Comparative Statics of Games Between Relatives

IGAL MILCHTAICH

Department of Economics, Bar-Ilan University,
Ramat-Gan 52900, Israel

E-mail: milchti@mail.biu.ac.il
Web Page: http://faculty.biu.ac.il/~milchti

July 2005

Abstract

According to Hamilton's theory of kin selection, species tend to evolve behavior such that each organism appears to be attempting to maximize its inclusive fitness. In particular, two neighbors are likely to help each other if the cost of doing so is less than the benefit multiplied by r, their coefficient of relatedness. Since the latter is less than unity, mutual altruism benefits both neighbors. However, is it theoretically possible that acting so as to maximize the inclusive, rather than personal, fitness may <u>harm</u> both parties. This may occur in strategic symmetric pairwise interactions (more specifically, $n \times n$ games), in which the outcome depends on both sides' actions. In this case, the equilibrium outcome may be less favorable to the interactants' personal fitness than if each of them acted so as to maximize the latter. This paper shows, however, that such negative effect of relatedness on fitness is incompatible with evolutionary stability. If the symmetric equilibrium strategies are evolutionarily stable, a higher coefficient of relatedness can only entail higher personal fitness for the two neighbors. This suggests that negative comparative statics as above are not likely to occur in nature.

Keywords: Altruism, kin selection, evolutionary game theory, inclusive fitness ESS, uninvadability

1. Introduction

The social environments of many organisms consist of individuals that, on average, are more closely related, and hence genetically more similar, to them than other individuals in the population. Therefore, any behavior of an individual I that affects the reproductive success of its neighbors and has a genetic basis is likely to have a disproportionate effect on genes that are identical to those causing it. More precisely, for each neighbor J, the effect is proportional to the <u>coefficient of relatedness</u> r between I and J. In a large, outbred population, this coefficient is given by the probability that a gene at an arbitrary locus in I is <u>identical by descent</u> to any of the genes at the same locus in J, i.e., the two genes are copies of the same gene in a recent common ancestor (Michod and Hamilton, 1980). Thus, r coincides with Wright's coefficient of relationship, which can be calculated from pedigree ties (Crow and Kimura, 1970; Wright, 1922), e.g., $r = 0.5$ for full sibs and $r = 0.25$ for half-sibs. The total effect of a particular behavior of I on the genes responsible for it is measured by the change in I's <u>inclusive fitness</u> (Hamilton, 1964). The inclusive fitness can be defined as a weighted sum, whose terms are the number of adult offspring produced by I and by each of its neighbors J, the latter weighted by the coefficient of relatedness between I and J.[1] Any possible strategy that an organism does not use cannot increase the inclusive fitness of an individual adopting it, since if it could, a rare gene causing this behavior would spread. Thus, the results of evolution are such that each organism appears to be attempting to maximize its inclusive fitness (Hamilton, 1964).
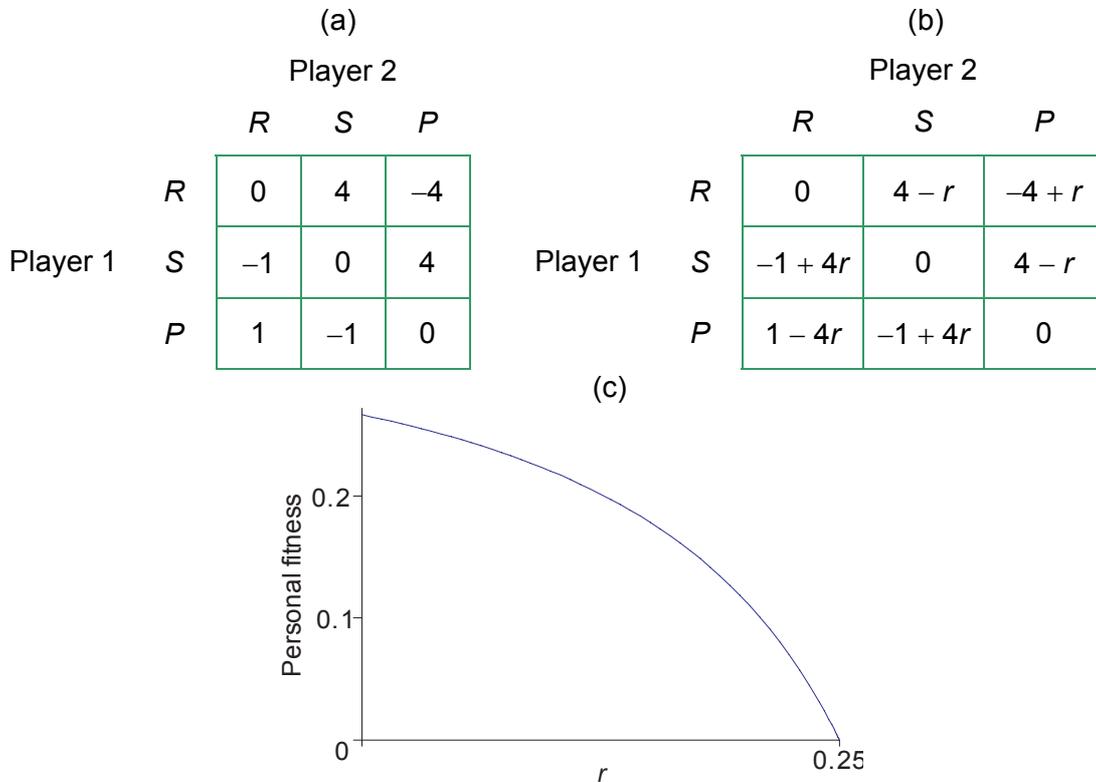
Hamilton's rule (Frank, 1998; Grafen, 1984; Hamilton, 1963, 1964) is a corollary of this principle. It says that an altruistic act is favored by natural selection if and only if

