA or of type Aa can be shown to approach 1. Therefore as ϵ approaches zero, ρ approaches the positive number $\frac{1}{\bar{p}}$. Then the ratio of AA genotypes to Aa genotypes in the surviving adult population is $\frac{\epsilon}{2\bar{p}(1-\epsilon)}$, which approaches zero as ϵ approaches zero.

Proof of Proposition 3:

When the A gene is recessive, an individual carrying mutant A genes get a payoff that differs from that received by normal aa genotypes only if this individual or its sibling is of genotype AA. There are three of these critical types of sibling pairs in which at least one sibling is AA. These are

- Critical type I. Both siblings are of genotype AA.
- Critical type II. One sibling is of genotype AA and the other is of genotype aa.
- Critical type III. One sibling is of genotype AA and the other is of genotype Aa.

Where both parents are heterozygotes, the probability that any pair of their offspring is of critical type I is 1/16, the probability that the pair is of critical type II is 1/8, and the probability that the pair is of critical type III is 1/4. In each sibling pair of critical type I, a total of four A genes are found. In each sibling pair of critical type II, two A genes are found, and in each sibling pair of critical type III, three A genes are found. Of those A genes that are found in a critical type of sibling pair, the fraction located in critical type I sibling pairs is

$$\frac{4 \times 1/16}{4 \times 1/16 + 2 \times 1/8 + 3 \times 1/4} = 1/5,$$

and the fraction located in critical type II sibling pairs is

$$\frac{2 \times 1/8}{4 \times 1/16 + 2 \times 1/8 + 3 \times 1/4} = 1/5.$$

The remaining 3/5 of critical type pairs will be of critical type III.

In critical type I sibling pairs, both siblings take action x, so each A gene gets a payoff of $\Pi(x,x)$. In type II sibling pairs, each of these A genes is found in an AA genotype who chooses action x while its sibling chooses action $\bar{x}$. Hence each A gene in a type II sibling pair gets a payoff of $\Pi(x,\bar{x})$. In type III sibling pairs, two thirds of the type A genes are carried by AA genotypes who take action x, while their siblings, who are Aa genotypes, take action $\bar{x}$. These A genes get payoffs of $\Pi(x,\bar{x})$. One third of the A genes found in critical pairs of type III are carried by Aa genotypes whose siblings are AA genotypes. These A genes are carried by individuals who take action $\bar{x}$, while their siblings take action x. Accordingly they get payoffs of $\Pi(\bar{x},x)$. Putting these facts together, we find that in the offspring of two parents of genotype Aa, the expected payoff experienced by Type A genes conditional on it being found in a sibling pair of one of the critical types is:

$$W(x,\bar{x}) = \frac{1}{5}\Pi(x,x) + \frac{3}{5}\Pi(x,\bar{x}) + \frac{1}{5}\Pi(\bar{x},x).$$

Since offspring who do not belong to a critical type of sibling pair get the same payoff as normal offspring, it follows that among the offspring of two Aa parents, the expected payoff to type A genes will be greater or smaller than the expected payoff to normal individuals depending on whether $W(x,\bar{x})$ is greater or smaller than $W(\bar{x},\bar{x})$.

The only types of parent pairs that can produce offspring of genotype AA are matches between two parents of genotype Aa or matches in which one parent is of genotype AA and the other is either of genotype Aa or of genotype AA. According to Lemmma 1, the ratio of adults of type AA to adults of type Aa approaches zero as the fraction of mutant genes in the population approaches zero. Therefore when the mutant gene is rare, almost all families that produce offspring of genotype AA will consist of two parents of genotype Aa. It follows that a necessary condition for a population of $\bar{x}$-strategists to resist invasion by recessive mutants is that $W(\bar{x},\bar{x}) \geq W(x,\bar{x})$ for all $x \in S$.

