Comparative statics of altruism and spite

Igal Milchtaich *

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

A R T I C L E   I N F O

A B S T R A C T

The equilibrium outcome of a strategic interaction may depend on the weight players place on other players' payoffs or, more generally, on some social payoff that depends on everyone's actions. A positive, negative or zero weight represents altruism, spite or complete selfishness, respectively. As it turns out, even in a symmetric interaction the equilibrium level of social payoff may be lower for a group of altruists than for selfish or spiteful groups. In particular, a concern for others' payoffs may paradoxically lower these payoffs. However, this can only be so if the equilibrium strategies involved are unstable. If they are stable, the social payoff can only increase or remain unchanged with an increasing degree of altruism. In these results, 'stability' stands for a general notion of static stability, which includes a number of established ones, such as evolutionarily stable strategy, as special cases.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Altruism and spite represent deviations in opposite directions from complete selfishness, or total disregard to the effect of one's actions on others' welfare. A person is altruistic or spiteful towards another if he is willing to bear a cost in order to benefit or harm the other person, respectively. This paper considers the question of the welfare consequences of such preferences. Is social welfare in a group in which everyone is equally altruistic or spiteful towards the others necessarily higher or lower, respectively, than in a group in which everyone is only concerned with his own good? Unlike much of the related literature (e.g., Frank, 1988; Ridley, 1997) this question only involves the consequences of deviations from complete selfishness, not their origin or evolution. The common degree of altruism or spite r is viewed as an exogenous parameter, representing, for example, a shared moral value or social attribute. The parameter quantifies the extent to which each individual i internalizes social welfare, for example, the ratio between the weight i attaches to the payoff of every other individual j and the weight of his own, personal payoff. It thus determines the individuals' preferences over action profiles and in particular their best responses to the others' actions. Comparative statics examines the resulting effect on the actual, material payoffs.

The propensity for acting altruistically or spitefully may have a biological basis. In particular, with family members it may be the result of kin selection. According to Hamilton's rule (Hamilton, 1963, 1964; Frank, 1998), natural selection favors acts that maximize the actor's inclusive fitness, which is his own fitness augmented by r times that of each of the other affected individuals, where r is their coefficient of relatedness. This stems from the fact that helping a relative assists the propagation of the actor's own genes; as the coefficient of relatedness increases, so does the number of shared genes.

---

* Fax: +972 3 7384034.
E-mail address: igal.milchtaich@biu.ac.il.
URL: http://faculty.biu.ac.il/~milchti.

The inclusive fitness has the functional form described above. The fitness of each individual enters linearly, and the weight attached to it, which is the corresponding coefficient of relatedness r, is exogenous, specifically, determined by the family tree. Thus, comparative statics analysis might reveal, for example, how the expected consequences of a particular interaction involving two or more individuals depends on their relatedness, e.g., whether they are full or half-siblings.

The simple, linear model of interdependent preferences assumed in this paper excludes a number of more sophisticated models suggested in the literature (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), which might better predict people's behavior in certain experimental settings. In particular, it does not cover "psychological games" (Geanakoplos et al., 1989; Rabin, 1993), where preferences for physical outcomes are affected by a person's beliefs about the other's actions and discrepancies between these beliefs and the actual actions. However, a noteworthy aspect of the theoretical phenomena described in this paper is that they are evident already for values of r arbitrarily close to 0. This suggests that similar phenomena might also be exhibited by more complicated models where a person's utility has a first-order approximation, linear in the individuals' material payoffs, that is valid in the limit of weak altruism or spite (Levine, 1998). In this linear approximation, the coefficient r expresses the ratio between the *marginal* contributions to the person's utility of (i) the material payoff of each of the other individuals and (ii) the person's own material payoff.

The flip side of the limitations described above is that the simple linear model is readily extendible to internalization of any kind of social payoff, be it an index of social welfare, which depends only on the personal payoffs, a determinant of social welfare, such as the level of some public good, or any other variable that is determined by the individuals' actions. In this general setting, the common altruism coefficient r is the weight everyone attaches to the social payoff, and the weight attached to the personal payoffs is $1 - r$. Comparative statics examines the effect of changing preferences, in the form of increasing or decreasing r, on the actual level of social payoff.

As this paper shows, the question of whether altruism has a positive effect on the social payoff and spite a negative effect has a simple, affirmative answer only in the case of non-strategic interactions, in which the optimal action for each individual does not depend on the others' actions. In strategic interactions, or games, altruism and spite do not necessarily have the effects one would expect. For example, even in a symmetric two-player game with a unique, symmetric equilibrium, the players' personal payoff at equilibrium may be higher if they are both selfish rather than mildly caring, and even higher if they resent one another. Thus, altruism in a strategic interaction may paradoxically result in real, material losses for all parties.[1]

A central finding in this paper is that a crucial factor affecting the nature of comparative statics is the stability or instability of the strategies involved. In particular, in a symmetric setting, continuously increasing the weight that players place on each other's payoff can only increase the payoffs or leave them unchanged if the strategies involved are stable, but has the opposite effect if the strategies are definitely unstable (this term is defined below). This finding is akin to Samuelson's (1983) "correspondence principle", which maintains that conditions for stability often coincide with those under which comparative statics analysis leads to what are usually regarded as "normal" conclusions, such as the conclusion that an increase in demand for a commodity results in a rise in its equilibrium price (Lindbeck, 1992). Since comparative statics considers equilibria in different games, whereas stability is a property of equilibrium strategies in a specific game, the finding that the latter conveys information about the former is not all that obvious.[2]

In the works of Samuelson and others, 'stability' refers to dynamic, asymptotic stability. It therefore depends on the dynamical system used to model the evolution of the players' off-equilibrium behavior. By contrast, in this paper, 'stability' means static stability. This arguably more fundamental concept only considers the players' off-equilibrium *incentives*, and does not involve any assumptions about the translation of these incentives into concrete changes of actions. One example of a static notion of stability, applicable to symmetric $n \times n$ games, is evolutionarily stable strategy, or ESS. Another example, applicable to symmetric games with a unidimensional set of strategies, is continuously stable strategy, or CSS, which is essentially equivalent to the requirement that, at the equilibrium point, the graph of the best-response function, or reaction curve, intersects the forty-five degree line from above. If the intersection is from below, the symmetric equilibrium strategy is definitely unstable. These two examples of static stability are in fact essentially special cases of a general notion of static stability, proposed in Milchtaich (2012), which is applicable to any symmetric N-player game or population game with a non-discrete strategy space. The effects of altruism and spite on the social payoff turn out to be related to this general notion of static stability rather than to any special dynamic one.

The layout of the paper is as follows. The next section defines the modified game, which is the tool used in this paper to model internalization of a social payoff. The modified game has a single parameter, the altruism coefficient r, which expresses the degree of internalization. Section 3 presents the distinction between local comparative statics, which concern small, continuous changes to the altruism coefficient and the corresponding equilibria, and global comparative statics, which allow for large, discrete changes. Examples of the former are presented, which show that altruism can increase or decrease the social payoff, and the effect correlates with the stability or instability of the corresponding equilibrium strategies. A general definition of static stability of strategies in symmetric and population games is presented in Section 4. Section 5 gives a number of results that connect this notion of stability with comparative statics, first in a general setting and then for

---

[1] That altruism may theoretically lead to socially inefficient outcomes in asymmetric two-player strategic interactions, even if both individuals are equally altruistic towards each other, has long been recognized (Lindbeck and Weibull, 1988; Corts, 2006). That the same is true for symmetric games seems to be less well known (but see the remarks in Stark, 1989).

[2] Note that this refers to the stability of the equilibrium, and not to that of altruism itself, e.g., in the sense of Bester and Güth (1998).

specific classes of games such as symmetric $n \times n$ games. It also identifies the special form that static stability takes in each of these classes. Section 6 lays out a comparable analysis for global comparative statics. Section 7 considers both local and global comparative statics in asymmetric games. The relation between the static stability notion used here and dynamic stability is discussed in Section 8. This relation bears on the likeliness of "paradoxical" comparative statics. In particular, it is shown that stability with respect to the replicator dynamics does not prevent altruism from making everyone worse off. Appendix A presents a useful connection, which has an implication for comparative statics, between static stability in symmetric $n \times n$ games and stability with respect to perturbations of the game parameters.

2. Internalization of social payoff

Individuals in an interacting group may be unified in a desire to maximize some common social payoff that is determined by their choice of strategies, may only care about their own payoffs, or may give some weight to both goals. If the weights are the same for all individuals, each individual i may be viewed as seeking to maximize his *modified payoff*

$$h_i^r = (1 - r)h_i + rf, \tag{1}$$

where h_i is i's own, *personal* payoff, f is the social payoff, and $r \leqslant 1$ is the *altruism coefficient* (Levine, 1998) or coefficient of effective sympathy (Edgeworth, 1881), which quantifies the degree of internalization of the social payoff.[3]

The social payoff f in (1) can in general be any function of the players' strategies. An important special case is when f is expressible as a function of the personal payoffs. This is so in particular in the following simple and rather standard model of linearly interdependent preferences in a one-shot strategic interaction. The model is less general than, e.g., Levine's (1998) model, in which different individuals can be more or less altruistic and their attitudes are reflected in the ways others treat them. Here the players' preferences are all interdependent in the same manner. Specifically, the dependencies are expressible by a single parameter $r \leqslant 1$, which specifies the weight that each player i attaches to the payoff h_j of each of the other players j relative to the weight he attaches to his own payoff h_i. A positive r expresses concern for others' welfare, zero expresses total selfishness, and a negative r expresses envy or spite (Morgan et al., 2003). Thus, the expression player i seeks to maximize is

$$h_i^r = h_i + r \sum_{j \neq i} h_j. \tag{2}$$

Comparison with (1) shows that the social payoff f in this case is the aggregate payoff $\sum_j h_j$.

In this paper, the altruism coefficient is viewed as exogenously given, for example, a shared moral standard. Different values of r represent different groups rather than different kinds of individuals within a group, and varying the coefficient corresponds to a cross-group or cross-society comparison.[4]

An example of an interacting group as above is a (human or non-human) family. In the biological theory of kin selection, an expression similar to (2) gives the *inclusive fitness* of an individual interacting with relatives. In this case, the altruism coefficient r is the *coefficient of relatedness* between i and j, which is, for example, 0.5 for full siblings and 0.25 for half-siblings (Crow and Kimura, 1970). Thus, (2) expresses the inclusive fitness when the interaction involves only equally related individuals, e.g., siblings. In small populations, the coefficient of relatedness may also take on negative values, which represent less-than-average relatedness. In this case, the possibility of spiteful behavior arises (Hamilton, 1970).

The rest of this section concerns the internalization of social payoff in the special case of symmetric and population games. The general case of asymmetric games is not studied in detail in this paper. However, a partial analysis is presented in Section 7.

2.1. Symmetric and population games

In a *symmetric N-player game*, all players share the same strategy space X and their payoffs are specified by a single function $g : X^N \to \mathbb{R}$ that is invariant to permutations of its second through Nth arguments. If one player uses strategy x and the others use $y, z, \ldots, w$, in any order, the first player's payoff is $g(x, y, z, \ldots, w)$. A strategy y is a (symmetric Nash) *equilibrium strategy* in g, with the equilibrium payoff $g(y, y, \ldots, y)$, if it is a best response to itself: for every strategy x,

$$g(y, y, \ldots, y) \geqslant g(x, y, \ldots, y). \tag{3}$$

In the context of symmetric games, social payoffs $f : X^N \to \mathbb{R}$ are assumed symmetric functions, that is, invariant to permutations of their N variables. With an altruism coefficient $r \leqslant 1$, the modified payoff function is

---

[3] The upper limit of 1 means that the weight attached to the personal payoff is nonnegative. In some contexts, it may be natural or necessary to limit the altruism coefficient also from below, in particular by requiring $r > -1$. See footnote 12 and Section 6.

[4] The evolution and origin of altruism and spite are outside the scope of this paper. The model is not an evolutionary one, and it is not suitable for studying the effects that different individuals' attitudes towards others have on their own success. The model and corresponding comparative statics analysis may however have relevance for *group selection*. See Section 8.

$$g^r = (1 - r)g + rf. \tag{4}$$

If the strategy profile is symmetric, that is, all players use the same strategy, they necessarily also receive identical personal and modified payoffs. If in addition the social payoff f is the aggregate payoff, then it is simply N times the personal payoff.