---

[1] According to Hamilton's (1964) definition of inclusive fitness, from this sum, a second weighted sum has to be subtracted, whose terms are the number of I's offspring produced only because of help received from neighbors, and the number of neighbors' offspring that would have been produced in the absence of I's influence, the latter weighted as above. The subtracted terms may be difficult to measure (Grafen, 1982; Lucas at al., 1996; Queller, 1996), and (as pointed out by Grafen, 1982, 1984) are sometimes omitted from the definition of inclusive fitness. Such omissions are inconsequential as far as changes in I's inclusive fitness caused by its own behavior are concerned, since by definition the omitted terms are not affected by this behavior.

$$-c + r\,b > 0, \tag{1}$$

where c is the cost to the actor's fitness, b is the benefit to the recipient's fitness, and r is the coefficient of relatedness between them. This inequality expresses the condition that the altruistic act increases the actor's inclusive fitness. It follows from Hamilton's rule that closely related neighbors will act altruistically towards one another whenever more distantly related ones (whose r is smaller) would do, but not conversely. Each such act benefits the recipient more than it costs the actor (since $r \leq 1$). Therefore, if each neighbor is equally likely to be in a position of potential donor or recipient, both neighbors' <u>personal</u> fitness, i.e., the number of their adult offspring, should be higher the more closely they are related.

(a)

Player 2

		R	S	P
	R	0	4	−4
Player 1	S	−1	0	4
	P	1	−1	0

(b)

Player 2

		R	S	P
	R	0	$4 - r$	$-4 + r$
Player 1	S	$-1 + 4r$	0	$4 - r$
	P	$1 - 4r$	$-1 + 4r$	0

(c)

Fig. 1. Relatedness and personal fitness in a generalized rock–scissors–paper game. (a) The payoff matrix. The cells show the change in player 1's personal fitness for each choice of pure strategies by the two players. (b) The inclusive fitness game. The cells show the change in player 1's inclusive fitness, defined as the change in 1's own fitness augmented by r times that of player 2, where r is the coefficient of relatedness between the two players. For each $0 \leq r \leq 0.25$, this game has a unique Nash equilibrium, in which both players use the same mixed strategy. (c) The personal fitness at equilibrium in the inclusive fitness game. Personal fitness decreases with increasing r, which indicates (see the text) that the equilibrium strategies are not evolutionarily stable.

4

The main question this paper addresses is whether this positive relation between relatedness and personal fitness extends to general symmetric interactions. The case considered above is special in that the interaction is non-strategic (Milchtaich, 2005), or additive. This means that the effects of the altruistic act on the actor and the recipient do not depend on how the latter behaves. In a strategic symmetric interaction, or game, the same act can have different effects on the actor depending on the other player's behavior. This case differs significantly from the non-strategic one in that two individuals, each acting so as to maximize its inclusive fitness, may have <u>lower</u> personal fitness at a symmetric equilibrium the higher their coefficient of relatedness. The example in Fig. 1 demonstrates this. However, this paper shows that such negative effect of relatedness on fitness is unlikely to occur in nature. This is because it necessarily involves symmetric equilibrium strategies that are not evolutionarily stable. With stable strategies, the conclusion reached in the special, non-strategic, case generalizes. That is, the effect of a symmetric game on the personal fitness of related players can only be the same or more positive than for unrelated ones, and it can only increase with increasing relatedness or remain unchanged.

2. Comparative statics

A symmetric two-player game with n pure strategies is specified by an $n \times n$ payoff matrix $\mathbf{A} = [a_{ij}]$. If one (the row) player uses strategy i and the other (the column player) uses j, the former's personal payoff changes by a_{ij} and the latter's by a_{ji}. The change in the first player's inclusive fitness is $a_{ij} + r\,a_{ji}$, where r is the coefficient of relatedness between the two players. For an individual using a mixed strategy

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

which assigns probability x_i to each pure strategy i, against a relative using a mixed strategy $\mathbf{y}$, the expected changes in the personal and inclusive fitness are given by $\mathbf{x}'\mathbf{A}\,\mathbf{y}$ and $\mathbf{x}'(\mathbf{A} + r\,\mathbf{A}')\,\mathbf{y}$, respectively, where the prime (′) denotes transposition. The symmetric game with payoff matrix $\mathbf{A} + r\,\mathbf{A}'$ is referred to below as the <u>inclusive fitness game</u>. A mixed strategy $\mathbf{x}$ is a symmetric Nash equilibrium strategy in this

game if an individual facing a relative using **x** would maximize its inclusive fitness by doing the same. In other words, the equilibrium condition is that, for all mixed strategies **y**,

$$(\mathbf{x}' - \mathbf{y}')(\mathbf{A} + r\,\mathbf{A}')\,\mathbf{x} \geq 0. \tag{2}$$

A symmetric equilibrium in the inclusive fitness game changes both players' <u>personal</u> fitness by $\mathbf{x}'\mathbf{A}\,\mathbf{x}$. In the generalized rock–scissors–paper game example in Fig. 1, this change is positive for $r = 0$ but decreases to zero as r approaches 0.25. Thus, the personal fitness at equilibrium is higher for unrelated players than, e.g., half-sibs using strategies maximizing their inclusive fitness.

 Plotting the fitness effect of an interaction as a function of the coefficient of relatedness is an example of comparative statics analysis. Economists are familiar with the idea, known as the correspondence principle (Samuelson, 1983), that the conditions under which such analysis leads to conclusions usually regarded as "normal" often coincide with the conditions for stability. For example, in a competitive market, the equilibrium price of a good (which equates supply and demand) increases with increasing demand if the equilibrium is stable, but not if it is unstable. The negative effect of varying r on fitness in Fig. 1(c) may similarly be attributed to the failure of the Nash equilibrium strategies involved to satisfy the more stringent condition of being evolutionarily stable strategies (ESSs) in the corresponding inclusive fitness games. By definition (Maynard Smith, 1982), a symmetric Nash equilibrium strategy **x** in a game with payoff matrix $\mathbf{A} + r\,\mathbf{A}'$ is an ESS if and only if, for every mixed strategy $\mathbf{y} \neq \mathbf{x}$ for which (2) holds as equality, the following inequality holds:

$$(\mathbf{x}' - \mathbf{y}')(\mathbf{A} + r\,\mathbf{A}')\,\mathbf{y} > 0. \tag{3}$$

Such a strategy **x** is referred to below as an <u>inclusive fitness ESS</u>. An equivalent definition, formulated directly in terms of the base payoff matrix **A**, is given by the following proposition, the proof of which is given in Appendix A.

