ROOTED POTENTIAL AND STATIC STABILITY IN GAMES

Igal Milchtaich, Department of Economics, Bar-Ilan University, Ramat Gan 5290002, Israel
igal.milchtaich@biu.ac.il

April 2023

Rooted potential is a generalization of (exact) potential that exists in every game. In a potential game, it coincided with the potential. In general games, it varies according to the choice of root, which is any strategy (in symmetric and population games) or strategy profile (in asymmetric games). Roots that are local maxima of the rooted potential have interesting stability features. Depending on the class of games considered, the set of equilibrium roots with this property may coincide with the evolutionarily stable strategies, ESS (in symmetric $n \times n$ games); with the continuously stable strategies, CSS (in symmetric games with a unidimensional strategy space); with the risk dominant equilibria (in finite symmetric 2×2 games); or with the strict equilibria (in bimatrix games). Local maximization of the rooted potential may thus be viewed as a general, unifying notion of *static* stability. Like the earlier notions of static stability it generalizes, it is formulated solely in terms of the players' *incentives* to change their strategies close to the point in question. This differs from dynamic stability, which is specific to an extraneous law of motion that dictates the players' *reactions* to the incentives. Depending on the choice of the latter, dynamic stability may be weaker than static stability or the two may be incomparable. Static stability is also connected with comparative statics of altruism. In general, internalization of the aggregate payoff or some other measure of social payoff by all players may paradoxically result in a decrease of that payoff. But this is never so if the equilibria involved satisfy static stability.

Keywords: Static stability, evolutionarily stable strategy, continuously stable strategy, risk dominance, potential games, comparative statics, altruism.

1 Introduction

Common-payoff games are N-player games where the players' payoff functions are identical: they always get equal payoffs. The maximum points of the common payoff function are (Nash) equilibria, as they Pareto dominate all other strategy profiles. However, some other strategy profiles may also be equilibria. Which raises the question of whether there is any general sense in which such equilibria are less reasonable, compelling or stable than the payoff-maximizing ones.[1] A similar question arises in the more general setting of potential games (Monderer and Shapley 1996). In these games, each player's payoff can be presented as the sum of a common payoff function P, referred to as the potential, and a

[1] A simple example of the two kinds of equilibria is provided by the two pure equilibria in the game

$$\begin{array}{cc} & \begin{array}{cc} L & R \end{array} \\ \begin{array}{c} T \\ B \end{array} & \begin{pmatrix} 3,3 & 0,0 \\ 0,0 & 1,1 \end{pmatrix}. \end{array}$$

When discussing their notion of strategic stability, Kohlberg and Mertens (1986) state that they have nothing to say about the distinction between these equilibria, which they maintain has to do with the pre-play bargaining game, and hence with cooperative theory, rather than with the game itself.

second function that only depends on the other players' strategies. It is still true that the maxima of P are equilibria. Whether there is anything substantially special about these equilibria is even less clear than in the common-payoff case. Consider, for example, the pair of two-player games (often referred to as stag-hunt games)

$$\begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 3,3 & -1,0 \\ 0,-1 & 0,0 \end{pmatrix} \end{array} \quad \text{and} \quad \begin{array}{c} \\ T \\ B \end{array} \begin{array}{c} \begin{array}{cc} L & R \end{array} \\ \begin{pmatrix} 0,0 & 0,-3 \\ -3,0 & 1,1 \end{pmatrix} \end{array}.$$

In the left game, the equilibrium (T, L) is Pareto dominant and the equilibrium (B, R) is "safe". For the right game, it is the other way around. Yet the two games have the same potential P, which is the common payoff in the game in footnote 1.

One way to look at the above questions is to consider unilateral deviations. When a single player switches strategy, the change in that player's payoff is equal to the change in the potential P. When several players do so one after the other, the sum of the payoff changes is given by the difference $P(y) - P(x)$, where x is the initial strategy profile and y is the terminal one. This difference can be described as the overall incentive to move from x to y. It is the negative of the overall incentive to move along the reverse path, from y to x, and can also be considered the "work" done along the latter path, with incentives viewed as forces acting on the players. (The path reversal represents the convention that the work is positive when there is an *opposing* force. Correspondingly, it is $-P$ here that would be considered the potential in physics.) The strategy profile y is a strict maximum point of P if and only if this work is positive for all $x \neq y$. Such y will be referred to as globally stable. This property of a strategy profile distinguishes between the two equilibria in the above examples. If one player deviates from the equilibrium (B, R) by playing T or L, that player's incentive to return is weaker than the other player's incentive to also deviate. Put differently, the players' overall incentive to move to the opposite equilibrium (T, L) is positive. Thus, the two equilibria differ in that only the latter is globally stable, a distinction that applies to all games above.[2]