A *population game*, as defined in this paper, is formally a symmetric two-player game such that X is a convex set in a (Hausdorff real) linear topological space (for example, the unit simplex in a Euclidean space) and $g(x, y)$ is continuous in y for all $x \in X$. However, the game is interpreted not as representing an interaction between two specific players but as one involving an (effectively) infinite population of individuals who are "playing the field". This means that an individual's payoff $g(x, y)$ depends only on his own strategy x and on the *population strategy y*. The latter may be, for example, the population's *mean* strategy with respect to some nonatomic measure, which attaches zero weight to each individual. In this case, the equilibrium condition,

$$g(y, y) = \max_{x \in X} g(x, y), \tag{5}$$

means that, in a *monomorphic* population, where everyone plays y, single individuals cannot increase their payoff by choosing any other strategy. Alternatively, a population game g may describe a dependence of an individual's payoff on the *distribution* of strategies in the population (Bomze and Pötscher, 1989), with the latter expressed by the population strategy y. In this case, X consists of *mixed strategies*, that is, probability measures on some underlying space of allowable actions or (pure) strategies, and $g(x, y)$ is linear in x and expresses the expected payoff for an individual whose choice of strategy is random with distribution x. Provided the space X is rich enough, the equilibrium condition (5) now means that the population strategy y is supported in the collection of all best response pure strategies. In other words, the (possibly) *polymorphic* population is in an equilibrium state.

Since players in a population game are individually insignificant in that they cannot affect the population strategy, internalization of a social payoff that only depends on the latter would be inconsequential if it meant, literally, consideration for the effect of ones' choice of strategy on the social payoff. However, internalization may change the players' behavior if it means consideration for the *marginal* effect on the social payoff (Chen and Kempe, 2008). To formalize this idea, suppose that the social payoff is given by a continuous function $\phi : \hat{X} \to \mathbb{R}$ whose domain $\hat{X}$ is the *cone* of the strategy space X, which consists of all elements of the form tx, with $x \in X$ and $t > 0$. Suppose also that ϕ has a directional derivative in every direction $\hat{x} \in \hat{X}$ and that the derivative depends continuously on the point $\hat{y} \in \hat{X}$ at which it is computed. In other words, the *differential $d\phi : \hat{X}^2 \to \mathbb{R}$* exists, where

$$d\phi(\hat{x}, \hat{y}) = \frac{d}{dt}\Big|_{t=0^+} \phi(t\hat{x} + \hat{y}),$$

and $d\phi$ is continuous in the second argument. The modified payoff can then be defined by setting

$$f = d\phi \tag{6}$$

in (4). Note that in the present context the social payoff is ϕ, not f, and that the latter is in general not a symmetric function.

An important example of a social payoff in a population game with a payoff function g (that satisfies the required technical conditions) is $\phi(y) = g(y, y)$, which represents the population's *mean payoff* (see Example 4 below). Another example is the production level of some public good (see Example 2).

3. Comparative statics

An increase or decrease in the altruism coefficient r means a change in the weight players attach to the social payoff relative to their personal payoff. This may result in a change of equilibrium strategies, and hence also of the social payoff. In the case of a symmetric game and a social payoff that is the aggregate payoff, changes in the latter precisely mirror changes in the players' personal payoffs.

The comparative statics of altruism in a symmetric or population game g may be examined either globally or locally. *Global comparative statics* compare the social payoff at the equilibria in the modified and unmodified games, g^r and g, or more generally, the equilibria in g^r and g^s, with $r \neq s$. The comparison is global in that it is not restricted to equilibria in g^r that are close to particular equilibria in g^s or to small changes in the altruism coefficient, that is, r close to s. The unrestricted nature of global comparative statics means that significant results can be obtained only in some cases. This paper considers such comparative statics mainly in the context of two-player games (Section 6). The paper's focus is on *local comparative statics*, which concern the way the social payoff at a given equilibrium in a given modified game g^s (or, in the special case $s = 0$, the unmodified game g) changes when the altruism coefficient continuously increases or decreases from s. For this comparison to be meaningful, a continuous function has to exist that assigns to every altruism coefficient r close to s an equilibrium in g^r, which coincides with the given equilibrium for $r = s$. This effectively rules out games with discrete strategy spaces, since if strategies are isolated, such a continuous function is necessarily constant. Assuming that the strategy spaces are non-discrete and that a function as above exists,[5] the question is whether the social

---

[5] For a sufficient condition for existence in the special case of symmetric $n \times n$ games, see Section 5.1.
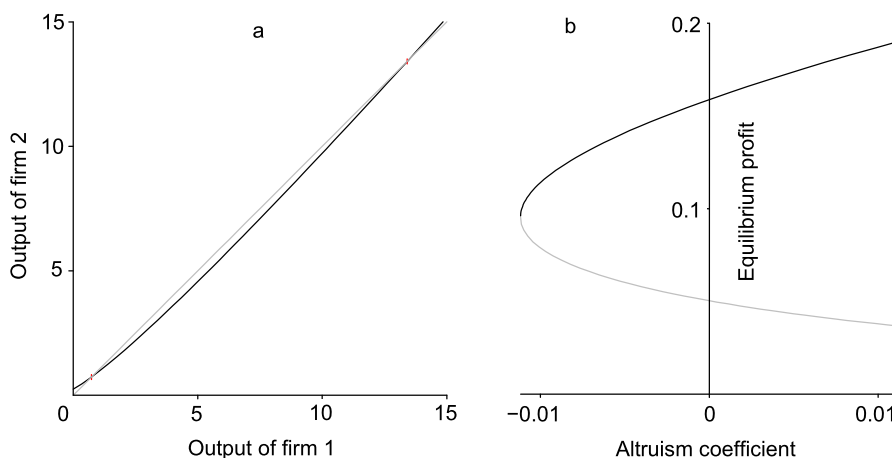
Fig. 1. The Cournot duopoly game in Example 1. (a) The reaction curve. For every output level for firm 1, the unique output level for firm 2 that maximizes the latter's modified payoff is shown for the altruism coefficient $r = 0.05$ (black curve). The two points, marked by short vertical lines, at which the curve intersects the forty-five degree line (grey) are the (symmetric) equilibria. The lower-left and upper-right points represent stable and unstable equilibrium output levels, respectively. When the altruism coefficient r changes, the two points move in opposite directions. (b) The equilibrium profits. The firms' equilibrium profit depends on the altruism coefficient r. It also depends on whether the equilibrium output level is stable or unstable. For the former (black, upper curve), the profit increases with increasing r, and for the latter (grey, lower curve), it decreases.

payoff increases or decreases with increasing (or decreasing) r. As the following example shows, even in a single game both possibilities may occur.

Example 1 *(Symmetric Cournot competition).* Firms 1 and 2 produce an identical good at zero cost. They simultaneously decide on their respective output levels q_1 and q_2 and face a downward sloping, convex demand curve given by the price (or inverse demand) function

$$P(Q) = \left[(Q + 0.4)\ln(Q + 1.4)\right]^{-\frac{3}{2}},$$

where $Q = q_1 + q_2$ is the total output. The profit of each firm i is its revenue $q_i P(Q)$. The social payoff is defined as the total revenue $Q P(Q)$. Thus, with an altruism coefficient r, firm i's modified payoff is

$$(q_i + r q_j)P(Q),$$

where j is the other firm. If the altruism coefficient r increases from 0 all the way to 1, the duopoly effectively becomes a monopoly and the firms' profits increase. However, this is not necessarily so for a small increase in r. For every output level of a firm, there is a unique, nonzero output level for its competitor that maximizes the latter's profit. The same is true for the modified payoff, for every r less than about 0.5 (see Fig. 1a). For r close to 0, there are precisely two equilibria, which are both symmetric, i.e., the firms' output levels, and hence also their profits, are equal. In one equilibrium, the output level is below 2, and in the other, it is above 2. The first equilibrium output level continuously decreases with increasing altruism coefficient, and consequently the profits increase. However, at the second equilibrium, increasing altruism has the opposite effects on the output and profit (Fig. 1b).

The modified payoffs of two competing firms may conceivably represent real profits for owners, e.g., if they own stock in both companies. However, in Example 1 and elsewhere in this paper, the modified payoffs are not *assumed* real entities. The question asked is how internalization of the social payoff affects it. The above example shows that the effect may go in both directions. The comparative statics seen in this example can be understood by examining the geometry of the reaction curve (Fig. 1a), which is the graph of the best-response function. Increasing the altruism coefficient r lowers the curve, since it amplifies the negative effect that a firm's output level, which adversely affects the competitor's profit, has on the firm's modified payoff. At the lower-output equilibrium, where the upward-sloping reaction curve is less steep than the forty-five degree line, lowering the curve moves the equilibrium point downwards and to the left. Thus, the equilibrium output decreases. At the higher-output equilibrium, where the reaction curve is steeper than the forty-five degree line, its shift has the opposite effect on the equilibrium output level.

Significantly, the same geometrical property of the reaction curve also determines whether the equilibrium output level is stable. It is stable if the reaction curve is less steep than the forty-five degree line and unstable (even definitely unstable, in a sense defined below) if the curve is steeper than the line. Thus, stable or definitely unstable equilibrium strategies entail "positive" or "negative" local comparative statics, respectively, in the sense of the effect of a continuously increasing altruism coefficient on the social payoff. A similar phenomenon occurs in the next two examples.

Example 2 *(Public good game).* An infinite population of identical players, represented by the unit interval $[0, 1]$, face the decision of what part $0 \leqslant x \leqslant 1$ of their unit endowment of private good they contribute for the production of some public good. The amount of public good produced is $\phi(y)$, where y is the aggregate contribution, which is the integral of the players' individual contributions (assuming they constitute an integrable function on $[0, 1]$) and $\phi : [0, \infty) \to \mathbb{R}$ is the twice continuously differentiable production function. The payoff of a player contributing x is given by

$$g(x, y) = \phi(y) - x. \tag{7}$$

Suppose that contribution of public good is socially desirable, that is, $\phi' > 1$. This means that players might have an incentive to contribute if they cared enough about the effect on the other players. However, since individual players cannot possibly affect the population strategy (that is, the aggregate contribution) y, for a modification of the personal payoff (7) to be effective it should involve the *marginal* product

$$d\phi(x, y) = \frac{d}{dt}\bigg|_{t=0^+} \phi(tx + y) = x\phi'(y). \tag{8}$$

Thus, with an altruism coefficient $0 < r < 1$, the modified payoff is

$$g^r(x, y) = (1 - r)(\phi(y) - x) + rx\phi'(y). \tag{9}$$

Since this expression is linear in the player's own contribution x, if $\phi'(0) \leqslant 1/r - 1$ or $\phi'(1) \geqslant 1/r - 1$ then contributing 0 or 1, respectively, is an equilibrium strategy. An intermediate value $0 < y^r < 1$ is an equilibrium strategy if and only if $x = y^r$ is a solution of

$$\phi'(x) = \frac{1}{r} - 1. \tag{10}$$

Assuming $\phi''(y^r) \neq 0$, there is in this case a unique solution in a neighborhood of y^r of an equation similar to (10) in which r is replaced by any close altruism coefficient, and the relation between the altruism coefficient and the corresponding equilibrium strategy satisfies

$$\frac{dy^r}{dr} = -\frac{1}{r^2 \phi''(y^r)}. \tag{11}$$

Hence,

$$\frac{d}{dr} \phi(y^r) = -\frac{1 - r}{r^3 \phi''(y^r)}. \tag{12}$$

This shows that whether internalization of the (positive) social effect of contribution of private good actually results in higher aggregate contribution and level of public good depends on whether the production function is concave ($\phi'' < 0$) or convex ($\phi'' > 0$) in a neighborhood of the equilibrium strategy y. In the former case, internalization has the above effect, but in the latter, it paradoxically has the opposite, negative effect on the production of public good and hence on social welfare.

Significantly, concavity or convexity of the production function also determines the effect of small changes in the population strategy on the players' incentives. If, for example, the aggregate contribution y slightly increases above the equilibrium level y^r (which is the solution of (10)), then the coefficient of x in (9) becomes negative or positive if $\phi''(y^r)$ has that sign. In the former case, this creates an incentive for players to cut down their contribution of private good, and in the latter, the incentive is to (further) increase it. Thus, concavity or convexity of ϕ may be interpreted as entailing stability or instability, respectively, of the equilibrium strategy.

Example 3 *(Generalized rock–scissors–paper games).* A symmetric 3×3 game g has the following payoff matrix A:

$$\begin{pmatrix} 0 & -3 & 2 \\ 6 & 0 & -2 \\ -1 & 3 & 0 \end{pmatrix}. \tag{13}$$

The social payoff f is the aggregate payoff, which is given by the payoff matrix $A + A^{\mathrm{T}}$. For $-1 < r < 0.5$, the modified game g^r has a unique equilibrium, which is symmetric and completely mixed. The corresponding personal payoff (which is half the social payoff) is equal to $(30(1 - r)^2 + 3r)/(73(1 - r)^2 + 3r)$, which is a positive and continuously increasing function of r in the above interval. The same is true with A replaced by $-A$, except that in this case the equilibrium payoff has the opposite sign and it hence decreases with increasing r. Thus, altruism, or internalization of the other player's payoff, has the opposite effect on the player's personal payoffs in these two games. As in the previous two examples, comparative statics align with stability. It can be shown (see Section 5.1) that with the payoff matrix A all the equilibrium strategies involved are evolutionarily stable (ESSs), and with $-A$, they are unstable (indeed, definitely evolutionarily unstable; see below).