Proposition. *For specified* $0 < r \leq 1$, *a mixed strategy* **x** *is an inclusive fitness ESS if and only if it has a neighborhood in the space of mixed strategies where, for all* $\mathbf{y} \neq \mathbf{x}$,

$$\mathbf{x}'\mathbf{A}\,\mathbf{x} > \mathbf{y}'\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,\mathbf{x}). \tag{4}$$

In other words, an inclusive fitness ESS **x** is characterized by the property that the expression on the right-hand side of (4), if seen as a function of **y**, has a strict local maximum at **y** = **x**. That expression gives the expected change in fitness of a player whose opponent uses the same mixed strategy **y** as it does with probability r, and with probability $1 - r$ plays **x**. Thus, the Proposition implies that **x** is an inclusive fitness ESS if and only if it would locally maximize the personal fitness of such a player. This shows that being an inclusive fitness ESS is a local rather than global property of a strategy. In this, it differs from the uninvadability criterion derived in Section 3 below.

In a generalized rock–scissors–paper game, a necessary and sufficient condition for the existence of an ESS is that, whenever the two players use different pure strategies, the sum of the their payoffs is positive (Hofbauer and Sigmund, 1998, Theorem 7.7.1; Weissing, 1991, Theorem 4.6). The inclusive fitness game in Fig. 1(b) does not satisfy this condition for any $0 \leq r \leq 0.25$, since the sum of the players' payoffs when they play R and P is negative, $-3 - 3r$. Therefore, in this example, none of the symmetric Nash equilibrium strategies is an inclusive fitness ESS. As the next theorem shows, if the equilibrium strategies were ESSs, the personal fitness could only increase with increasing r or remain unchanged. The proof of the Theorem is given in Appendix B.

Theorem. *Suppose that, for specified $-1 \leq r_1 < r_2 \leq 1$, there is a continuous function assigning for each $r_1 \leq r \leq r_2$ a corresponding inclusive fitness ESS, $\mathbf{x}(r)$, such that $\mathbf{x}(r_1) \neq \mathbf{x}(r_2)$. Then,*

$$\mathbf{x}(r_2)' \, \mathbf{A} \, \mathbf{x}(r_2) > \mathbf{x}(r_1)' \, \mathbf{A} \, \mathbf{x}(r_1). \tag{5}$$

This comparative statics result is rather general. It applies to any symmetric two-player game in which the players can use a finite number of pure strategies and any mixed strategy. It is applicable to different populations, in each of which the players' relatedness is always the same, and different from that in the other population; and to single populations, in which individuals use conditional strategies that prescribe different actions depending on the players' relatedness. It also covers negative as well as positive coefficients of relatedness. Negative relatedness ($r < 0$) may occur, and possibly lead to spiteful behavior, in small populations, if the two interacting individuals are genetically less similar to each other than to random members of the

population (Grafen, 1985; Hamilton 1970, 1971, 1972). The Theorem also allows for multiple ESSs. Regardless of the number of inclusive fitness ESSs for r_1, normally (specifically, if these ESSs are regular), for r_2 sufficiently close to r_1, for each of them there is a continuous function $\mathbf{x}(r)$ as above connecting it with some inclusive fitness ESS for r_2 (Milchtaich, 2005). By the Theorem, each of these pairs of corresponding inclusive fitness ESSs, $\mathbf{x}(r_1)$ and $\mathbf{x}(r_2)$, satisfies (5), unless $\mathbf{x}(r_1) = \mathbf{x}(r_2)$.

The inequality (5) need not hold for <u>all</u> pairs of distinct inclusive fitness ESSs with $r_1 < r_2$. The assumption that the two are connected by a line of inclusive fitness ESSs corresponding to intermediate values of r is indispensable. A simple example showing this is the two-strategy coordination game in which both players' payoff is 1 or 1/2 if they both choose strategy 1 or 2, respectively, and zero if they choose different strategies. For any r, both pure strategies are inclusive fitness ESSs. However, if $\mathbf{x}(r_1)$ and $\mathbf{x}(r_2)$ are chosen as strategies 1 and 2, respectively, the opposite inequality to (5) holds.

In general, evolutionary stability is only a sufficient condition for the expected personal fitness to increase with continuously increasing r. However, in the important special case of symmetric 2×2 games, it is also close to being necessary. In these games, the payoff matrix can always be put in the following normalized form, by subtracting a constant from all entries:

$$\mathbf{A} = \begin{bmatrix} 0 & b \\ -c & b - c + d \end{bmatrix}, \tag{6}$$

where a, b, c and d can be positive or negative. If the game has a completely mixed (i.e., not pure) symmetric Nash equilibrium, the equilibrium strategy is an ESS if and only if $d < 0$. This condition entails <u>strategic substitutability</u> (Bulow et al., 1985): the profitability of switching from one pure strategy to the other is lower if the other player uses the latter, rather than the former, strategy. In the hawk–dove game (Maynard Smith, 1982), d is always negative, and equal to each individual's expected loss of fitness due to injury if both opponents escalate the fight. By the Theorem, this implies that, in this game, the expected personal fitness at the symmetric equilibrium can only increase with increasing r. By contrast, in the prisoner's dilemma (where $b > c - d > 0$ and $b + 2d > c > 0$; Maynard Smith, 1982), d can be negative, positive, or zero. In all three cases, (i) if $r \le 0$, the game has a unique Nash equilibrium, in which