Most games do not admit a potential, which means that the work done when moving between strategy profiles may depend on the path taken, that is, on the order of moves. (Similar path dependence occurs in non-conservative force fields, which do not have a potential function.) The notion of the overall incentive to move from y to x can be naturally extended to the general setting by averaging the sum of the payoff changes over all possible orderings of the players. Viewed as a function of x, this average defines the *rooted* potential, with y as the root. Global stability of y is generalized to the condition that the rooted potential is negative for all $x \neq y$. A weaker, local condition, referred to simply as stability, is that this is so in a vicinity of the root, that is, for all $x \neq y$ in some neighborhood of y. This local version applies when a system of neighborhoods, or a topology, is defined for the strategies in the game.

Stability is a fundamentally different concept than equilibrium. A body is at equilibrium when it experiences no net force. A strategy profile is an equilibrium if there is no incentive for any unilateral deviation. Stability differs in taking into consideration forces, or incentives, at other, nearby points, which are reached after small displacements. Thus, an object perched at the top of a rock is at equilibrium, but this equilibrium is unstable. The aim of this paper is to explore the meaning and implications of stability, as defined above, in games. As

[2] It is no coincidence that (T, L) is also the risk dominant equilibrium. See below.

shown below, this concept justifies its name in that it is a very broad generalization of a number of earlier, more special, concepts that may be viewed as conveying stability in specific classes of games. These concepts, like the one introduced here, concern static stability, in the sense that they are formulated entirely in terms of incentives. As the players incentives are specified by the payoff functions, all forms of static stability are intrinsic to the game. They can be determined without referring to, and without requiring, additional assumptions concerning dynamics, or the fashion by which incentives lead to changes in the players' behavior.

The notions of rooted potential and stability proposed here are formalized in Section 2. These concepts are universal in that they are applicable to any game, and do not require the strategy spaces or the payoff functions to have any specific form or structure. Moreover, variants of them are applicable to symmetric games and to population games, where stability refers to strategies rather than strategy profiles. A formal connection, established in Section 2.1.1, between stability of strategy profiles in asymmetric games (as described above) and stability of strategies in symmetric games is that the former can be reduced to the latter by symmetrizing the game. In addition, the two concepts are directly comparable in asymmetric games that are essentially symmetric in the sense that the players are interchangeable (Section 2.2). This comparison and several other facts indicate that, in a sense, stability is a weaker requirement in symmetric games than in asymmetric ones.

The exact connections between potential and rooted potential are laid out in Section 3. In particular, it is shown that the rooted potential in an asymmetric game is a potential if and only if it only changes by an additive constant whenever the roots changes. A Similar result holds in population games that admit a potential (Section 3.2). This result relies on a novel definition of potential for such games which significantly extends and generalizes the standard one.

Section 4 is the heart of the paper. It looks at several earlier notions that are also formulated entirely in terms of the players' incentives and can thus be categorized under the heading of static stability, and examines whether they are comparable with (that is, weaker, stronger, or equivalent to) stability. Each comparison is restricted to the class(es) of games where the notion under examination is well defined or most meaningful.

The simplest kind of games for which static stability of any sort is applicable are finite symmetric two-player games with only two strategies, a and b (Section 4.1). If both strategies are equilibrium strategies, then normally exactly one of the two symmetric equilibria, (a, a) or (b, b), is risk dominant. Risk dominance is easily seen to be equivalent to global stability of the corresponding strategy, a or b.

The second existing concept examined is local superiority (Section 4.2). In a symmetric game, a strategy y is locally superior if, when all the players use any nearby strategy x, a unilateral deviation to y is beneficial. An extension to strategy profiles in an asymmetric game can be obtained by symmetrizing the game. In general, local superiority in asymmetric games is a stronger property than stability: the former implies the latter but not the other way around. The two properties are equivalent in the special case of asymmetric N-player games that are the mixed extensions of finite games.[3] (For $N = 2$, these are the bimatrix, or

[3] In a finite game, by definition only pure strategies are considered. The mixed extension is obtained by allowing mixed strategies.