As this paper shows, the connection seen in these examples between stability and comparative statics is in fact a very general phenomenon. This connection is laid out in several steps in the following sections. The first step is presentation of a suitable general notion of static stability in symmetric and population games. This notion, introduced in Milchtaich (2012), unifies and generalizes several well-established stability notions that are specific to particular classes of games, such as evolutionary stability and the geometrical criterion for stability used in Example 1.

4. Stability in symmetric and population games

Inequality (3) in the equilibrium condition expresses a player's lack of incentive to be the first to move from strategy y to an alternative strategy x. Stability, as defined below, adopts a broader perspective: it takes into consideration the incentive to be the jth player to move from y to x, for all j between 1 and N, the number of players in the game. These incentives are all given the same weight. Thus, stability requires that, when the players move one-by-one from y to x, the corresponding changes of payoff are negative *on average*.

Definition 1. A strategy y in a symmetric N-player game $g : X^N \to \mathbb{R}$ is *stable*, *weakly stable* or *definitely unstable* if it has a neighborhood where, for every strategy $x \neq y$, the inequality

$$\frac{1}{N} \sum_{j=1}^{N} \big(g(x, \underbrace{x, \ldots, x}_{j-1 \text{ times}}, \underbrace{y, \ldots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \ldots, x}_{j-1 \text{ times}}, \underbrace{y, \ldots, y}_{N-j \text{ times}}) \big) < 0, \tag{14}$$

a similar weak inequality or the reverse (strict) inequality, respectively, holds.

Inequality (14) may also be written, more symmetrically, as

$$\frac{1}{N} \sum_{j=1}^{N} \big(g(\underbrace{x, \ldots, x}_{j \text{ times}}, y, \ldots, y) - g(\underbrace{y, \ldots, y}_{j \text{ times}}, x, \ldots, x) \big) < 0. \tag{15}$$

Stability, as defined here, is a local concept. It refers to neighborhood systems of strategies or equivalently to a topology on the strategy space X. The topology may be explicitly specified or it may be understood from the context. The latter applies when X may be naturally viewed as a subset of a Euclidean space, e.g., a simplex or an interval in the real line. In this case, the default topology on X is the relative one, so that proximity between strategies is measured by the Euclidean distance between them.

In some classes of games (see Sections 5.1 and 5.3), stability of a strategy automatically implies that it is an equilibrium strategy. However, in general, neither of these conditions implies the other. The difference is partially due to equilibrium being a global condition: all alternative strategies, not only neighboring ones, are considered. However, it persists if 'equilibrium' is replaced by 'local equilibrium', with the obvious meaning. A *stable equilibrium strategy* is a strategy that satisfies both conditions. In a symmetric two-player game, where (14) can be written as

$$\frac{1}{2} \big(g(x, x) - g(y, x) + g(x, y) - g(y, y) \big) < 0, \tag{16}$$

a strategy y is a stable equilibrium strategy if and only if it has a neighborhood where for every $x \neq y$ the inequality

$$pg(x, x) + (1 - p)g(x, y) < pg(y, x) + (1 - p)g(y, y)$$

holds for all $0 < p \leqslant 1/2$. This condition means that x affords a lower expected payoff than y against an uncertain strategy that may be x or y, with the former no more likely than the latter.

Stability in population games may be defined by a variant of Definition 1 that replaces the number of players using strategy x or y with the size of the subpopulation to which the strategy applies, p or $1 - p$ respectively. Correspondingly, the sum in (14) is replaced with an integral.

Definition 2. A strategy y in a population game $g : X^2 \to \mathbb{R}$ is *stable*, *weakly stable* or *definitely unstable* if it has a neighborhood where, for every strategy $x \neq y$, the inequality

$$\int_0^1 \big(g(x, px + (1 - p)y) - g(y, px + (1 - p)y) \big) \, dp < 0, \tag{17}$$

a similar weak inequality or the reverse (strict) inequality, respectively, holds.

For some games, Definitions 1 and 2 are both potentially applicable, depending on whether the game is viewed as a symmetric or as a population game. An important example of this is symmetric $n \times n$ games (Section 5.1). However, it is not difficult to see that, in this particular example (and similar ones; see Milchtaich, 2012), the point of view is immaterial: the two definitions are equivalent.

5. Stability and local comparative statics

The following theorem is the basic local comparative statics result for symmetric games, for which stability is given by Definition 1. Both the definition and the theorem refer to a topology on the strategy space X, which has to be the same one. As indicated, in many games only one topology is natural or interesting. However, the result holds for *any* topology on X.

Theorem 1. *For a symmetric N-player game $g : X^N \to \mathbb{R}$ and a social payoff function $f : X^N \to \mathbb{R}$ that are both Borel measurable,[6] and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leqslant 1$, suppose that there is a continuous and finitely-many-to-one[7] function that assigns a strategy y^r to each $r_0 \leqslant r \leqslant r_1$ such that the function $\pi : [r_0, r_1] \to \mathbb{R}$ defined by*

$$\pi(r) = \frac{1}{N} f(y^r, y^r, \dots, y^r) \tag{18}$$

is absolutely continuous.[8] If, for every $r_0 < r < r_1$, y^r is a stable, weakly stable or definitely unstable strategy in the modified game g^r, then π is strictly increasing, nondecreasing or strictly decreasing, respectively.

Proof. The proof uses the following identity, which holds for all r, s and strategies x and y:

$$(r - s) \frac{1}{N} \big(f(x, x, \dots, x) - f(y, y, \dots, y) \big)$$

$$= (1 - s) \frac{1}{N} \sum_{j=1}^{N} \big(g^r(\underbrace{x, \dots, x}_{j \text{ times}}, y, \dots, y) - g^r(\underbrace{y, \dots, y}_{j \text{ times}}, x, \dots, x) \big)$$

$$+ (1 - r) \frac{1}{N} \sum_{j=1}^{N} \big(g^s(\underbrace{y, \dots, y}_{j \text{ times}}, x, \dots, x) - g^s(\underbrace{x, \dots, x}_{j \text{ times}}, y, \dots, y) \big). \tag{19}$$

For $-(r_1 - r_0)/2 < \epsilon < (r_1 - r_0)/2$, consider the (Borel) set $U_\epsilon \subseteq [r_0, r_1]$ defined by

$$U_\epsilon = \left\{ r_0 + |\epsilon| \leqslant r \leqslant r_1 - |\epsilon| \; \Big| \; \sum_{j=1}^{N} \big(g^r(\underbrace{y^{r+\epsilon}, \dots, y^{r+\epsilon}}_{j \text{ times}}, y^r, \dots, y^r) - g^r(\underbrace{y^r, \dots, y^r}_{j \text{ times}}, y^{r+\epsilon}, \dots, y^{r+\epsilon}) \big) < 0 \right\}. \tag{20}$$

Suppose that, for every $r_0 < r < r_1$, y^r is a stable strategy in g^r. Since the function $r \mapsto y^r$ is finitely-many-to-one, for every $r_0 < r < r_1$ the inequality $y^s \neq y^r$ holds for all $s \neq r$ in some neighborhood of r. Hence, for $\epsilon \neq 0$ sufficiently close to 0 (including negative ϵ), $r \in U_\epsilon$ (cf. (15)). It follows that the Lebesgue measure of U_ϵ tends to $r_1 - r_0$ as ϵ tends to 0. The same clearly holds for the set $U_{-\epsilon}$, hence also for $U_{-\epsilon} - \epsilon = \{r - \epsilon \mid r \in U_{-\epsilon}\} \subseteq [r_0, r_1]$ (which is obtained from $U_{-\epsilon}$ by translation, and thus has the same measure), and therefore also for the set

$$V_\epsilon = U_\epsilon \cap (U_{-\epsilon} - \epsilon).$$

Therefore, for (Lebesgue-)almost every $r_0 < r < r_1$, the relation $r \in V_{1/k}$ holds for infinitely many positive integers k. For each r and k satisfying this relation, the strict inequality in (20) holds for $\epsilon = 1/k$ (since $r \in U_{1/k}$), and a similar inequality holds for $\epsilon = -1/k$ with r replaced by $r + 1/k$ (since $r + 1/k \in U_{-1/k}$). If k is sufficiently large (specifically, greater than $1/(1 - r)$), the two inequalities together imply that

$$\left(1 - r - \frac{1}{k} \right) \sum_{j=1}^{N} \big(g^r(\underbrace{y^{r+\frac{1}{k}}, \dots, y^{r+\frac{1}{k}}}_{j \text{ times}}, y^r, \dots, y^r) - g^r(\underbrace{y^r, \dots, y^r}_{j \text{ times}}, y^{r+\frac{1}{k}}, \dots, y^{r+\frac{1}{k}}) \big)$$

$$+ (1 - r) \sum_{j=1}^{N} \big(g^{r+\frac{1}{k}}(\underbrace{y^r, \dots, y^r}_{j \text{ times}}, y^{r+\frac{1}{k}}, \dots, y^{r+\frac{1}{k}}) - g^{r+\frac{1}{k}}(\underbrace{y^{r+\frac{1}{k}}, \dots, y^{r+\frac{1}{k}}}_{j \text{ times}}, y^r, \dots, y^r) \big) < 0. \tag{21}$$

---

[6] Borel measurability of a function means that the inverse image of every open set is a Borel set (Rana, 2002, Ex. 7.3.13). Every continuous function is Borel measurable.

[7] A function is finitely-many-to-one if the inverse image of every point is a finite set.

[8] A sufficient condition for absolute continuity is that the function is continuously differentiable.

It follows from (18) and the identity (19), applied to $s = r + 1/k$, $y = y^r$ and $x = y^{r+1/k}$, that the expression on the left-hand side of (21) is equal to $-(1/k)(\pi(r+1/k) - \pi(r))$. Thus, (21) gives

$$\pi\left(r + \frac{1}{k}\right) > \pi(r). \tag{22}$$

If π is differentiable at r and (22) holds for infinitely many k's, then $\pi'(r) \geqslant 0$. Since π, being an absolutely continuous function, is differentiable almost everywhere in $[r_0, r_1]$ and satisfies

$$\pi(s) = \pi(r_0) + \int_{r_0}^{s} \pi'(r)\,dr, \quad r_0 \leqslant s \leqslant r_1$$

(Yeh, 2006, Theorem 13.17), this proves that π is nondecreasing. To prove that it is in fact strictly increasing it suffices to show that (r_0, r_1) has no open subinterval where π is constant. Any point r lying in such a subinterval satisfies $\pi(r+1/k) - \pi(r) = 0$ for all large enough k, which implies that $r \notin V_{1/k}$ for all such k. Since it is proved above that the latter does *not* hold for almost all $r \in (r_0, r_1)$, this proves that a subinterval as above does not exist.

If, for every $r_0 < r < r_1$, y^r is a weakly stable strategy in g^r, then the conclusion that π is nondecreasing still holds, and the only change required in the above proof is changing the strict inequalities in (20), (21) and (22) to weak ones. If each y^r is definitely unstable, then a proof very similar to that above shows that π is strictly decreasing. The only change required here is reversing the strict inequalities in (20), (21) and (22). $\quad\square$

If each of the strategies y^r in Theorem 1 is an equilibrium strategy in the corresponding modified game g^r (which the theorem does not require) and the social payoff f is the aggregate payoff, then, by (18), $\pi(r)$ gives the players' personal payoff at the equilibrium of the modified game. Thus, in this case, an increase in $\pi(r)$ spells (uniform) Pareto improvement. Theorem 1 can therefore be interpreted as saying that, under certain technical assumptions, stability of the equilibrium strategies guarantees that all players benefit from gradually becoming more altruistic, if the change in preferences is simultaneous and to the same degree for all players and it affects the equilibrium strategy. If the equilibrium strategies are definitely unstable, altruism has the opposite effect on the players' personal payoff.

Changes in the altruism coefficient may also leave the social payoff unchanged. A trivial example of this is when there are no personal payoffs, $g = 0$. In Theorem 1, ineffective change of the altruism coefficient is excluded by the assumption that the assignment of an equilibrium strategy y^r to each r is finitely-many-one-to-one. The assignment is also assumed to be continuous, which means that (unlike for global comparative statics; see below) two equilibrium strategies can be compared only if they are connected in the strategy space by a curve whose points are equilibrium strategies for intermediate values of the altruism coefficient. In the case of multiple equilibria, this guarantees that an equilibrium in one modified game is compared with the "right", or corresponding, equilibrium in the other game. Even with all of these assumptions, stability or definite instability are not *necessary* conditions for the social payoff to increase or decrease, respectively, with increasing altruism. It is shown in Section 5.1 that, if the equilibrium strategies are neither stable nor definitely unstable, both kinds of comparative statics are possible.

A similar result to Theorem 1 holds for population games, for which the meaning of stability is given by Definition 2. The choice of topology is restricted in this case by the definition of population game, which requires X to be a (convex) set in a linear topological space such that the continuity conditions spelled out in Section 2.1 hold for the game and the social payoff.