both players defect; (ii) both players cooperating is a Nash equilibrium for $r = 1$; and (iii) completely mixed symmetric equilibria exist in one or (normally) more inclusive fitness games with $0 < r < 1$. Whether the expected personal fitness corresponding to the latter increases or decreases with increasing r depends on evolutionary stability, which is determined by the sign of d (see Fig. 2).
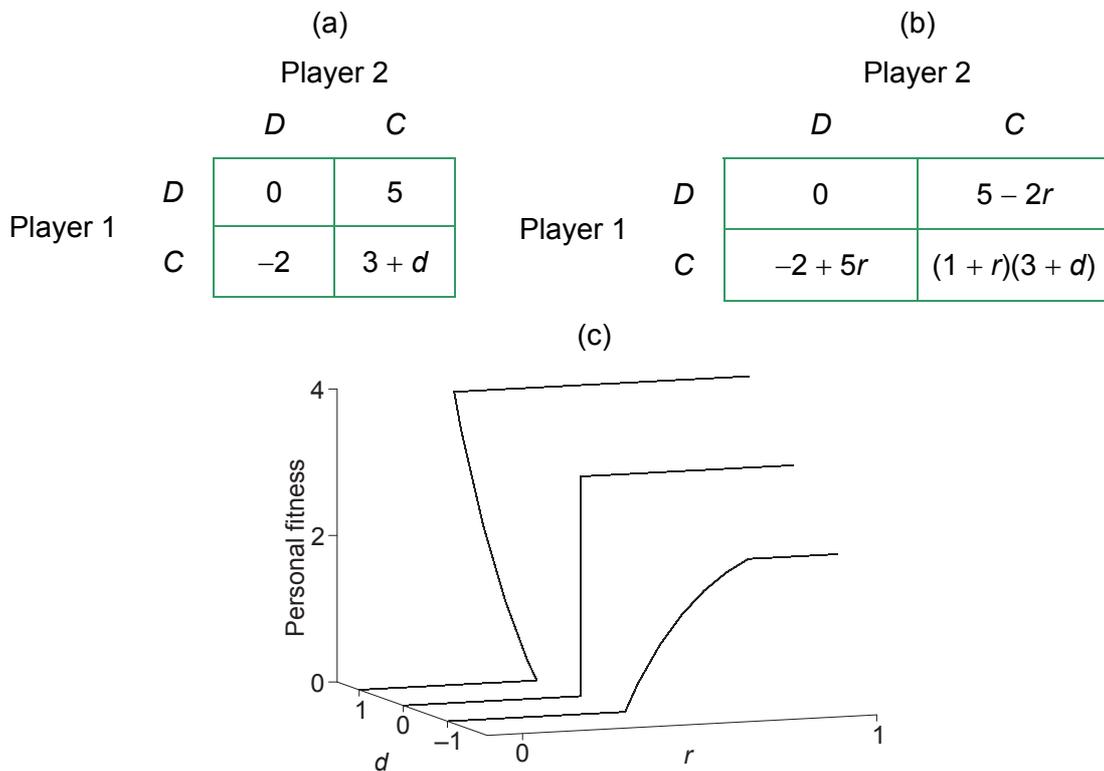
Fig. 2. Three kinds of prisoner's dilemma. (a) The payoff matrix. The parameter d is between -1 and 1. Choosing to cooperate, C, rather than defect, D, always lowers a player's personal fitness and raises that of the other player. The inclusive fitness can be either higher or lower. (b) The inclusive fitness game. Depending on d and the coefficient of relatedness r, this game has one or more symmetric Nash equilibria, which can be completely mixed (i.e., both D and C used with positive probability) or pure. (c) The personal fitness at equilibrium in the inclusive fitness game. For a completely mixed equilibrium, the personal fitness can increase ($d < 0$) or decrease ($d > 0$) with increasing r. In the former case, the equilibrium strategies are evolutionarily stable (i.e., ESSs), and in the latter, they are unstable.

3. Uninvadable strategies

"Evolutionary stability" is often used in a broader sense than in Section 2. Namely, an evolutionarily stable strategy is such that, if all the members of a population adopt it,

no mutant strategy can invade (Hammerstein, 1998, p. 7; Maynard Smith, 1982, p. 204). In order not to confuse this notion of stability with that considered above, it is referred to below as <u>uninvadability</u>. In a symmetric game between relatives, an inclusive fitness ESS is not necessarily uninvadable. However, as Hines and Maynard Smith (1979) showed (building on the work of Grafen, 1979), the <u>converse</u> is true: every uninvadable strategy is an inclusive fitness ESS. This suggests that related players are only likely to use a strategy that is: (i) a Nash equilibrium strategy in the corresponding inclusive fitness game, i.e., a maximizer of each player's inclusive fitness; and, moreover, (ii) an ESS in that game. This conclusion, which may be viewed as an extension of Hamilton's rule to strategic symmetric interactions, justifies the assumptions made in the Proposition and Theorem in the previous section.

The connection between uninvadable strategies and inclusive fitness ESSs is illustrated by the simple example of the additive prisoner's dilemma, in which $d = 0$ in (6). By Hamilton's rule, in this game, cooperation or defection is favored by natural selection if the expression on the left-hand side of (1) is positive or negative, respectively. Since that expression gives the change in inclusive fitness of a player switching from defection to cooperation, in both cases the same strategy is also the unique inclusive fitness ESS. A less trivial example, involving a strategic interaction, is provided by Cavalli-Sforza and Feldman's (1978) multiplicative model of diploid sib-to-sib altruism, in which $d = -b\,c < 0$ (Queller, 1985). Uyenoyama and Feldman (1982) showed that, in this exact population genetic model, there is a unique level of expression of altruism which, when associated with a homozygous genotype at fixation, precludes invasion by genotypes showing any other level of altruistic tendency. The uninvadable strategy (or "strong form ESS") is to perform the altruistic act with probability $x = (1/3)(1/c - 2/b)$ (or 0 or 1, if this expression is negative or greater than unity, respectively). It follows from the Proposition that this is also the unique inclusive fitness ESS. This is because, for any other mixed strategy, whereby the altruistic act is performed with probability y, the expression on the right-hand side of (4) (where **A** is given by (6), $d = -b\,c < 0$, and $r = 0.5$) is equal to $(1 - c\,y)(1 + b\,(x + y)/2)$ plus a constant. This quadratic function of y has a unique maximum point at $y = (1/2)(1/c - 2/b - x)$, which coincides with x if and only if $x = (1/3)(1/c - 2/b)$. For another example of a game between diploid sibs in which an uninvadable strategy is necessarily an inclusive fitness ESS, see Tao and Lessard (2002).