Proof of Proposition 4:

In each of the possible cases where the parents have a total of two or more A genes, the probability that any offspring has at least one A gene exceeds 1/2. Therefore the expected payoff to a mutant offspring of parents who carry two or more A genes is $\lambda\Pi(x,x) + (1-\lambda)\Pi(x,\bar{x})$ where $\lambda > 1/2$. Since, by assumption $\Pi(x,x)/2 + \Pi(x,\bar{x})/2 = \Pi(\bar{x},\bar{x})$, it follows that $\Pi(x,x) < \Pi(\bar{x},\bar{x})$. This is the desired second-order necessary condition for stable monomorphic equilibrium.

Proof of Lemma 3:

Since $V(x,y) = \frac{1}{2}\Pi(x,x) + \frac{1}{2}\Pi(x,y)$, it follows that $V_1(x,y) = \frac{1}{2}\Pi_1(x,x) + \frac{1}{2}\Pi_2(x,x) + \frac{1}{2}\Pi_1(x,y)$. Therefore $V_1(x,x) = \Pi_1(x,x) + \frac{1}{2}\Pi_2(x,x)$. Since $H(x,y) = \Pi(x,y) + \frac{1}{2}\Pi(y,x)$, $H_1(x,y) = \Pi_1(x,y) + \frac{1}{2}\Pi_2(y,x)$ and therefore $H_1(x,x) = \Pi_1(x,x) + \frac{1}{2}\Pi_2(x,x) = V_1(x,x)$. Differentiating both sides of the equality in Lemma 2, one finds that $W_1(x,y) = \frac{2}{5}H_1(x,y) + \frac{2}{5}V_1(x,y) = \frac{4}{5}H_1(x,y)$.

Proof of Proposition 6:

From Fact 1 it follows that there exists at least one symmetric Nash equilibrium for each of these functions. The remaining task is to show that the Nash equilibria for the three functions coincide. From Lemma 3 we see that the gradients of $V(x,y)$, $H(x,y)$, and $W(x,y)$ with respect to x are all proportional to each other when $x = y$. At a Nash equilibrium for any one of these functions, the gradient of the function with respect to x will be zero. Therefore the gradients for the other two functions will also be zero. For a concave function, a zero gradient is a sufficient as well as a necessary condition for a global maximum. If $V(x,y)$ and $H(x,y)$ are both concave in X, then since the function W is a positive linear combination of V and H, $W(x,y)$ will also be concave in X. Therefore a Nash equilibrium for any of these three functions, $\bar{x}$ satisfies sufficiency conditions for Nash equilibrium for the other two.

Proof of Corollary to Proposition 6:

The Jacobean matrices for $H(x,y)$, and $V(x,y)$ with respect to x are respectively:

$$\begin{aligned} H_{11}(x,y) &= \Pi_{11}(x,y) + \frac{1}{2}\Pi_{22}(y,x) \\ V_{11}(x,y) &= \frac{1}{2}\Pi_{11}(x,y) + \Pi_{12}(x,x) + \frac{1}{2}\Pi_{22}(y,y) + \frac{1}{2}\Pi_{11}(x,y). \end{aligned}$$

The functions $H(x,y)$, $V(x,y)$, and $W(x,y)$ are concave in x if and only if the matrices $H_{11}(x,y)$, $V_{11}(x,y) \leq 0$, and $W_{11}(x,y)$ are negative semi-definite. This will be the case if $\Pi_{11}(x,y)$, $\Pi_{12}(x,y)$, and $\Pi_{22}(x,y)$.

References

Becker, Gary (1976) ‘‘Altruism, Egoism, and Fitness: Economics and Sociobiology,’’ *Journal of Economic Literature*, **14**, .

Bergstrom, Carl and Bergstrom, Theodore (1993) ‘‘The Evolution of Altruism Among Siblings who Play Games Like Prisoners’ Dilemma,’’ working paper.

Bergstrom, Theodore, and Stark, Oded (1993) ‘‘How Altruism Can Prevail in Evolutionary Environments,’’ *American Economic Review*, **83**, 149-155.

Binmore, Ken (1992) *Fun and Games*. Lexington, Ma: D. C. Heath.