Theorem 2. *For a population game $g : X \times X \to \mathbb{R}$ and a social payoff $\phi : \hat{X} \to \mathbb{R}$ such that both g and $d\phi$ are Borel measurable, and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leqslant 1$, suppose that there is a continuous and finitely-many-to-one function that assigns a strategy y^r to each $r_0 \leqslant r \leqslant r_1$ such that the function $\pi : [r_0, r_1] \to \mathbb{R}$ defined by*

$$\pi(r) = \phi(y^r) \tag{23}$$

is absolutely continuous. If, for every $r_0 < r < r_1$, y^r is a stable, weakly stable or definitely unstable strategy in the modified game g^r, then π is strictly increasing, nondecreasing or strictly decreasing, respectively.

Proof. A key difference between population games and symmetric two-player games is that, for the former, stability and instability in g^r are defined by the sign of the expression obtained by substituting g^r for g in the left-hand side of (17) rather than (16). However, the first expression can be given a form similar to the second one, namely,

$$\frac{1}{2}\left(\tilde{g}^r(x, x) - \tilde{g}^r(y, x) + \tilde{g}^r(x, y) - \tilde{g}^r(y, y)\right),$$

by defining the function $\tilde{g}^r$ as

$$\tilde{g}^r(x, y) = 2 \int_0^1 g^r(x, px + (1 - p)y) \, dp - g^r(x, x). \tag{24}$$

This formal similarity means that almost the entire proof of Theorem 1 can be reused. The only missing part is a proof of an identity similar to (19), with $\tilde{g}$ substituted for g on the right-hand side (where $N = 2$) and the left-hand side replaced by $(r - s)(\phi(x) - \phi(y))$. By (24), the above substitution of $\tilde{g}$ for g gives

$$(1 - s) \int_0^1 \left(g^r(x, px + (1 - p)y) - g^r(y, px + (1 - p)y) \right) dp$$

$$+ (1 - r) \int_0^1 \left(g^s(y, py + (1 - p)x) - g^s(x, py + (1 - p)x) \right) dp.$$

Since $g^r = (1 - r)g + r \, d\phi$ and a similar equation holds for s, this expression simplifies to

$$(r - s) \int_0^1 \left(d\phi(x, px + (1 - p)y) - d\phi(y, px + (1 - p)y) \right) dp. \tag{25}$$

For all $0 < p < 1$, the integrand in (25) is equal to

$$\frac{d}{dp} \phi(px + (1 - p)y)$$

(Milchtaich, 2012, Section 8). Therefore, (25) is equal to $(r - s)(\phi(x) - \phi(y))$.   □

When considering particular classes of symmetric and population games, more special versions of Theorems 1 and 2 can be obtained by replacing the general stability or definite instability condition with a condition that is equivalent to, or at least implies, that property in the class under consideration. Unlike Definitions 1 and 2, these "native" notions of stability or instability may depend on structures and special properties of the game that are not necessarily present in other classes. Some of the better-known classes of games are considered in the following subsections.

5.1. Symmetric $n \times n$ games

An important class of symmetric two-player games, which may also be viewed as population games, is symmetric $n \times n$ games. In these games, both players share a common finite set of n actions, and a (mixed) strategy $x = (x_1, x_2, \ldots, x_n)$ specifies the probability x_i with which a player chooses the ith action, for $i = 1, 2, \ldots, n$. The set of all actions i with $x_i > 0$ is the *support* of x. A strategy is *pure* or *completely mixed*, respectively, if its support contains only a single action i (in which case the strategy itself may also be denoted by i) or all n actions. The payoff function g in a symmetric $n \times n$ game is bilinear and is hence completely specified by the $n \times n$ *payoff matrix* $A = (g(i, j))_{i,j=1}^n$. With strategies viewed as column vectors (and superscript T denoting transpose),

$$g(x, y) = x^{\mathrm{T}} A y.$$

A standard notion of stability in a symmetric $n \times n$ game g is evolutionary stability (Maynard Smith, 1982). A strategy y is an *evolutionarily stable strategy* (ESS) or a *neutrally stable strategy* (NSS) if, for every strategy $x \neq y$, for sufficiently small $\epsilon > 0$ the inequality

$$g(y, \epsilon x + (1 - \epsilon)y) > g(x, \epsilon x + (1 - \epsilon)y)$$

or a similar weak inequality, respectively, holds. A completely mixed equilibrium strategy y is *definitely evolutionarily unstable* (Weissing, 1991) if $g(y, x) < g(x, x)$ for all $x \neq y$. As the following theorem shows, evolutionary stability and instability are equivalent to the corresponding notions in Definitions 1 and 2 (which, as indicated, are equivalent for symmetric $n \times n$ games).

Theorem 3. *(See Milchtaich, 2012.) A strategy in a symmetric $n \times n$ game g is stable or weakly stable if and only if it is an ESS or an NSS, respectively. A completely mixed equilibrium strategy is definitely unstable if and only if it is definitely evolutionarily unstable.*

A characterization of stability and weak stability in the modified game is obtained from Theorem 3 by replacing g with g^r. Alternatively, stability in g^r may be expressible directly in terms of the unmodified payoffs. Such an expression is

given by the following proposition, which concerns a positive level of altruism when the social payoff f is the aggregate payoff. As it shows, a stable strategy y in the modified game is characterized by the property that it affords a higher expected personal payoff than any other strategy close to it for a player whose opponent either mimics him and uses whatever (mixed) strategy x he uses, or uses strategy y; the former with probability r and the latter with probability $1 - r$. This characterization is somewhat similar to Myerson et al.'s (1991) notion of δ-viscous equilibrium. The main difference is that the latter only takes into consideration alternative pure strategies. In particular, if y itself is pure, and inequality (26) below holds for every pure strategy $x \neq y$, then y is a δ-viscous equilibrium for $\delta = r$, but it is not necessarily even an equilibrium strategy in g^r.

Proposition 1. *(See Milchtaich, 2006.) For a symmetric $n \times n$ game g, a social payoff that is the aggregate payoff, and an altruism coefficient $0 < r \leqslant 1$, a strategy y is an ESS or an NSS in the modified game g^r if and only if the inequality*

$$g(y, y) > g(x, rx + (1 - r)y) \tag{26}$$

or a similar weak inequality, respectively, holds for all strategies $x \neq y$ in some neighborhood of y.

Theorem 3 and Proposition 1 give the stability condition in Theorem 1 a concrete, special meaning. (For another, somewhat simpler, special version of that theorem, see Milchtaich, 2006.) As indicated (see Section 3), local comparative statics as in Theorem 1 are meaningful only if it is possible to continuously map altruism coefficients to equilibrium strategies in the corresponding modified games. As it turns out, for symmetric $n \times n$ games the existence of such a mapping is automatically guaranteed by an only slightly stronger stability condition. An ESS y is said to be a *regular* ESS if every action that is a best response to y is in its support, in other words, if (y, y) is a quasi-strict equilibrium (van Damme, 1991). Proposition A.1 in Appendix A immediately gives the following.

Corollary 1. *Let g be a symmetric $n \times n$ game, f a bilinear social payoff (for example, the aggregate payoff) and s an altruism coefficient. For every regular ESS y^s in g^s there is a continuous function that assigns to each altruism coefficient r in a neighborhood of s a regular ESS y^r in the game g^r, which is moreover the unique equilibrium strategy in g^r with the same support as y^s.*

As an example of the application of Theorem 1 to symmetric $n \times n$ games, consider the game g with the payoff matrix A given by (13). With f that is the aggregate payoff and $-1 < r < 0.5$, the modified game g^r is a generalized rock–scissors–paper game (that is, its payoff matrix has the same sign pattern as in the familiar, simple game). It hence has a unique equilibrium, which is symmetric and completely mixed (Hofbauer and Sigmund, 1998). Straightforward computation shows that the three coordinates of the equilibrium strategy y^r are determined by the altruism coefficient r as non-constant rational functions. Replacing A by $-A$ clearly does not change the equilibrium strategy but only reverses the sign of the equilibrium payoff. As indicated, this replacement also turns the equilibrium personal payoff from an increasing function of r to a decreasing function. This finding is accounted for by Theorem 1. The unique equilibrium strategy in a generalized rock–scissors–paper game with payoff matrix $B = (b_{ij})_{i,j=1}^3$ is evolutionarily stable or definitely evolutionarily unstable if and only if (i) the sum $c_{ij} = b_{ij} + b_{ji}$ is positive or negative, respectively, for all $1 \leqslant i < j \leqslant 3$ and (ii) the three numbers $\sqrt{|c_{12}|}$, $\sqrt{|c_{13}|}$, $\sqrt{|c_{23}|}$ correspond to the lengths of the sides of a triangle (Weissing, 1991, Theorem 4.6). The condition for stability clearly holds for the game with the payoff matrix (13) and for the corresponding modified games, and the condition for definite instability holds when the matrix is replaced by its negative.

Similar examples show that stability or definite instability are not *necessary* conditions for positive or negative comparative statics, respectively.[9] With the payoff matrix A given by

$$\begin{pmatrix} 0 & 3 & -2 \\ -2 & 0 & 2 \\ 1 & -1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 3 & -2 \\ -2 & 0 & 2 \\ 1 & -3 & 0 \end{pmatrix}, \tag{27}$$

and with f defined as the aggregate payoff, the modified game g^r is again a generalized rock–scissors–paper game for all $-1 < r < 0.5$. For the left payoff matrix, at equilibrium the personal payoff is equal to $(2(1 - r)^2 - r)/(29(1 - r)^2 - 5r)$, and it is hence determined by r as a (positive and) decreasing function. For the right matrix, the corresponding payoff is $-(6(1 - r)^2 - r)/(41(1 - r)^2 - 5r)$, and it is hence (negative and) increasing. However, since for both matrices $a_{12} + a_{21} > 0$ but $a_{13} + a_{31} < 0$, none of the (completely mixed) equilibrium strategies involved is stable or definitely unstable.

These examples raise the question of whether, for symmetric $n \times n$ games, with the social payoff defined as the aggregate payoff, and with $-1 < r \leqslant 1$, the personal (equivalently, social) payoff at the completely mixed equilibrium in the modified game is always a monotonic (either nondecreasing or nonincreasing) function of the altruism coefficient. For 2×2 games, 3×3 games with a non-singular payoff matrix, and symmetric $n \times n$ games that are potential games (see Section 5.3), it can be shown that the answer is affirmative. However, this is not so in general. For example, in the 4×4 game g with the (non-singular) payoff matrix

---

[9] However, these conditions are close to being necessary in the special case of symmetric 2×2 games. See Section 6.1.

$$\begin{pmatrix} 2 & 5 & 1 & 0 \\ -7 & -2 & 9 & 8 \\ -3 & 7 & 9 & -9 \\ 9 & 2 & -4 & -5 \end{pmatrix},$$

for which the modified game g^r has a unique completely mixed equilibrium strategy for every $0 \leqslant r \leqslant 1$, the personal payoff at equilibrium decreases for $0 < r < 0.584$ but increases for $0.584 < r < 1$.

5.2. Games with a unidimensional strategy space

In a symmetric two-player game or population game g in which the strategy space is an interval in the real line $\mathbb{R}$, stability or instability of an equilibrium strategy, in the sense of either Definition 1 or 2, has a simple, familiar meaning. If g is twice continuously differentiable, and with the possible exception of certain borderline cases, an equilibrium strategy is stable or definitely unstable if, at the (symmetric) equilibrium point, the reaction curve intersects the forty-five degree line from above or below, respectively. Stability is also essentially equivalent to the notion of continuously stable strategy, or CSS (Eshel and Motro, 1981; Eshel, 1983).

Theorem 4. *(See Milchtaich, 2012.) Let g be a symmetric two-player game or population game with a strategy space X that is a (finite or infinite) interval in the real line, and y an equilibrium strategy lying in the interior of X such that g has continuous second-order partial derivatives*[10] *in a neighborhood of the equilibrium point (y, y). If*

$$g_{11}(y, y) + g_{12}(y, y) < 0, \tag{28}$$

then y is stable and a CSS. If the reverse inequality holds, then y is definitely unstable and not a CSS.

The reaction curve is the collection of all strategy profiles (x, y) such that y is a best response to x. If (y, y) is an interior equilibrium as in Theorem 4, then (the second-order maximization condition) $g_{11}(y, y) \leqslant 0$ holds, since y is a best response to itself. If the inequality is strict, then (28) can be written as

$$-\frac{g_{12}(y, y)}{g_{11}(y, y)} < 1.$$

This inequality or the reverse one, respectively, says that at the equilibrium point the slope of the reaction curve is less or greater than the slope of the forty-five degree line (which equals 1). The former and the latter, respectively, holds for the low- and the high-output equilibria of the modified games in Example 1 (see Fig. 1a). Therefore, the comparative statics in that example are accounted for by Theorem 1.

In Example 2, Eq. (9) gives $g_{11}^r = 0$ and $g_{12}^r(x, y) = r\phi''(y)$. Therefore, the inequality obtained by substituting g^r for g in (28) or the reverse inequality holds if the production function ϕ is concave or convex, respectively. This shows that the comparative statics in Example 2 are accounted for by Theorem 2.

The following proposition, while much more special than Theorems 1 and 2, goes beyond them by pointing to a direct *quantitative* connection between comparative statics and an expression similar to that on the left-hand side of (28). In particular, in Example 2, in conjunction with (11) the proposition gives (12).