Hines and Maynard Smith's (1979) analysis, as well as Grafen's (1979), concerns only asexual or haploid sexual populations. However, it is shown below that (as the last examples suggest) their finding holds for diploid sexual populations as well. Below, the analysis of the sexual-population case follows a proof of a somewhat modified version of Hines and Maynard Smith's result for asexual populations.

Consider a large population consisting of a set of parthenogenetic clones, each adopting either a mixed (or, as a special case, pure) strategy $\mathbf{x}$ or a mutant strategy $\mathbf{y}$ in a symmetric game with an $n \times n$ payoff matrix $\mathbf{A}$. A fraction $0 < r < 1$ of all contests is between members of the same clone (which always use the same strategy) and the rest between randomly assorted opponents. Thus, r is the probability that the two opponents will be identical by descent, and so has some resemblance to the coefficient of relatedness in sexual populations (Maynard Smith, 1982, Appendix F). If the fraction of the population using the mutant strategy $\mathbf{y}$ is ε, then, given that a player's strategy is $\mathbf{x}$ or $\mathbf{y}$, the conditional probability that the opponent uses the same strategy is given, respectively, by

$$\mathsf{P}(\mathbf{x} \mid \mathbf{x}) = r + (1 - r)\,(1 - \varepsilon) \ \text{ or } \ \mathsf{P}(\mathbf{y} \mid \mathbf{y}) = r + (1 - r)\,\varepsilon.$$

The effect of the game on a player's fitness can be calculated from these conditional probabilities. For a player using strategy $\mathbf{x}$ or $\mathbf{y}$, the expected change in personal fitness is given, respectively, by

$$\mathbf{x}'\,\mathbf{A}\,(r\,\mathbf{x} + (1 - r)\,((1 - \varepsilon)\,\mathbf{x} + \varepsilon\,\mathbf{y}))$$

or

$$\mathbf{y}'\,\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,((1 - \varepsilon)\,\mathbf{x} + \varepsilon\,\mathbf{y})).$$

The former is greater if and only if

$$\mathbf{x}'\,\mathbf{A}\,\mathbf{x} > \mathbf{y}'\,\mathbf{A}\,((1 - r)\,\mathbf{x} + r\,\mathbf{y}) + \varepsilon\,(1 - r)\,(\mathbf{x}' - \mathbf{y}')\,\mathbf{A}\,(\mathbf{x} - \mathbf{y}).$$

The mutant strategy $\mathbf{y}$ decreases in frequency when rare if and only if this inequality holds for sufficiently small $\varepsilon > 0$. This gives the following uninvadability criterion (Hines and Maynard Smith, 1979): A necessary and sufficient condition for a strategy $\mathbf{x}$ to be uninvadable by any mutant is that, for every $\mathbf{y} \neq \mathbf{x}$, either

$$\mathbf{x}'\,\mathbf{A}\,\mathbf{x} > \mathbf{y}'\,\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,\mathbf{x}), \tag{7}$$

or

$$\mathbf{x}' \mathbf{A} \mathbf{x} = \mathbf{y}' \mathbf{A} (r \mathbf{y} + (1 - r) \mathbf{x}) \ \text{ and } \ (\mathbf{x}' - \mathbf{y}') \mathbf{A} (\mathbf{x} - \mathbf{y}) < 0. \qquad (8)$$

In fact, (8) is redundant: The above uninvadability criterion actually implies that (8) does <u>not</u> hold for any $\mathbf{y}$. This is because applying this criterion to the mid-point strategy $\mathbf{z} = (1/2) \mathbf{x} + (1/2) \mathbf{y}$ (instead of $\mathbf{y}$) gives, in particular, that

$$\mathbf{x}' \mathbf{A} \mathbf{x} - \mathbf{z}' \mathbf{A} (r \mathbf{z} + (1 - r) \mathbf{x}) \geq 0.$$

This inequality is equivalent to

$$\mathbf{x}' \mathbf{A} \mathbf{x} - \mathbf{y}' \mathbf{A} (r \mathbf{y} + (1 - r) \mathbf{x}) + (r/2) (\mathbf{x}' - \mathbf{y}') \mathbf{A} (\mathbf{x} - \mathbf{y}) \geq 0$$

(where the left-hand side is twice that in the previous one, as can be checked by writing both expressions in terms of $\mathbf{x}' \mathbf{A} \mathbf{x}$ etc.). The last inequality contradicts (8). Therefore, Hines and Maynard Smith's (1979) uninvadability criterion is equivalent to the requirement that (7) holds for all $\mathbf{y} \neq \mathbf{x}$. By the Proposition, a strategy $\mathbf{x}$ is an inclusive fitness ESS if an only if it has a <u>neighborhood</u> where (7) holds for all $\mathbf{y} \neq \mathbf{x}$. This shows that an uninvadable strategy is necessarily an inclusive fitness ESS (but not conversely). Indeed, a strategy is an inclusive fitness ESS if and only if it is "locally" uninvadable, i.e., immune to invasion by close strategies.

The first to derive a criterion for uninvadability (similar to (7)) in a game between relatives by arguing from a personal fitness approach was Grafen (1979). Hines and Maynard Smith (1979) accepted Grafen's argument, but maintained that the inclusive fitness approach is valuable in providing a necessary condition. Their recommended procedure was to seek a candidate for uninvadable strategy (or "Grafen ESS", as they call it) by finding an inclusive fitness ESS. The above analysis further clarifies the relation between these two approaches. As it shows, being an inclusive fitness ESS is equivalent to immunity to invasion by a <u>particular kind</u> of mutant mixed strategies, namely, those obtained by only slightly changing the probably of each pure strategy.