$m \times n$, games.) However, the only strategy profiles in these games with either property are the strict equilibria. In symmetric games, local superiority is in general incomparable with stability; neither property implies the other. Two special cases where local superiority is stronger than stability are equilibrium strategies in symmetric two-player games, and a large class of population games. An even more special case, in which local superiority and stability are equivalent, is symmetric two-player and population games that are the mixed extensions of finite games, that is, symmetric $n \times n$ games.

Stability in symmetric $n \times n$ games is also equivalent to another kind of static stability: evolutionary stability (Section 4.3). Thus, a strategy in such a game is stable if and only if it is locally superior if and only if it is an evolutionarily stable strategy, ESS. Interestingly, however, the same is not true in the multiplayer case, that is, for the mixed extensions of finite symmetric N-player games (Section 4.3.1). For general N, stability of a strategy is a sufficient condition for ESS and is a necessary condition for local superiority, but in both cases the reverse implication does not hold.

In games where a player's strategy space is a convex set in a Euclidean space, to that strategies are real numbers (in the unidimensional case) or vectors, it may be possible to present stability of equilibria or equilibrium strategies in a differential form, that is, as a condition involving partial derivatives of the payoff function(s). In symmetric two-player and population games with a unidimensional strategy space (Section 4.4), this differential condition coincides with that of yet another kind of static stability: continuous stability (Eshel and Motro 1981). Thus, an essentially necessary and sufficient condition for an equilibrium strategy to be stable is that it is a CSS. Geometrically, the differential condition means that, at the equilibrium point, the reaction (or best-response) curve intersects the forty-five degree line from above rather than below. In asymmetric games (Section 5), the differential condition for stability of an equilibrium y is negative definiteness of a particular square matrix $H(y)$ whose entries are second-order partial derivatives of the payoff functions. If the negative definiteness condition also holds for every other strategy profile, then the equilibrium is moreover unique.

In asymmetric games with unidimensional strategy spaces, static stability can be directly compared with certain kinds of dynamic stability, each of which corresponds to a particular law of motion, which specifies how the players' behavior changes in response to incentives to move (Section 6). In particular, for the dynamics where the rate of change of each player's strategy is proportional to the marginal payoff, the condition for asymptotic stability of an equilibrium y is *D*-stability of the same matrix $H(y)$ mentioned above. As *D*-stability of a square matrix is implied by negative definiteness but not conversely, this kind of dynamic stability is a weaker requirement than static stability. The same is not true for the dynamics where two players alternate in best responding to each other's strategy. Asymptotic stability with respect to these dynamics is not implied by static stability or vice versa. The conclusion, then, is that even in this simple kind of games, the comparison between static and dynamic stability is specific to the particular variety of the latter examined. This, of course, is hardly surprising. Unlike static stability, which only depends on intrinsic properties of the game, specifically, the players' incentives (which are expressed by their payoff functions), dynamic stability is defined with respect to an extraneous factor, the selected law of motion.

The reliance of static stability wholly on incentives makes it particularly suitable for comparative statics analysis, in particular, study of the welfare effects of altruism and spite (Section 7). Whether people in a group where everyone shares such sentiments are likely to fare better or worse than where people are indifferent to the others' payoffs strongly depends on the static stability or instability of the corresponding equilibria or equilibrium strategies (Milchtaich 2012 ,2021). If these are stable, then social welfare is likely to increase with increasing altruism or decreasing spite, but if they are (definitely) unstable, the effect goes in the opposite direction. Thus, 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, holds. However, this is so only if 'stability' refers to the particular notion of static stability presented in this paper. The principle may not hold for other kinds of stability. In particular, asymptotic stability with respect to the continuous-time replicator dynamics does not preclude a negative relation between altruism and social welfare, and instability does not preclude a positive relation.