Proposition 2. *For a symmetric two-player game or population game g with a strategy space X that is an interval (of real numbers), a corresponding social payoff function, and altruism coefficients r_0 and r_1 with $r_0 < r_1$, suppose that there is a continuously differentiable function that assigns an equilibrium strategy y^r in the modified game g^r to each $r_0 < r < r_1$ such that y^r lies in the interior of X and both g and f have continuous second-order partial derivatives in a neighborhood of (y^r, y^r). Then, at each point $r_0 < r < r_1$,*

$$\frac{d\pi}{dr} = -(1-r)\big(g_{11}^r(y^r, y^r) + g_{12}^r(y^r, y^r)\big)\left(\frac{dy^r}{dr}\right)^2, \tag{29}$$

where π is defined by (18) if g is a symmetric two-player game and by (23) if it is a population game.

Proof. Since y^r is an interior equilibrium strategy in g^r for every $r_0 < r < r_1$, it satisfies the first-order maximization condition

$$g_1^r(y^r, y^r) = 0. \tag{30}$$

Since

$$g_1^r = (1-r)g_1 + rf_1, \tag{31}$$

---

[10] Partial derivatives are denoted by subscripts. For example, g_{12} is the mixed second-order partial derivative of g.

differentiation of both sides of (30) with respect to r and multiplication by $(1 - r)\, dy^r/dr$ give

$$(1 - r)\big(-g_1(y^r, y^r) + f_1(y^r, y^r)\big)\frac{dy^r}{dr} + (1 - r)\big(g^r_{11}(y^r, y^r) + g^r_{12}(y^r, y^r)\big)\left(\frac{dy^r}{dr}\right)^2 = 0. \tag{32}$$

By (30) and (31), the first term in (32) is equal to $f_1(y^r, y^r)\, dy^r/dr$. If g is a symmetric two-player game, then the social payoff f is by definition symmetric, and therefore the last expression is equal to $(1/2)\, d/dr\, f(y^r, y^r)$. If g is a population game, then f is connected with the social payoff ϕ by (6), which by (8) gives

$$f(x, y) = x\phi'(y), \quad x, y \in X.$$

This equality implies that $f_1(y^r, y^r)\, dy^r/dr = \phi'(y^r)\, dy^r/dr = d/dr\, \phi(y^r)$. Therefore, for both kinds of games, (32) gives (29). □

5.3. Potential games

A symmetric N-player game $g : X^N \to \mathbb{R}$ is called an (exact) potential game if it has an (exact) *potential*, which is a symmetric function $F : X^N \to \mathbb{R}$ such that, whenever a single player changes his strategy, the change in the player's payoff is equal to the change in F. Thus, for any $N + 1$ (not necessarily distinct) strategies $x, y, z, \ldots, w$,

$$F(x, z, \ldots, w) - F(y, z, \ldots, w) = g(x, z, \ldots, w) - g(y, z, \ldots, w). \tag{33}$$

The potential F may itself be viewed as a symmetric N-player game, indeed, a doubly symmetric one.[11] It follows immediately from (33) that F and g have exactly the same equilibrium strategies, stable and weakly stable strategies, and definitely unstable strategies. Stability and instability have in this case a strikingly simple characterization, which follows immediately from the observation that the sum in (14) is equal to the difference $F(x, x, \ldots, x) - F(y, y, \ldots, y)$ divided by N.

Theorem 5. *In a symmetric N-player game with a potential F, a strategy y is stable, weakly stable or definitely unstable if and only if it is a strict local maximum point, a local maximum point or a strict local minimum point, respectively, of the function $x \mapsto F(x, x, \ldots, x)$. If $(y, y, \ldots, y)$ is a global maximum point of F itself, then y is in addition an equilibrium strategy.*

For population games, which represent interactions involving many players whose individual actions have negligible effects on the other players, the definition of potential may be naturally adapted by replacing the difference on the left-hand side of (33) with a derivative.

Definition 3. For a population game $g : X^2 \to \mathbb{R}$, a continuous function $\Phi : X \to \mathbb{R}$ is a *potential* if for all $x, y \in X$ and $0 < p < 1$ the following derivative exists and satisfies the equality:

$$\frac{d}{dp}\Phi\big(px + (1 - p)y\big) = g\big(x, px + (1 - p)y\big) - g\big(y, px + (1 - p)y\big).$$

As for symmetric games, stability and instability (here, in the sense of Definition 2) of a strategy y in a population game with a potential Φ is related to y being a local extremum point of the potential.

Theorem 6. *(See Milchtaich, 2012.) In a population game g with a potential Φ, a strategy y is stable, weakly stable or definitely unstable if and only if it is a strict local maximum point of Φ, a local maximum point of Φ or a strict local minimum point of Φ, respectively. In the first two cases, y is in addition an equilibrium strategy. If the potential Φ is strictly concave, an equilibrium strategy is a strict global maximum point of Φ, and necessarily the game's unique stable strategy.*

A sufficient condition for a continuous function Φ to be a potential for a population game g is that its differential (see Section 2.1) is equal to g. More precisely, this condition is relevant if Φ is defined on (or is extendable to) $\hat{X}$, the cone of the strategy space X.

Proposition 3. *(See Milchtaich, 2012.) Let $g : X^2 \to \mathbb{R}$ be a population game and $\Phi : \hat{X} \to \mathbb{R}$ a continuous function (on the cone of the strategy space) such that $d\Phi : \hat{X}^2 \to \mathbb{R}$ exists, is continuous in the second argument and satisfies*

$$g(x, y) = d\Phi(x, y), \quad x, y \in X.$$

Then the restriction of Φ to X is a potential for g.

---

[11] A symmetric game is *doubly symmetric* if it has a symmetric payoff function, which means that the players' payoffs are always equal.

Any social payoff f in a symmetric N-player game or social payoff ϕ in a population game is (trivially) a potential for some (other) game in the same class, namely, the modified game corresponding to $r = 1$. Thus, complete altruism turns any symmetric or population game into a potential game. Theorems 5 and 6 therefore have the following corollary, which extends a result of Hofbauer and Sigmund (1988) (see also Weibull, 1995, pp. 56–57) concerning doubly symmetric games.

Corollary 2. *For any symmetric N-player game g and social payoff f, a strategy y is stable, weakly stable or definitely unstable in the modified game g^1 $(= f)$ if and only if it has a neighborhood where for every strategy $x \neq y$ the inequality*

$$f(y, y, \ldots, y) > f(x, x, \ldots, x), \tag{34}$$

a similar weak inequality or the reverse inequality, respectively, holds. The same is true for any population game g and social payoff ϕ, except that in this case (where $g^1 = d\phi$) (34) is replaced by

$$\phi(y) > \phi(x).$$

If ϕ is strictly concave, a strategy is stable in g^1 if and only if it is an equilibrium strategy.

For population games, Corollary 2 generalizes a well-known result pertaining to nonatomic congestion games (Milchtaich, 2004).

Example 4 *(Nonatomic congestion game).* An infinite population of identical users, modeled as the unit interval $[0, 1]$, shares a finite number n of common resources (for example, road segments). The cost of using each resource j (for example, the time it takes to traverse the road) depends on the size of the set of its users and it is specified by a continuously differentiable and strictly increasing cost function $c_j : [0, \infty) \to [0, \infty)$. Each user t has to choose a subset of resources (for example, a route, comprising several road segments), which can be expressed as a binary vector $\sigma(t) = (\sigma_1(t), \sigma_2(t), \ldots, \sigma_n(t))$, where $\sigma_j(t) = 1$ or 0 indicates that resource j is included or is not included in t's choice, respectively. The vector must belong to a specified finite collection $\check{X} \subseteq \{0, 1\}^n$, which describes the allowable subsets of resources (for example, all routes from town A to town B). The population strategy y is the users' mean choice:

$$y = \int_0^1 \sigma(t)\, dt.$$

It lies in the convex hull of $\check{X}$,

$$X \overset{\text{def}}{=} \operatorname{co} \check{X} \subseteq \mathbb{R}^n,$$

and it is well-defined if for each j the set $\{0 \leqslant t \leqslant 1 \mid \sigma_j(t) = 1\}$ is measurable. The population strategy y determines the cost of each allowable subset of resources, and more generally, the cost of each (mixed) strategy $x = (x_1, x_2, \ldots, x_n) \in X$. Specifically, the negative of the latter, which is the payoff $g(x, y)$, is given by

$$g(x, y) = -\sum_{j=1}^n x_j c_j(y_j).$$

A social payoff for the population game $g : X^2 \to \mathbb{R}$ is the mean payoff ϕ, defined by

$$\phi(y) = g(y, y) = -\sum_{j=1}^n y_j c_j(y_j).$$

The modified payoff is given by

$$g^r(x, y) = (1 - r)g(x, y) + r\, d\phi(x, y) = -\sum_{j=1}^n x_j \big(c_j(y_j) + r y_j c_j'(y_j)\big).$$

In particular, setting $r = 1$ means replacing the resources' costs (the $c_j(y_j)$'s) with the corresponding marginal social costs $(d/dy_j(y_j c_j(y_j)))$. If the latter are all strictly increasing functions, ϕ is strictly concave. In this case, it follows from Corollary 2 that the unique maximum point in X of the mean payoff ϕ is the unique equilibrium (as well as the unique stable) strategy in the game g^1.

Corollary 2 may be viewed as a rudimentary comparative statics result. As it shows, every stable strategy in the "extreme" modified game g^1, where the players' only concern is the effect or marginal effect of their choice of strategies on the social payoff, is, in a sense, a local maximizer of the social payoff, and every local (and, in particular, global) maximizer is a weakly stable strategy in g^1. (For symmetric $n \times n$ games, 'strategy' means mixed strategy. A similar assertion does not hold for pure strategies.) This result may be interpreted as entailing that complete altruism maximizes social welfare. (For a somewhat similar result, see Bernheim and Stark, 1988.) Corollary 2 is however much more limited than Theorems 1 and 2. In particular, it says nothing about comparative statics with low levels of altruism, indeed, about any modified game g^r with $r < 1$.

The stability conditions in Theorems 1 and 2 take a particularly simple form if the symmetric N-player game or population game g is itself a potential game, with a potential F or Φ, respectively. In this case, for any altruism coefficient r, the modified game g^r is also a potential game, with the potential

$$F^r = (1 - r)F + rf \quad \text{or} \quad \Phi^r = (1 - r)\Phi + r\phi. \tag{35}$$

By Theorems 5 and 6, a strategy y^r in g^r is stable if and only if it is a strict local maximum point of the function $x \mapsto F^r(x, x, \ldots, x)$ or of the function Φ^r. As r increases, so does the weight attached to the social payoff f or ϕ in (35). It is hence not very surprising that $\pi(r)$, defined by (18) or (23), also increases. Similar intuition applies to weakly stable and to definitely unstable strategies.

In the special case of a symmetric two-player game g that is an $n \times n$ game, it can be shown that a necessary and sufficient condition for g to be a potential game is that the $(n - 1) \times (n - 1)$ matrix defining the quadratic form Q in (43) below is symmetric. (The potential itself is given by a symmetric $n \times n$ matrix.) This condition holds trivially if $n = 2$, which proves that all symmetric 2×2 games are potential games. The following proposition shows that for symmetric $n \times n$ potential games, and with the social payoff defined as the aggregate payoff, completely mixed equilibrium strategies corresponding to different values of the altruism coefficient lie side-by-side along a straight line in the strategy space. In other words, the line segment connecting any pair of such strategies consists of equilibrium strategies corresponding to intermediate values of the altruism coefficient. Along that line, the players' personal payoff either does not change or changes monotonically (cf. the remarks at the end of Section 5.1).

Proposition 4. *For a symmetric $n \times n$ potential game g, the aggregate payoff as the social payoff, and altruism coefficients r_0 and r_1 with $-1 < r_0 < r_1 \leqslant 1$, define for $r_0 \leqslant r \leqslant r_1$*

$$\sigma = \frac{1 - r}{1 + r}, \tag{36}$$

and denote the values corresponding to r_0 and r_1 by σ_0 and σ_1.[12] If the modified games g^{r_0} and g^{r_1} have completely mixed equilibrium strategies y^{r_0} and y^{r_1}, respectively, then for every $r_0 \leqslant r \leqslant r_1$ the convex combination

$$y^r = \frac{\sigma - \sigma_1}{\sigma_0 - \sigma_1} y^{r_0} + \frac{\sigma_0 - \sigma}{\sigma_0 - \sigma_1} y^{r_1} \tag{37}$$

is a completely mixed equilibrium strategy in g^r. The corresponding personal payoff is given by

$$g(y^r, y^r) = g(y^{r_0}, y^{r_0}) + \frac{\sigma_0^2 - \sigma^2}{\sigma_0^2 - \sigma_1^2} \left(g(y^{r_1}, y^{r_1}) - g(y^{r_0}, y^{r_0}) \right), \tag{38}$$

and it is thus determined by r as a monotonic function in the interval $[r_0, r_1]$.