To demonstrate that an inclusive fitness ESS is a necessary condition for uninvadability also in the sexual-population case, similar logic can be used. That is, it suffices to show that any strategy that resists invasion by a particular kind of mutant genes is an inclusive fitness ESS. For sexually reproducing haploids whose strategy is determined by two alleles at a single locus (Hines and Maynard Smith, 1979), the arguments used in the asexual case carry over with little modification. In this case, r is the probability that the players' genes at that locus are identical by descent, i.e., are

copies of the same gene in a recent common ancestor. The more interesting case of a diploid sexual population is more difficult (Maynard Smith, 1982, Appendix F). Here, two alleles B_1 and B_2 at a single autosomal locus give three possible mixed strategies: $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ for B_1B_1, B_1B_2 and B_2B_2 individuals, respectively. A relatively simple case is that in which the gene effects are small and additive. This means that, for each pure strategy i, (i) the two homozygous types use i with close probabilities x_i and z_i, and (ii) the probability y_i that a heterozygote uses i is halfway between them. These assumptions imply weak selection: An individual's fraction of B_2 alleles, β (which can be 0, 1/2, or 1), has only a weak effect on behavior, and hence also on the individual's fitness. The additivity assumption, (ii), means that the dependence of the strategy on β is linear. That assumption, which can also be written as

$$\mathbf{y} = \frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{z}, \tag{9}$$

holds, for example, if only one of the two homologous genes, selected at random, is expressed in each individual, so that half the heterozygotes use the mixed strategy $\mathbf{x}$ and half use $\mathbf{z}$. Since this is a special case (in particular, it does not allow for dominance), resistance to invasion by such genes is only a necessary condition for a strategy $\mathbf{x}$ to be uninvadable. However, it is shown below that such partial uninvadability already implies that $\mathbf{x}$ is an inclusive fitness ESS.

Consider a large, randomly mating, outbred population of sexually reproducing diploids in which pairs of related players, all with the same coefficient of relatedness $0 < r < 1$, play a symmetric game with an $n \times n$ payoff matrix $\mathbf{A}$. Strategy is determined by two alleles, B_1 and a mutant allele B_2, satisfying assumptions (i) and (ii) above. As explained, this implies that, for a randomly selected pair of players, the strategies, $\mathbf{s}$ and $\mathbf{t}$, are connected with the respective fractions of B_2 alleles, β and γ, by the linear relations

$$\mathbf{s} = (1 - \beta)\,\mathbf{x} + \beta\,\mathbf{z} \quad \text{and} \quad \mathbf{t} = (1 - \gamma)\,\mathbf{x} + \gamma\,\mathbf{z}. \tag{10}$$

Since β and γ have the same expectation, which equals the frequency of B_2 in the population, p, the two players have equal expected strategies:

$$\mathbf{Es} = \mathbf{Et} = (1 - p)\,\mathbf{x} + p\,\mathbf{z}. \tag{11}$$

Since the players are related, their genotypes are (stochastically) dependent. Indeed,

the regression of γ on β is linear, and is given by $\mathsf{E}(\gamma \,|\, \beta) = r\beta + (1 - r)\,p$ (Jacquard, 1974, p. 118. For an extensive discussion of the view of relatedness as a regression coefficient, see Grafen, 1985, who also describes the earlier works of Hamilton and Price.) It follows, by (10) and (11), that the conditional expectation of the opponent's strategy $\mathbf{t}$, given $\mathbf{s}$, satisfies

$$\mathsf{E}(\mathbf{t} \,|\, \mathbf{s}) = r\,\mathbf{s} + (1 - r)\,\mathsf{E}\mathbf{s}. \tag{12}$$

This may be interpreted as meaning that with probability r an individual plays against an opponent using the same strategy $\mathbf{s}$ as it does because they are related, and with probability $1 - r$ it plays against the strategies according to their proportions in the population (Grafen, 1979).

It follows from (12) that the conditional expectation of the payoff (i.e., change in personal fitness) of a player with strategy $\mathbf{s}$ is given by

$$\mathsf{E}(\mathbf{s}'\mathbf{A}\,\mathbf{t} \,|\, \mathbf{s}) = \mathbf{s}'\mathbf{A}\,\mathsf{E}(\mathbf{t} \,|\, \mathbf{s}) = \mathbf{s}'\mathbf{A}\,(r\,\mathbf{s} + (1 - r)\,\mathsf{E}\mathbf{s}).$$

Substituting $\mathbf{x}$, $\mathbf{y}$ or $\mathbf{z}$ for $\mathbf{s}$ in this expression gives the expected payoff of $B_1 B_1$, $B_1 B_2$ or $B_2 B_2$ individuals, respectively. The frequency of the mutant allele B_2 will decrease if and only if these expected payoffs satisfy the following inequality (Crow and Kimura, 1970, p. 182):

$$p\left[\mathbf{z}'\mathbf{A}\,(r\,\mathbf{z} + (1 - r)\,\mathsf{E}\mathbf{s}) - \mathbf{y}'\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,\mathsf{E}\mathbf{s})\right]$$

$$+ (1 - p)\left[\mathbf{y}'\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,\mathsf{E}\mathbf{s}) - \mathbf{x}'\mathbf{A}\,(r\,\mathbf{x} + (1 - r)\,\mathsf{E}\mathbf{s})\right] < 0.$$

Using (9) and (11) (for eliminating $\mathbf{z}$ and $\mathsf{E}\mathbf{s}$), and after certain algebraic manipulations, this inequality gives the following equivalent one:

$$\mathbf{x}'\mathbf{A}\,\mathbf{x} > \mathbf{y}'\mathbf{A}\,(r\,\mathbf{y} + (1 - r)\,\mathbf{x}) + 2\,p\,(\mathbf{x}' - \mathbf{y}')\,\mathbf{A}\,(\mathbf{x} - \mathbf{y}).$$

It follows that B_2 will decrease in frequency when rare (i.e., $p \approx 0$) if and only if either (7) or (8) holds. Therefore, a necessary condition for a diploid population in which all the members play $\mathbf{x}$ to resist invasion by any mutant gene (and, in particular, any gene of the kind considered above) is that, for every $\mathbf{y} \neq \mathbf{x}$ close to $\mathbf{x}$, either (7) or (8) holds. A similar argument to one used above shows that this condition is in fact equivalent to the requirement that there is a neighborhood of $\mathbf{x}$ where (7) holds for all $\mathbf{y} \neq \mathbf{x}$. By the Proposition, this requirement is satisfied if and only if $\mathbf{x}$ is an inclusive fitness ESS. Therefore, the latter is a necessary condition for uninvadability.