2 Definitions and essential properties

In an N-player game h, each player i has a strategy space X_i and a payoff function $h_i: X \longrightarrow \mathbb{R}$, where $X = \prod_i X_i$ is the space of all strategy profiles. Given two strategy profiles, $x = (x_1, x_2, \dots, x_N)$ and $y = (y_1, y_2, \dots, y_N)$, and a permutation π of $(1,2, \dots, N)$, consider the path from y to x in which the players change their strategies in the order specified by π: player $\pi(1)$ moves first, from $y_{\pi(1)}$ to $x_{\pi(1)}$ (which may also be the same strategy), then player $\pi(2)$ moves, and so on. Summation of the movers' changes of payoff and averaging over the set Π of all permutations give the expression

$$P_y(x) \coloneqq \frac{1}{N!} \sum_{\pi \in \Pi} \sum_{j=1}^{N} \Big(h_{\pi(j)}(y \mid x_{\{\pi(1),\pi(2),\dots,\pi(j)\}}) - h_{\pi(j)}(x \mid y_{\{\pi(j),\pi(j+1),\dots,\pi(N)\}}) \Big), \qquad (1)$$

where $y \mid x_S$ denotes the strategy profile where the players in the set S play according to the strategy profile x and those outside S play according to y. (In $x \mid y_S$, the strategy profiles are interchanged.) Expression (1) may be interpreted as the overall incentive to move from y to x.

Definition 1 For a strategy profile y in a game h, the *y-rooted potential* is the function $P_y: X \longrightarrow \mathbb{R}$ defined by (1).

The somewhat unwieldy expression (1) can be put into simpler forms. For $N = 2$, it can be rearranged to read

$$P_y(x) = \frac{1}{2}\Big(\big(h_1(x_1, x_2) - h_1(y_1, x_2) + h_1(x_1, y_2) - h_1(y_1, y_2)\big) \\ + \big(h_2(x_1, x_2) - h_2(x_1, y_2) + h_2(y_1, x_2) - h_2(y_1, y_2)\big)\Big).$$

A general alternative form, which also suggests an alternative interpretation of rooted potential, can be obtained as follows. For strategy profiles x and y, define the *payoff of x-players when playing against y-players* as the quantity

$$\mathcal{H}(x, y) \coloneqq \sum_{j=1}^{N} \left[\frac{1}{\binom{N}{j}} \sum_{\substack{S \\ |S|=j}} \bar{h}_S(y \mid x_S) \right] = \sum_{S} \frac{1}{\binom{N}{|S|}} \bar{h}_S(y \mid x_S) = \sum_{S} \frac{1}{\binom{N}{|S|}} \bar{h}_{S^{\mathrm{c}}}(x \mid y_S), \qquad (2)$$

where $|S|$ denotes the number of players in a set S and $\bar{h}_S = (1/|S|) \sum_{i \in S} h_i$ is their average payoff, which is defined as 0 if $S = \emptyset$. (The third equality in (2) is obtained by replacing the summation variable S with the complementary set S^{C} and using the identity $x \mid y_S = y \mid x_{S^{\mathsf{C}}}$.) Note that the expression in square brackets is the average of $\bar{h}_S(\, y \mid x_S\,)$ over all sets of players S of size j. Thus, $\mathcal{H}(x, y)$ is equal to N times the expected payoff for an x-player, when the size of the set of x-players, the actual set, and the particular member examined are chosen at random one after the other, with uniform distributions. The proof of the next lemma shows that this expression is equal to the "positive part" of the y-rooted potential, while the "negative part" is obtained by interchanging x and y. Thus, the overall incentive to move from y to x is equal to the difference between the payoff of x-players when playing against y-players and the payoff of y-players when playing against x-players.

Lemma 1 For strategy profiles x and y, $P_y(x) = \mathcal{H}(x, y) - \mathcal{H}(y, x)$.

Proof. $\mathcal{H}(x, y)$ and $\mathcal{H}(y, x)$ are equal, respectively, to the "positive part" of (1), which is obtained by only considering the minuend in the parentheses, and to the "negative part", which only considers the subtrahend. To see this, note that in each part all payoffs have the form $h_i(\, y \mid x_S\,)$ or $h_i(\, x \mid y_S\,)$, with $i \in S$. Specifically, i is given by the equation $i = \pi(j)$, and S is given by the equation $S = \{\pi(1), \pi(2), \dots, \pi(j)\}$ in the positive part and by $S = \{\pi(j), \pi(j+1), \dots, \pi(N)\}$ in the negative part. In both parts, for every pair (S, i) with $i \in S$ there are precisely $(|S| - 1)!\,(N - |S|)!$ pairs (π, j) satisfying the two equations (as j is uniquely determined by $|S|$). Therefore, the positive part of (1) is equal to

$$\sum_{S \neq \emptyset} \sum_{i \in S} \frac{(|S| - 1)!\,(N - |S|)!}{N!} h_i(\, y \mid x_S\,) = \mathcal{H}(x, y)$$

and the negative part is given by a similar expression where x and y are interchanged. ■