Proof. The potential game g can be written as

$$g = F + \hat{g},$$

where F is a potential and $\hat{g}$ is a function that is constant in the first argument: for an arbitrary, fixed strategy z, $\hat{g}(x, y) = \hat{g}(z, y)$ for all strategies x and y. Since the social payoff is the aggregate payoff and F is symmetric, for an altruism coefficient r and strategies x and y

$$g^r(x, y) = g(x, y) + rg(y, x) = (1 + r)g(x, y) + r(\hat{g}(z, x) - \hat{g}(z, y)).$$

A completely mixed strategy y is an equilibrium strategy in g^r if and only if the value of $g^r(x, y)$ does not depend on the strategy x, which is the case if and only if the function

---

[12] The definition may be viewed as an alternative parameterization of altruism in symmetric two-player games, for $-1 < r \leqslant 1$. It is not difficult to see that, since the social payoff here is twice the average payoff $\bar{g}$, the modified payoff can be written as a positive multiple of $\sigma g + (1 - \sigma)\bar{g}$. Since in the last expression the nonnegative parameter σ is the weight attached to the player's own payoff g, it may be called the *selfishness coefficient*. The coefficient is negatively related to the altruism coefficient r: it increases as the latter decreases.

$$c_{r,y}(\cdot) = g(\cdot, y) + \frac{r}{1+r}\hat{g}(z, \cdot)$$

is constant. It follows from (36) and (37) after some algebra that

$$c_{r,y^r}(\cdot) = \frac{\sigma - \sigma_1}{\sigma_0 - \sigma_1}c_{r_0,y^{r_0}}(\cdot) + \frac{\sigma_0 - \sigma}{\sigma_0 - \sigma_1}c_{r_1,y^{r_1}}(\cdot).$$

If y^{r_0} and y^{r_1} are equilibrium strategies in the respective modified games, this identity proves that the same is true for y^r. The corresponding personal payoff $g(y^r, y^r)$ can then be computed by replacing y^r with the expression in (37). After some algebra, this substitution gives the expression on the right-hand side of (38) plus a third term that can be written as

$$\frac{(\sigma_0 - \sigma)(\sigma - \sigma_1)}{(\sigma_0 + \sigma_1)(\sigma_0 - \sigma_1)^2}\left(\sigma_1(1+\sigma_0)\left(g^{r_0}\left(y^{r_1}, y^{r_0}\right) - g^{r_0}\left(y^{r_0}, y^{r_0}\right)\right) + \sigma_0(1+\sigma_1)\left(g^{r_1}\left(y^{r_0}, y^{r_1}\right) - g^{r_1}\left(y^{r_1}, y^{r_1}\right)\right)\right).$$

Since y^{r_0} and y^{r_1} are completely mixed equilibrium strategies, the last expression is equal to zero. □

6. Global comparative statics

Global comparative statics differ from local comparative statics in not being limited to continuous changes in the altruism coefficient or in the corresponding strategies. This makes the analysis applicable also to games with finite strategy spaces, which is not the case for local comparative statics (see Section 3). Formally, the topology on the strategy space X is put out of the way, so to speak, by taking it to be the trivial topology: the only neighborhood of any strategy is the entire space. Stability, weak stability or definite instability of a strategy with respect to this topology automatically implies the same with respect to any other topology on X, and is referred to as *global* stability, weak stability or definitely instability, respectively. The previous sections demonstrate a strong association between stability and positive local comparative statics. As might be expected, a similar association exists between global stability and positive global comparative statics.

Theorem 7. *For a symmetric two-player game g, a social payoff f, and altruism coefficients r and s with $r < s \leqslant 1$, if two distinct strategies y^r and y^s are globally stable in g^r and g^s, respectively, then*

$$f\left(y^s, y^s\right) > f\left(y^r, y^r\right). \tag{39}$$

If the strategies are globally weakly stable, then

$$f\left(y^s, y^s\right) \geqslant f\left(y^r, y^r\right). \tag{40}$$

If they are globally definitely unstable, then

$$f\left(y^s, y^s\right) < f\left(y^r, y^r\right). \tag{41}$$

Proof. The proof uses the following identity, which holds for all $(r, s$ and$)$ x and y:

$$(r - s)\left(f(x, x) - f(y, y)\right) = (1 - s)\left(g^r(x, x) - g^r(y, x) + g^r(x, y) - g^r(y, y)\right)$$
$$+ (1 - r)\left(g^s(y, y) - g^s(x, y) + g^s(y, x) - g^s(x, x)\right).$$

For $x = y^s$ and $y = y^r$, the right-hand side is negative, nonpositive or positive if these strategies are globally stable, globally weakly stable or globally definitely unstable, respectively, in the corresponding modified games. □

Global stability or weak stability holds *automatically* for all equilibrium strategies in every symmetric two-player game that satisfies the corresponding condition in the following proposition. Since there can obviously be at most one globally stable strategy, in the former case the game also has at most one equilibrium strategy.

Proposition 5. *If a symmetric two-player game g satisfies the* symmetric substitutability *condition that, for every pair of distinct strategies x and y,*

$$g(x, x) - g(y, x) < g(x, y) - g(y, y), \tag{42}$$

then every equilibrium strategy in g is globally stable. If g satisfies the weak symmetric substitutability *condition, in which the strict inequality (42) is replaced by a weak one, then every equilibrium strategy in g is globally weakly stable.*

Proof. An equilibrium strategy y by definition satisfies the two-player version of (3) for every strategy x. That inequality and (42) or its weak version together imply (16) or its weak version, respectively. □

The term symmetric substitutability (Bergstrom, 1995) reflects the following interpretation of (42): switching from strategy y to strategy x increases a player's payoff less (or decreases it more) if the opponent uses x than if he uses y. An alternative, related interpretation of the symmetric substitutability condition is that *coordination* decreases the players' payoffs. This interpretation is based on the fact that the difference between the left- and right-hand sides of (42) is equal to four times the difference between (i) each player's expected payoff if the two players *jointly* randomize 50–50 between x and y, and so always choose the same strategy, and (ii) the expected payoff if the players *independently* randomize 50–50 between x and y. More generally, suppose that for some finite list of distinct strategies $x^1, x^2, \ldots$ both players use each strategy x^i with the same (marginal) probability $p_i > 0$ (with $\sum_i p_i = 1$). Then, the difference between each player's expected payoff if the strategy choices of the two players are perfectly correlated and the expected payoff if the choices are independent is given by $\sum_{i<j} p_i p_j (g(x^i, x^i) - g(x^j, x^i) - g(x^i, x^j) + g(x^j, x^j))$. A sufficient condition for this expression to be negative or nonpositive is that (42) or its weak version, respectively, holds whenever $x \neq y$.

When applied to the modified game, Proposition 5 shows that a sufficient condition for global stability of an equilibrium strategy in g^r is that this game satisfies the symmetric substitutability condition. Clearly, if the set of all r values for which this condition holds is not empty, then it is a (finite or infinite) interval. In the special case where f is the aggregate payoff, the symmetric substitutability condition holds for $r = 0$ (that is, for the unmodified game) if and only if it holds for all $-1 < r \leqslant 1$. This is because the inequality obtained from (42) by replacing g with g^r is equivalent to that obtained by multiplying both sides by $1 + r$.

6.1. Symmetric $n \times n$ games

One class of symmetric two-player games for which Theorem 7 can be given a more concrete form is symmetric $n \times n$ games.

Definition 4. For a symmetric $n \times n$ game g, altruism *globally increases payoffs*, *globally weakly increases payoffs* or *globally decreases payoffs for completely mixed equilibria* if, with the social payoff f defined as the aggregate payoff, for every pair of altruism coefficients r and s with $-1 < r < s \leqslant 1$ and distinct equilibrium strategies y^r and y^s in g^r and g^s, (39) holds, (40) holds, or (41) holds provided y^r and y^s are completely mixed, respectively.

Proposition 6. *For a symmetric $n \times n$ game g, consider the quadratic form $Q : \mathbb{R}^{n-1} \to \mathbb{R}$ defined by*

$$Q(\zeta_1, \zeta_2, \ldots, \zeta_{n-1}) = \sum_{i,j=1}^{n-1} \big(g(i, j) - g(n, j) - g(i, n) + g(n, n)\big) \zeta_i \zeta_j. \tag{43}$$

1. *If Q is negative definite, then altruism globally increases payoffs, and for every $-1 < r \leqslant 1$, the modified game g^r has a unique equilibrium strategy, which is globally stable (hence, an ESS).*
2. *If Q is negative semidefinite, then altruism globally weakly increases payoffs, and for every $-1 < r \leqslant 1$, every equilibrium strategy in g^r is globally weakly stable (hence, an NSS).*
3. *If Q is positive definite, then altruism globally decreases payoffs for completely mixed equilibria, and for every $-1 < r \leqslant 1$, the modified game g^r has at most one completely mixed equilibrium strategy, which is definitely (evolutionarily) unstable.*

Proof. For (mixed) strategies $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$, the difference between the left- and right-hand sides of (42) is equal to $\sum_{i,j=1}^{n} g(i, j) \zeta_i \zeta_j$, where $\zeta_i = x_i - y_i$. This sum is equal to that in (43), since $\sum_{i=1}^{n} \zeta_i = 0$. Therefore, Assertions 1 and 2 follow from Theorem 7, Proposition 5, the comment that immediately precedes this subsection and the fact that every symmetric $n \times n$ game has a symmetric equilibrium. If Q is positive definite, then $-Q$ is negative definite, and therefore the conclusion in Assertion 1 holds for the game $-g$. It only remains to note that $-g$ and g share the same completely mixed equilibrium strategies, and that stability of a strategy in the former is equivalent to definite instability in the latter. $\square$

The quadratic form Q defined by (43) is particularly simple if $n = 2$. In this case, Q is negative definite, negative semidefinite or positive definite if and only if $g(1, 1) - g(2, 1) - g(1, 2) + g(2, 2)$ is negative, nonpositive or positive, respectively. The last expression is the difference between the sum of the two diagonal entries of the payoff matrix and the sum of the two off-diagonal entries. A negative or positive difference expresses *strategic substitutability* or *complementarity*, respectively, in the sense that the profitability of switching from one pure strategy i to the other j decreases or increases, respectively, as the probability that the other player uses j increases (cf. (42)). Symmetric 2×2 games with strategic substitutability include Chicken and the battle-of-the-sexes game, in its symmetric form. By Proposition 6, in such games a higher altruism coefficient entails a weakly higher personal payoff at (the symmetric) equilibrium. By contrast, in games with strategic complementarity, if the equilibria are completely mixed, the personal payoff can only decrease or remain unchanged when r increases. In the prisoner's dilemma, both strategic substitutability and complementarity are possible, and in addition there is a borderline case in which the personal equilibrium payoff abruptly increases at some critical r (Milchtaich, 2006, Fig. 2a).

6.2. Games with a unidimensional strategy space

In a symmetric two-player game g with a unidimensional strategy space, a sufficient condition for symmetric substitutability is that g is *strictly submodular*, that is,

$$g(x, x') - g(y, x') < g(x, y') - g(y, y') \tag{44}$$

whenever $x > y$ and $x' > y'$. A sufficient condition for weak symmetric substitutability is that g is *submodular*, that is, satisfies a similar condition with the strict inequality in (44) replaced by a weak one. If the strategy space is a (finite or infinite) interval and g has continuous second-order partial derivatives, submodularity is equivalent to $g_{12} \leqslant 0$ (everywhere) and a sufficient condition for strict submodularity is $g_{12} < 0$ (everywhere). Thus, if the latter condition holds, and f is the aggregate payoff, then it follows from Proposition 5 that, for every altruism coefficient $-1 < r \leqslant 1$, an equilibrium strategy y in the modified game g^r is necessarily globally stable. This result may be viewed as a limited global version of Theorem 4. Note, however, that for an equilibrium strategy y in g the inequality $g_{12}(y, y) < 0$ (and, a fortiori, the requirement that $g_{12} < 0$ everywhere) is a stronger condition than (28).

In the special case of a symmetric Cournot duopoly game (see Example 1), symmetric substitutability is also implied by a simpler and somewhat weaker condition than strict submodularity. In such a game, the profit $g(x, y)$ of a producer with output level x competing against an identical producer with output level y is $xP(x + y) - C(x)$, where P is the price (or inverse demand) function and C is the cost function. Therefore, (42) can be written as

$$\frac{1}{2}\big(2xP(2x) + 2yP(2y)\big) < (x + y)P(x + y). \tag{45}$$

(Note that the inequality does not involve C. This is because the production cost is a function of the firm's own output only.) A sufficient (and, if the price function is continuous, also necessary) condition for (45) to hold for every pair of distinct points x and y in the (finite or infinite) interval of possible output levels is that the total revenue is a strictly concave function of the total output. Similarly, if the total revenue is concave, the weak-inequality version of (42) always holds. It then follows from Proposition 5 and Theorem 7 that negative comparative statics (exemplified by the lower curve in Fig. 1b) cannot occur, so that moving from duopoly towards (not necessarily all the way to) effective monopoly cannot decrease the firms' profits. This proves the following.