4. Discussion

The results in the last section show that, in a symmetric game between relatives, any strategy that is not an inclusive fitness ESS can be invaded. Therefore, a phenotypically monomorphic population is unlikely to use such a strategy. By the Theorem, comparative statics as in Fig. 1(c), in which relatedness has a negative effect on fitness, necessarily involve equilibrium strategies that are not inclusive fitness ESSs. This suggests that such comparative statics are unlikely to occur in nature. Thus, for the same game in two different populations, in one of which the players are more closely related than in the other, the expected contribution to personal fitness is likely to be the same or higher in the first population. A similar comparative statics result applies to single populations, in which individuals use conditional strategies that prescribe different actions depending on their relatedness.

These conclusions need not hold in settings in which mutation (or other sources of variation) or selection do not occur. In particular, it is unclear to what extent they hold for human social interactions. In this case, the coefficient r, rather than reflecting kinship, may be interpreted as modeling in an elementary manner the type and intensity of one person's sentiments towards the other. Examples such as in Fig. 1 (and those in Milchtaich, 2005) show that, with unstable equilibria, two people's material gains from a strategic symmetric interaction may be higher if they are both selfish ($r = 0$) rather than mildly caring ($0 < r < 1$), and even higher if they resent each other ($r < 0$). However, if social dynamics tend to exclude unstable equilibria, this kind of "paradoxical" comparative statics is unlikely to occur.

The above results only concern symmetric $n \times n$ games, which have a finite number of pure strategies, a continuum of mixed ones, and bilinear payoff functions. Similar results may also hold for certain classes of symmetric two-player games with non-bilinear payoff functions. However, they do not hold in all such games. In some games, a similar connection to that established above between uninvadable strategies and inclusive fitness ESSs does not hold (Mesterton-Gibbons, 1996).

Acknowledgements

Appendix A. Proof of the Proposition

It has to be shown that, for specified $n \times n$ payoff matrix $\mathbf{A}$ and $0 < r \le 1$, the condition in the Proposition is necessary and sufficient for a strategy $\mathbf{x}$ to be an inclusive fitness ESS, i.e., an ESS in the game with payoff matrix $\mathbf{A} + r \mathbf{A'}$.

To prove that the condition is sufficient, suppose that it holds for $\mathbf{x}$, and let $\mathbf{y}$ be another mixed strategy. By the condition, for sufficiently small $0 < \delta < 1$, the mixed strategy $\mathbf{z} = (1 - \delta) \mathbf{x} + \delta \mathbf{y}$ satisfies

$$\mathbf{x'} \mathbf{A} \mathbf{x} - \mathbf{z'} \mathbf{A} (r \mathbf{z} + (1 - r) \mathbf{x}) > 0. \tag{A.1}$$

This inequality is equivalent to

$$(1 + (1 - \delta) r) (\mathbf{x'} - \mathbf{y'}) (\mathbf{A} + r \mathbf{A'}) \mathbf{x} + \delta r (\mathbf{x'} - \mathbf{y'}) (\mathbf{A} + r \mathbf{A'}) \mathbf{y} > 0 \tag{A.2}$$

(where the left-hand side is equal to $(1 + r)/\delta$ times that in (A.1), as can be checked by writing both expressions in terms of $\mathbf{x'} \mathbf{A} \mathbf{x}$ etc.). It follows that either (2) holds as a strict inequality, or it holds as equality and (3) also holds. Since this is true for any $\mathbf{y} \neq \mathbf{x}$ (not necessarily close to $\mathbf{x}$), $\mathbf{x}$ is an inclusive fitness ESS.

To prove that the condition in the Proposition is necessarily, consider a mixed strategy $\mathbf{x}$ that is an ESS in the game with payoff matrix $\mathbf{A} + r \mathbf{A'}$. There is a neighborhood of $\mathbf{x}$ where, for all $\mathbf{y} \neq \mathbf{x}$, both (2) and (3) hold (Weibull, 1995, proposition 2.6). These two inequalities imply that (A.2) holds for $\delta = 1$, and hence, (A.1) holds for $\mathbf{z} = \mathbf{y}$.

Appendix B. Proof of the Theorem

The proof is based on the following algebraic identity, which can easily be shown to hold for every pair of mixed strategies $\mathbf{x}$ and $\mathbf{y}$ and numbers r and s:

$$(1 + r) (r - s) (\mathbf{x'} \mathbf{A} \mathbf{x} - \mathbf{y'} \mathbf{A} \mathbf{y}) = (r - s) (\mathbf{x'} - \mathbf{y'}) (\mathbf{A} + r \mathbf{A'}) \mathbf{x} \tag{B.1}$$

$$+ (1 - r s) (\mathbf{x'} - \mathbf{y'}) (\mathbf{A} + r \mathbf{A'}) \mathbf{y} + (1 - r^2) (\mathbf{y'} - \mathbf{x'}) (\mathbf{A} + s \mathbf{A'}) \mathbf{y}.$$

Suppose that $\mathbf{x}$ is an ESS in the game with payoff matrix $\mathbf{A} + r \mathbf{A'}$. There is a neighborhood of $\mathbf{x}$ in the space of mixed strategies where, for all $\mathbf{y} \neq \mathbf{x}$, both (2) and (3) hold (Weibull, 1995, proposition 2.6). These inequalities imply that, if $-1 \le s < r \le 1$, the sum of the first and second terms on the right-hand side of (B.1) is positive.