Binmore, Ken and Samuelson, Larry (1992) ‘‘Evolutionary Stability in Repeated Games Played by Finite Automata,’’ *Journal of Economic Theory*, **57**, 278-305.

Boyd, Robert and Richerson, Peter (1985) *Culture and the Evolutionary Process*. Chicago: Univesity of Chicago Press.

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

Dawkins, R. (1976) *The Selfish Gene*. New York: Oxford University Press.

Foster, Dean and Young, H. Peyton (1990) ‘‘Stochastic Evolutionary Game Dynamics,’’ *Theoretical Population Biology*, **38**, 219-232.

Frank, R. (1988) *Passions within Reason*. New York: Norton.

Friedman, Daniel (1991) ‘‘Evolutionary Games in Economics,’’ *Econometrica*, **59**, 637-666.

Fudenberg, D., and Tirole, J. (1991) *Game Theory*. Cambridge, Ma: MIT Press.

Haig, David (1992) ‘‘Genomic Imprinting and the Theory of Parent-Offspring Conflict,’’ *Developmental Biology*, **3**, 153-160.

Hamilton, W. D. (1964a) ‘‘The Genetical Evolution of Social Behavior. I,’’ *Journal of Theoretical Biology*, **7**, 1-16.

Hamilton, W. D. (1964b) ‘‘The Genetical Evolution of Social Behavior. II.,’’ *Journal of Theoretical Biology*, **7**, 17-52.

Hansson, Ingemar and Stuart, Charles (1990) ‘‘Malthusian Selection of Preferences,’’ *American Economic Review*, **80**, 529-544.

Hines, W. G. S. and Maynard Smith, J. (1979) ‘‘Games Between Relatives,’’ *Journal of Theoretical Biology*, **79**, 19-30.

Hirshleifer, J. (1977) ‘‘Economics from a Biological Viewpoint,’’ *Journal of Law and Economics*, **20**, 1-52.

Hirshleifer, J. (1978) ‘‘Natural Economy Versus Political Economy,’’ *Journal of Social and Biological Structures*, **1**, 319-337.

Kandori, Michihiro, Mailath, George, Rob, Rafael (1993) ‘‘Learning, Mutation, and Long Run Equilibrium in Games,’’ *Econometrica*, **61**, 85-122.

Karlin, S. (1969) *Equilibrium Behavior of Genetic Models with Nonrandom Mating*. London: Gordon and Breach.

Maynard Smith, J. (1974) ‘‘The Theory of Games and the Evolution of Animal Conflicts,’’ *Journal of Theoretical Biology*, **47**, 209-221.

Maynard Smith, J. (1978) ‘‘Optimization Theory in Evolution,’’ *Annual Review of Ecology and Systematics*, **9**, 31-56.

Maynard Smith, J. (1982) *Evolution and the Theory of Games*. New York: Cambridge University Press.

Milgrom, Paul and Roberts, John (1990) ‘‘Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,’’ *Econometrica*, **58**, 1255-1278.

Rasmusen, Eric (1989) *Games and Information*. New York: Blackwell.

Robson, Arthur (1992a) ‘‘Status, the Distribution of Wealth, Private and Social Attitudes Toward Risk,’’ *Econometrica*, **60**, 837-858.

Robson, Arthur (1992b) ‘‘The Biological Basis of Expected Utility, Knightian Uncertainty, and the Ellsberg Paradox,’’ Working Paper, Economics Department, University of Western Ontario, London Ontario.

Rousseau, Jean Jacques (1992) ‘‘Discourse on the Origin and Foundation of Inequality among Men (Second Discourse),’’ in *Collected Writings of Rosseau, Vol. 3,* ed. . Hanover, NH: University Press of New England.

Trivers, R. (1985) *Social Evolution*. Menlo Park, Ca.: Benjamin/Cummings.

Vives, Xavier (1990) ‘‘Nash Equilibrium with Strategic Complementarities,’’ *Journal of Mathematical Economics*, **19**, 305-321.

Young, H. Peyton (1993) ‘‘The Evolution of Conventions,’’ *Econometrica*, **61**, 57-84.