Corollary 3. *In a symmetric Cournot duopoly game, each of the following two conditions, the latter being weaker than the former, implies that altruism $(0 < r \leqslant 1)$ and spite $(-1 < r < 0)$ can only increase or decrease, respectively, the firms' equilibrium profit or leave it unchanged, relative to the case in which each firm is only concerned with its own profit $(r = 0)$:*

 (i) *A firm's profit is given by a submodular function of the two firms' output levels.*
(ii) *The total revenue is a concave function of the total output.*

It is instructive to compare Corollary 3 with the results of Koçkesen et al. (2000). These authors show that, in a symmetric Cournot duopoly game that is strictly submodular, a firm with negatively interdependent preferences obtains a strictly higher profit than does a competitor with independent preferences in any equilibrium. The difference between the two firms' preferences is that the latter is only concerned with its own profit while the former also seeks a high ratio ρ between its own and the average profit. Complete selfishness, $r = 0$, corresponds to independent preferences, whereas weak spite, i.e., negative r close to 0, gives negatively interdependent preferences if the ratio ρ is not too close to zero. Thus, with a strictly submodular profit, if only one firm is spiteful, it is likely to do better than its competitor does. However, by Corollary 3, if both firms have such preferences, they do not have higher profits than two firms with independent preferences.

The result that mutual spite cannot be beneficial does not necessarily hold when the conditions in Corollary 3 are not met. As Fig. 1b shows, in Example 1, where the total revenue is not concave (since it tends to zero when the total output increases), duopolists stuck at the inefficient (high-output, low-profit) equilibrium would actually benefit from a low level of mutual spite. The reason, as indicated, is that spite encourages increased production by emphasizing its negative effect on the competitor, and thus raises the reaction curve. Since at the inefficient, unstable equilibrium the slope of the reaction curve is greater than 1 (see Fig. 1a), the result is actually *lower* equilibrium outputs and consequently a higher profit.

7. Asymmetric games

An asymmetric game differs from a symmetric one in that each of the players i has his own strategy space X_i and payoff function h_i. Correspondingly, for an asymmetric N-player game $h = (h_1, h_2, \ldots, h_N) : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$, a social payoff function is any (rather than necessarily symmetric) function $f : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}$. For an altruism coefficient $r \leqslant 1$, the modified game $h^r = (h_1^r, h_2^r, \ldots, h_N^r) : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$ is defined by (1).

A complete analysis of comparative statics in asymmetric N-player games is not yet attainable. This is mainly because of the unavailability of a suitable notion of (static) stability, which in the asymmetric context is a property of strategy profiles rather than strategies. A partial workaround is to view strategy profiles as strategies in another, symmetric N-player

game, namely, the game obtained by symmetrizing the asymmetric one. In that game, the players' roles are not fixed: all $N!$ assignments of players to roles in the asymmetric game are possible. A player's payoff is defined as the average of his payoff over the set Π of all assignments. Each assignment $\rho \in \Pi$ is a permutation of $(1, 2, \ldots, N)$: $\rho(i)$ is the player assigned to role i $(= 1, 2, \ldots, N)$. A strategy for a player in the symmetric game is a strategy profile $x = (x_1, x_2, \ldots, x_N)$ in the asymmetric game: it specifies the strategy x_i the player would use in each role i. A more formal definition follows.

Definition 5. *Symmetrization* of an asymmetric N-player game $h = (h_1, h_2, \ldots, h_N) : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$ gives the symmetric N-player game $g : X^N \to \mathbb{R}$ in which the strategy space X is the product space $X_1 \times X_2 \times \cdots \times X_N$ and, for all $x^1 = (x_1^1, x_2^1, \ldots, x_N^1), x^2 = (x_1^2, x_2^2, \ldots, x_N^2), \ldots, x^N = (x_1^N, x_2^N, \ldots, x_N^N) \in X$,[13]

$$g(x^1, x^2, \ldots, x^N) = \frac{1}{N!} \sum_{\rho \in \Pi} h_{\rho^{-1}(1)} (x_1^{\rho(1)}, x_2^{\rho(2)}, \ldots, x_N^{\rho(N)}). \tag{46}$$

Any choice of topologies for the strategy spaces $X_1, X_2, \ldots, X_N$ specifies a topology for X, namely, the product topology.

Symmetrization in a sense preserves the original game's equilibria. It is not difficult to see that a strategy profile in an asymmetric N-player game h is an equilibrium if and only if it is a (symmetric) equilibrium strategy in the symmetric game g obtained by symmetrizing h. In this case, the equilibrium payoff in g equals the players' average equilibrium payoff in h.

The following theorem links local comparative statics in an asymmetric game with stability in the symmetric game obtained from it by symmetrization. The theorem holds for any choice of topologies for the players' strategy spaces in the asymmetric game; stability and instability in the corresponding symmetric game are with respect to the product topology.

Theorem 8. *For an asymmetric N-player game $h : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$ and a social payoff function $f : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}$ that are both Borel measurable, and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leqslant 1$, suppose that there is a continuous and finitely-many-to-one function that assigns a strategy profile y^r to each $r_0 \leqslant r \leqslant r_1$ such that the function $\pi : [r_0, r_1] \to \mathbb{R}$ defined by*

$$\pi(r) = f(y^r) \tag{47}$$

is absolutely continuous. If, for every $r_0 < r < r_1$, y^r is stable, weakly stable or definitely unstable as a strategy in the game obtained from the modified game h^r by symmetrization, then π is strictly decreasing, nonincreasing or strictly increasing, respectively.

Proof. The modified game h^r is a linear combination of two games, h and a game in which all players have the (same) payoff function f. Symmetrizing the latter gives a doubly symmetric game with a payoff function $\overline{f}$, which may be viewed as a social payoff for the game g obtained by symmetrizing h. It is easy to see that both $\overline{f}$ and g are Borel measurable. Eq. (46) shows symmetrization to be a linear operator. Therefore, the game obtained by symmetrizing h^r is $(1 - r)g + r\overline{f}$, and it thus coincides with the modified game g^r.[14] If, for every $r_0 < r < r_1$, y^r is a stable, weakly stable or definitely unstable strategy in g^r and the function $r \mapsto \overline{f}(y^r, y^r, \ldots, y^r)$ is absolutely continuous, then by Theorem 1 this function is strictly increasing, nondecreasing or strictly decreasing, respectively. It only remains to note that, by Definition 5, $\overline{f}(x, x, \ldots, x) = f(x)$ for all strategy profiles x in h. □

Theorem 8 is applicable both to asymmetric games and to asymmetric equilibria in symmetric games. The symmetric games need simply be viewed as (a special kind of) asymmetric games. However, the theorem does not give Theorem 1 as a special case. This is because the stability, weak stability and definite instability conditions in Theorem 8 are in a sense much more stringent than the corresponding conditions in Theorem 1. If a symmetric game g is viewed as an asymmetric game h, then symmetrization of the latter gives a second symmetric game $\overline{g}$, in which each strategy $x = (x_1, x_2, \ldots, x_N)$ is a strategy profile in g and vice versa. If the strategy profile is symmetric, that is, $x_1 = x_2 = \cdots = x_N$, then stability of x_1 in g is *not* a sufficient condition (but it is a necessary one) for stability of x in $\overline{g}$. For example, if g is a symmetric $n \times n$ game, x may be stable in $\overline{g}$ (which, parenthetically, is not an $m \times m$ game, for any m) only if x_1 is a pure strategy.

The last result holds more generally. For any bimatrix game h, a strategy profile x is a stable strategy in the game obtained from h by symmetrization if and only if it is a strict equilibrium in h (Selten, 1980). Since strict equilibria are pure, and there are only finitely many pure strategy profiles, for bimatrix games (and similar games with more than two players) Theorem 8 is inapplicable: the stability condition is inconsistent with the finitely-many-to-one condition.

One class of asymmetric games to which Theorem 8 *is* applicable is games with unidimensional strategy spaces. The differential condition for stability of equilibrium in these games is given by the following proposition, whose proof is omitted.

Proposition 7. *Let $h = (h_1, h_2, \ldots, h_N) : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$ be an asymmetric N-player game in which the players' strategy spaces are subsets of the real line, and $x = (x_1, x_2, \ldots, x_N)$ an equilibrium lying in the interior of $X_1 \times X_2 \times \cdots \times X_N$ with a*

---

[13] Note that superscripts here index strategies in the symmetric game while subscripts refer to roles in the asymmetric one.

[14] Thus, in a sense, the operations of modifying a game and symmetrization commute.

neighborhood where $h_1, h_2, \ldots, h_N$ have continuous second-order derivatives. A sufficient condition for stability or definite instability of x as a strategy in the game obtained from h by symmetrization is that, at the point x, the Jacobian matrix of the marginal payoffs,

$$H = \begin{pmatrix} (h_1)_{11} & \cdots & (h_1)_{1N} \\ \vdots & \ddots & \vdots \\ (h_N)_{N1} & \cdots & (h_N)_{NN} \end{pmatrix}, \tag{48}$$

is negative definite or positive definite,[15] respectively. A necessary condition for weak stability is that the matrix is negative semidefinite.

Theorem 8 and Proposition 7 together point to a connection between local comparative statics in games with unidimensional strategy spaces and properties of the matrix H^r obtained from (48) by replacing h with the modified game h^r. An essentially stronger, quantitative connection between that matrix and comparative statics is given by the following proposition, whose proof is rather similar to that of Proposition 2.

Proposition 8. *For an asymmetric N-player game $h = (h_1, h_2, \ldots, h_N) : X_1 \times X_2 \times \cdots \times X_N \to \mathbb{R}^N$ with strategy spaces that are subsets of the real line, a social payoff function f, and altruism coefficients r_0 and r_1 with $r_0 < r_1$, suppose that there is a continuously differentiable function that assigns an equilibrium y^r in the modified game h^r to each $r_0 < r < r_1$ such that y^r lies in the interior of $X_1 \times X_2 \times \cdots \times X_N$ and has a neighborhood where $h_1, h_2, \ldots, h_N$ and f have continuous second-order partial derivatives. Then, at each point $r_0 < r < r_1$,*

$$\frac{d\pi}{dr} = -(1-r)\left(\frac{dy^r}{dr}\right)^{\mathrm{T}} H^r \left(\frac{dy^r}{dr}\right),$$

where π is defined by (47) and the matrix H^r is evaluated at y^r.

7.1. Global comparative statics in asymmetric games

Global comparative statics are less relevant for asymmetric games than for symmetric ones. This is because a condition analogous to (even) weak symmetric substitutability does not hold for most asymmetric two-player games. A notable exception is games in which such a condition holds as equality:

$$h_i(x_1, x_2) - h_i(y_1, x_2) = h_i(x_1, y_2) - h_i(y_1, y_2) \quad \text{for all } x_1, x_2, y_1, y_2 \text{ and } i = 1, 2. \tag{49}$$

Games satisfying this condition are non-strategic in that the change in the payoff of a player switching strategies (from x_1 to y_1 in the case of player 1 or from x_2 to y_2 in the case of player 2) is independent of the opponent's strategy. It is easy to see that, for any social payoff f and altruism coefficient r, if (49) holds and a similar condition holds with h_i replaced by f, then the same is true with h replaced by h^r. If f is the aggregate payoff $h_1 + h_2$, then the second condition is clearly redundant: if (49) holds, then a similar condition automatically holds for the modified game h^r, for any altruism coefficient r.

It is not difficult to show that, if an asymmetric game $h = (h_1, h_2)$ satisfies (49), then every equilibrium in h is a weakly stable strategy in the game obtained by symmetrizing h. The following proposition shows that, in such a game h, altruism can only positively affect the social payoff or leave it unchanged.