If, in addition, **y** is an ESS in the game with payoff matrix $\mathbf{A} + s\,\mathbf{A}'$, then the whole right-hand side of (B.1) is positive, and hence

$$\mathbf{x}'\mathbf{A}\,\mathbf{x} - \mathbf{y}'\mathbf{A}\,\mathbf{y} > 0. \tag{B.2}$$

The rest of the proof is essentially a continuity argument. Since, by assumption, the function $r \mapsto \mathbf{x}(r)$ is continuous, there is some mixed strategy **x** such that:

(i)   for some $r_1 \le r \le r_2$, $\mathbf{x}(r) = \mathbf{x}$, and,

(ii)   for all $r_1 \le s \le r_2$, $\mathbf{x}(s)'\mathbf{A}\,\mathbf{x}(s) \ge \mathbf{x}'\mathbf{A}\,\mathbf{x}$.

To prove (5), it suffices to show that, necessarily, $\mathbf{x} = \mathbf{x}(r_1)$ (which implies that $\mathbf{x}(r_2)'\mathbf{A}\,\mathbf{x}(r_2) \ne \mathbf{x}(r_1)'\mathbf{A}\,\mathbf{x}(r_1)$, for otherwise (i) and (ii) would hold also for $\mathbf{x} = \mathbf{x}(r_2)$). Consider the smallest $r_1 \le r \le r_2$ for which $\mathbf{x}(r) = \mathbf{x}$. If $r > r_1$, then, for all $r_1 \le s < r$, $\mathbf{x}(s) \ne \mathbf{x}$. On the other hand, by the continuity assumption, for s tending to r, $\mathbf{x}(s)$ tends to **x**. Therefore, it follows from the first part of the proof that, if s is less than but sufficiently close to r, (B.2) holds for $\mathbf{y} = \mathbf{x}(s)$. However, this contradicts (ii). This contradiction shows that, necessarily, $r = r_1$, and hence $\mathbf{x} = \mathbf{x}(r_1)$, as had to be shown.

References

Bulow, J. I., Geanakopolos, J. D., Klemperer, P. D., 1985. Multimarket oligopoly: strategic substitutes and complements. J. Polit. Economy 93, 488−511.

Cavalli-Sforza, L. L., Feldman, M. W., 1978. Darwinian selection and 'altruism'. Theor. Popul. Biol. 14, 268–280.

Crow, J. F., Kimura, M., 1970. An Introduction to Population Genetics Theory. Harper & Row, New York.

Frank, S. A., 1998. Foundations of Social Evolution. Princeton University Press, Princeton.

Grafen, A., 1979. The hawk–dove game played between relatives. Anim. Behav. 27, 905–907.

Grafen, A., 1982. How not to measure inclusive fitness. Nature 298, 425–426.

Grafen, A., 1984. Natural selection, kin selection and group selection. In: Krebs, J. R., Davies, N. B. (Eds.), Behavioural Ecology, second edition, Blackwell, Oxford, pp. 62−84.

Grafen, A., 1985. A geometric view of relatedness. In: Dawkins, R., Ridley, M. (Eds.), Oxford Surveys in Evolutionary Biology, vol. 2, Oxford University Press, Oxford, pp. 28–89.

Hamilton, W. D., 1963. The evolution of altruistic behavior. Am. Nat. 97, 354–356.

Hamilton, W. D., 1964. The genetical evolution of social behaviour I and II. J. Theor. Biol. 7, 1–52.

Hamilton, W. D., 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228, 1218–1220.

Hamilton, W. D., 1971. Selection of selfish and altruistic behavior in some extreme models. In: Eisenberg, J. F., Dillon, W. S. (Eds.), Man and Beast: Comparative Social Behavior, Smithsonian Institute Press, Washington, DC, pp. 57–91.

Hamilton, W. D., 1972. Altruism and related phenomena, mainly in social insects. Ann. Rev. of Ecol. and System. 3, 193–232.

Hammerstein, P., 1998. What is evolutionary game theory? In: Dugatkin, L. A., Reeve, H. K. (Eds.), Game Theory and Animal Behavior, Oxford University Press, Oxford, pp. 3–15.

Hines, W. G. S., Maynard Smith, J., 1979. Games between relatives. J. Theor. Biol. 79, 19–30.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.

Jacquard, A., 1974. The Genetic Structure of Populations. Springer-Verlag, Berlin.

Lucas, J. R., Creel, S. R., Waser, P. M., 1996. How to measure inclusive fitness, revisited. Anim. Behav. 51, 225–228.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Mesterton-Gibbons, M., 1996. On the war of attrition and other games among kin. J. Math. Biol. 34, 253–270.

Michod, R. E., Hamilton, W. D., 1980. Coefficients of relatedness in socio-biology. Nature 288, 694–697.

Milchtaich, I., 2005. Comparative statics of altruism and spite. Bar-Ilan University preprint, http://faculty.biu.ac.il/~milchti/papers/sympathy.pdf.

Queller, D. C., 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. Nature 318, 366–367.

Queller, D. C., 1996. The measurement and meaning of inclusive fitness. Anim. Behav. 51, 229–232.

Samuelson, P. A., 1983. Foundations of Economic Analysis, enlarged edition. Harvard University Press, Cambridge, MA.

Tao, Y., Lessard, S., 2002. Frequency-dependent selection in sexual family-structured populations. J. Theor. Biol. 217, 525–534.

Uyenoyama, M. K., Feldman, M., 1982. Population genetic theory of kin selection. II. The multiplicative model. Am. Nat. 120, 614–627.

Weibull, J. W., 1995. Evolutionary Game Theory. MIT Press, Cambridge, MA.

Weissing, F. J., 1991. Evolutionary stability and dynamic stability in a class of evolutionary normal form games. In: Selten, R. (Ed.), Game Equilibrium Models I, Springer-Verlag, Berlin, pp. 29–97.

Wright, S., 1922. Coefficients of inbreeding and relationship. Am. Nat. 56, 330–338.