Proposition 9. *Let $h = (h_1, h_2) : X_1 \times X_2 \to \mathbb{R}^2$ be an asymmetric two-player game, f a social payoff, and r and s altruism coefficients with $r < s \leqslant 1$. If a condition similar to (49) holds for both modified games h^r and h^s, then for any pair of corresponding equilibria $y^r = (y_1^r, y_2^r)$ and $y^s = (y_1^s, y_2^s)$*

$$f(y^s) \geqslant f(y^r).$$

Proof. The proof uses the following identity, which holds for all $(r, s$ and$)$ strategy profiles x and y:

$$\begin{aligned} 2(r-s)&\big(f(x_1, x_2) - f(y_1, y_2)\big) \\ &= 2(1-s)\big(h_1^r(x_1, y_2) - h_1^r(y_1, y_2)\big) + 2(1-s)\big(h_2^r(y_1, x_2) - h_2^r(y_1, y_2)\big) \\ &\quad + 2(1-r)\big(h_1^s(y_1, y_2) - h_1^s(x_1, y_2)\big) + 2(1-r)\big(h_2^s(y_1, y_2) - h_2^s(y_1, x_2)\big) \\ &\quad + (1-s)\big(h_1^r(x_1, x_2) - h_1^r(y_1, x_2) - h_1^r(x_1, y_2) + h_1^r(y_1, y_2)\big) \\ &\quad + (1-s)\big(h_2^r(x_1, x_2) - h_2^r(y_1, x_2) - h_2^r(x_1, y_2) + h_2^r(y_1, y_2)\big) \end{aligned}$$

---

[15] By definition, H is negative or positive definite if the symmetric matrix $(1/2)(H + H^{\mathrm{T}})$ has the same property. Negative definiteness implies that H is *D-stable* (but not conversely). The latter is often taken to be the criterion of stability in dynamic contexts (e.g., Dixit, 1986).

$$+ (1-r)\big(h_1^s(x_1, x_2) - h_1^s(y_1, x_2) - h_1^s(x_1, y_2) + h_1^s(y_1, y_2)\big)$$
$$+ (1-r)\big(h_2^s(x_1, x_2) - h_2^s(y_1, x_2) - h_2^s(x_1, y_2) + h_2^s(y_1, y_2)\big).$$

For the equilibria $x = y^s$ and $y = y^r$, the first four terms on the right-hand side are nonpositive. If the assumption concerning (49) holds, then the last four terms are equal to zero, so that the right-hand side is nonpositive. □

A simple example illustrating the last result is *non-strategic altruism*, which is of considerable importance to the theory of kin selection (for references, see Milchtaich, 2006). An altruistic act confers a benefit b on the recipient at a cost c to the actor, with $b > c > 0$. Therefore, with f that is the aggregate payoff, the act changes the actor's modified payoff by $-c + rb$. If the altruism coefficient r is greater than c/b, this means that the actor's payoff in the modified game is maximized by acting altruistically. Increasing the altruism coefficient can therefore only increase the aggregate payoff of leave it unchanged.

An increase in the aggregate payoff spells a gain for both individuals if the interaction between them is symmetric in that each has an equal chance to be in the position of a potential actor or receiver. Such a symmetric interaction is still non-strategic, since each player's payoff is *additively separable*. Specifically, the payoff is the sum of a nonpositive term (cost) that is 0 or $-c$, depending on the individual's own decision of whether to act altruistically, and a nonnegative term (benefit) that is 0 or b, depending on the other individual's decision; there is no interaction term. The game is thus a prisoner's dilemma, specifically, the borderline case between strategic substitutability and complementarity mentioned at the end of Section 6.1.

8. Dynamic stability and comparative statics

The main finding of this paper is that negative comparative statics, whereby altruism paradoxically negatively affects the equilibrium social payoff, are unlikely if the equilibrium strategies involved are statically stable, that is, stable in the sense described in Section 4. A corollary of this finding is that, in groups or societies in which the dynamics of strategy choices tend to exclude statically unstable equilibrium strategies, so that dynamic stability implies static stability, altruism is only likely to make the group members better off. This result has particular relevance for groups that compete with each other, so that the effect of altruism on social welfare may affect the group's probability of survival. In this case, the above finding suggests that altruism may be favored by group selection. Thus, dynamic stability, which refers to *intragroup* dynamics, may be consequential for *intergroup* dynamics.

Whether or not dynamic stability implies static stability depends on the particular dynamics involved. This is illustrated by the case of symmetric $n \times n$ games, for which the notion of static stability considered in this paper coincides with evolutionary stability (Theorem 3). In an animal population in which such a game g is played between pairs of related individuals with the same coefficient of relatedness r (e.g., full siblings, with $r = 0.5$; see Section 2), the dynamics are governed by mutation and natural selection. A strategy may be considered dynamically stable if it is uninvadable in the (population genetics) sense that, if all members of the population adopt it, no mutant strategy can invade. An uninvadable strategy is necessarily an ESS in the corresponding modified game g^r (but not conversely; see Hines and Maynard Smith, 1979; Milchtaich, 2006). As explained above, this means that in games between relatives in nature, negative comparative statics are unlikely. That is, if in a different population the same game g is played between somewhat more closely related individuals (which means a higher r), the outcome is likely to be either the same as or better than in the first population.

An alternative notion of dynamic stability in symmetric $n \times n$ games, viewed as population games, which is weaker (rather than stronger) than the static notion of evolutionary stability, is asymptotic stability under the continuous-time replicator dynamics (Hofbauer and Sigmund, 1998). The replicator equation gives the rate of change of the coordinates of the population strategy $y = (y_1, y_2, \ldots, y_n)$, which (in the simple version considered below) describes the frequency of use of each pure strategy in the population. Specifically, for a game with a payoff matrix $A = (a_{ij})$, the rate of change $\dot{y}_i$ $(i = 1, 2, \ldots, n)$ depends on the difference between the expected payoff from using (the pure) strategy i and the mean payoff:

$$\dot{y}_i = y_i \left(\sum_{j=1}^{n} a_{ij} y_j - \sum_{j,k=1}^{n} a_{jk} y_j y_k \right).$$

Asymptotic stability with respect to the replicator dynamics does not preclude negative comparative statics and instability does not preclude positive comparative statics. For example, in a generalized rock–scissors–paper game, the equilibrium strategy is globally asymptotically stable under the continuous-time replicator dynamics if and only if the equilibrium payoff is positive (Hofbauer and Sigmund, 1998, Theorem 7.7.2; Weissing, 1991, Theorem 5.6). In this case, the population converges to the equilibrium strategy from any initial interior point (i.e., a completely mixed population strategy). If the equilibrium payoff is negative, the equilibrium is unstable, and the population strategy converges to the boundary of the strategy space from any initial point other than the equilibrium strategy itself. For the left payoff matrix in (27), the equilibrium payoff in the modified game g^r is positive for all $-1 < r < 0.5$, and for the right matrix, it is negative. Hence, in the former case the corresponding equilibrium strategy is stable under the replicator dynamics (as well as under other natural dynamics; see Chamberland and Cressman, 2000), and in the latter, it is unstable. However, as shown in Section 5.1, the personal payoff decreases with increasing altruism coefficient for the left payoff matrix in (27) and increases for the right matrix.

This demonstrates the point made above: depending on the dynamics, the notion of static stability used in Theorems 1 and 2 may or may not be implied by dynamic stability. If the implication does not hold, then paradoxical, negative comparative statics are not necessarily unlikely.

Appendix A. Strong stability

A stable equilibrium strategy in a symmetric $n \times n$ game is normally also "strongly" stable in the sense that a continuous deformation of the payoff matrix changes the equilibrium strategy in a continuous manner. Specifically, the following proposition, which is essentially due to Selten (1983), shows that every *regular* ESS has this property. An ESS is said to be regular if its support includes every action that is a best response to it (equivalently, if the corresponding symmetric equilibrium is quasi-strict; see van Damme, 1991). The proposition has a corollary for local comparative statics; see Section 5.1.

Proposition A.1. *Let y be a regular ESS in a symmetric $n \times n$ game with a payoff matrix A. There is a neighborhood V of y in the strategy space and a neighborhood U of A in $\mathbb{R}^{n^2}$ such that*:

 (i) *every symmetric $n \times n$ game g with a payoff matrix in U has a unique equilibrium strategy in V,*
 (ii) *that strategy x is a regular ESS, its support is equal to that of y, and it is the only equilibrium strategy in g with that support, and*
(iii) *the mapping from U to V thus defined is continuous.*

Proof. According to the regularity assumption, every action i that is not in the support of the ESS y (i.e., $y_i = 0$) is not a best response to y. Therefore, there are neighborhoods V' and U' of y and A, respectively, such that every action i as above is also not a best response to any strategy $x \in V'$ in any symmetric $n \times n$ game with a payoff matrix $B \in U'$. If x is an equilibrium strategy for B, this means that its support is necessarily contained in that of y. Let $V \subseteq V'$ be a closed neighborhood of y that includes only strategies x whose support *contains* that of y (i.e., $x_i > 0$ for every action i with $y_i > 0$).

The regularity of y implies that it is an *essential* ESS (van Damme, 1991, Theorem 9.3.6). That is, every symmetric $n \times n$ game with a payoff matrix close to A has an ESS close to y. In particular, there is a neighborhood $U \subseteq U'$ of A such that every symmetric $n \times n$ game with a payoff matrix $B \in U$ has some ESS $x \in V$. As shown above, the support of x – indeed, of any equilibrium strategy for B lying in V – coincides with that of y. Since the support of an ESS cannot coincide with that of any other equilibrium strategy in the same game (van Damme, 1991, Lemma 9.2.4), conditions (i) and (ii) in the proposition hold. The mapping that assigns to each element of U the set of its equilibrium strategies in V is clearly upper semicontinuous. Since by (i) this mapping is singleton-valued, condition (iii) also holds. □

Acknowledgments

References

Bergstrom, T.C., 1995. On the evolution of altruistic ethical rules for siblings. Amer. Econ. Rev. 85, 58–81.

Bernheim, B.D., Stark, O., 1988. Altruism within the family reconsidered: Do nice guys finish last? Amer. Econ. Rev. 78, 1034–1045.

Bester, H., Güth, W., 1998. Is altruism evolutionarily stable? J. Econ. Behav. Organ. 34, 193–209.

Bolton, G.E., Ockenfels, A., 2000. ERC: A theory of equity, reciprocity and competition. Amer. Econ. Rev. 90, 166–193.

Bomze, I.M., Pötscher, B.M., 1989. Game Theoretical Foundations of Evolutionary Stability. Springer-Verlag, Berlin.

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

Chamberland, M., Cressman, R., 2000. An example of dynamic (in)consistency in symmetric extensive form evolutionary games. Games Econ. Behav. 30, 319–326.

Chen, P.-A., Kempe, D., 2008. Altruism, selfishness, and spite in traffic routing. In: Fortnow, L., Riedl, J., Sandholm, T. (Eds.), Proceedings of the 9th ACM Conference on Electronic Commerce. ACM, New York, pp. 140–149.

Corts, K.S., 2006. When altruism lowers total welfare. Econ. Philos. 22, 1–18.

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

Dixit, A., 1986. Comparative statics for oligopoly. Int. Econ. Rev. 27, 107–122.

Edgeworth, F.Y., 1881. Mathematical Psychics. Kegan Paul, London.

Eshel, I., 1983. Evolutionary and continuous stability. J. Theoretical Biol. 103, 99–111.

Eshel, I., Motro, U., 1981. Kin selection and strong evolutionary stability of mutual help. Theoretical Popul. Biol. 19, 420–433.

Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Quart. J. Econ. 114, 817–868.

Frank, R.H., 1988. Passions within Reason. Norton, New York.

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

Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1, 60–79.

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

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

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

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

Hofbauer, J., Sigmund, K., 1988. The Theory of Evolution and Dynamical Systems. London Math. Soc. Stud. Texts, vol. 7. Cambridge University Press, Cambridge, UK.

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

Koçkesen, L., Ok, E.A., Sethi, R., 2000. The strategic advantage of negatively interdependent preferences. J. Econ. Theory 92, 274–299.

Levine, D.K., 1998. Modeling altruism and spitefulness in experiments. Rev. Econ. Dynam. 1, 593–622.

Lindbeck, A., 1992. Presentation speech, 1970. In: Lindbeck, A. (Ed.), Nobel Lectures in Economic Sciences (1969–1980) – The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel. World Scientific, Singapore, pp. 53–55.

Lindbeck, A., Weibull, J.W., 1988. Altruism and time consistency: The economics of fait accompli. J. Polit. Economy 96, 1165–1182.

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

Milchtaich, I., 2004. Social optimality and cooperation in nonatomic congestion games. J. Econ. Theory 114, 56–87.

Milchtaich, I., 2006. Comparative statics of games between relatives. Theoretical Popul. Biol. 69, 203–210.

Milchtaich, I., 2012. Static stability in symmetric and population games. Mimeo.

Morgan, J., Steiglitz, K., Reis, G., 2003. The spite motive and equilibrium behavior in auctions. Contrib. Econ. Anal. Pol. 2 (1). Article 5.

Myerson, R.B., Pollock, G.B., Swinkels, J.M., 1991. Viscous population equilibria. Games Econ. Behav. 3, 101–109.

Rabin, M., 1993. Incorporating fairness into game theory and economics. Amer. Econ. Rev. 83, 1281–1302.

Rana, I.K., 2002. An Introduction to Measure and Integration, second ed. Grad. Stud. Math., vol. 45. American Mathematical Society, Providence, RI.

Ridley, M., 1997. The Origins of Virtue. Viking, New York.

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

Selten, R., 1980. A note on evolutionarily stable strategies in asymmetric animal conflicts. J. Theoretical Biol. 84, 93–101.

Selten, R., 1983. Evolutionary stability in extensive two-person games. Math. Soc. Sci. 5, 269–363.

Stark, O., 1989. Altruism and the quality of life. Amer. Econ. Rev. 79, 86–90.

van Damme, E., 1991. Stability and Perfection of Nash Equilibria, second ed. Springer-Verlag, Berlin.

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: Evolution and Game Dynamics. Springer-Verlag, Berlin, pp. 29–97.

Yeh, J., 2006. Real Analysis: Theory of Measure and Integration, second ed. World Scientific, New Jersey.