Of particular interest are strategy profiles y such that the overall incentive to move to any nearby strategy profile x is negative. That is, when the players move one-by-one to x, their moves on average harm them. This property is referred to as stability of y. In view of Lemma 1, stability also means that, when players only play according to x or according to y, those doing the former tend to fare worse.

Definition 2 A strategy profile y in an N-player game h is *stable*, *weakly stable* or *definitely unstable* if the y-rooted potential $P_y(x)$ is negative, nonpositive or positive, respectively, for all $x \neq y$ in some neighborhood of y. If a similar condition holds for *all* $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

Stability is a local concept. It refers to neighborhood systems of strategy profiles, in other words, to a topology on X.[4] That topology is the product topology: the product of the topologies on the players' strategy spaces. In principle, these topologies need to be explicitly specified, but in practice, they can often be understood from the context. This is so when there is a unique natural topology on each strategy space X_i.[5] In a game with a finite number of strategies, it may seem natural to consider the discrete topology on each strategy space,

[4] A subset of X is a *neighborhood* of a point x if its interior includes x.

[5] In particular, the latter applies when it is natural to view X_i as a subspace of a Euclidean space or some other standard topological space, so that its topology is the relative one. For example, if the strategy space is an interval in the real line $\mathbb{R}$, so that strategies are simply (real) numbers, a set of strategies is a neighborhood of a strategy y_i if and only if, for some $\varepsilon > 0$, every $x_i \in X_i$ with $|x_i - y_i| < \varepsilon$ is in the set.

that is, to view strategies as isolated. However, a more useful choice of topology in a finite game is the trivial, or indiscrete, topology. This choice effectively puts topology out of the way, since it means that the only neighborhood of any strategy is the entire strategy space. The trivial topology may be used also with an infinite X. Stability, weak stability or definite instability of a strategy profile y with respect to the trivial topology automatically implies it with respect to *any* topology, as it coincides with the global version of the property.

It follows immediately from the definition (by considering in (1) only strategy profiles that differ from y in a single coordinate) that every globally weakly stable strategy profile is a (Nash) equilibrium and every globally stable strategy profile is a strict equilibrium. (Note that there can be at most one globally stable strategy profile.) A strategy profile y that is stable but not globally so is still a *strict local equilibrium* in the sense that it has a neighborhood where for every $x \neq y$ the inequality

$$h_i(y) - h_i(\, y \mid x_i \,) > 0 \tag{3}$$

holds for every player i, where $y \mid x_i$ denotes the strategy profile that differs from y only in that player i uses strategy x_i. However, a stable y is not necessarily a strict equilibrium or even an equilibrium. The reverse implications also do not hold. As the next example shows, even in a two-player game, a strict equilibrium is not necessarily even weakly stable.

Example 1 *Games in the plane.* Players 1 and 2 have the same strategy space, the real line $\mathbb{R}$ (with the standard topology). Their payoff functions are

$$h_1(x_1, x_2) = -x_1^2 + 3x_1x_2 \text{ and } h_2(x_1, x_2) = -\frac{1}{2}x_2^2 - x_1x_2. \tag{4}$$

It is not difficult to see that the origin is the unique equilibrium, and it is moreover a strict equilibrium. When the players move from $(0,0)$ to any other strategy profile (x_1, x_2), player 1 may be the first to change his strategy, and in this case, the sum of the movers' payoff increments is $-x_1^2 - x_2^2/2 - x_1x_2$. If player 2 move first, the sum is $-x_2^2/2 - x_1^2 + 3x_1x_2$. Averaging the two sums gives $P_{(0,0)}(x_1, x_2) = -x_1^2 + x_1x_2 - x_2^2/2 = -(x_1 - x_2/2)^2 - x_2^2/4$. As the last expression is negative for all $(x_1, x_2) \neq (0,0)$, the equilibrium is globally stable. However, in the game obtained by dropping the second term in h_2, where the payoff functions are

$$h_1(x_1, x_2) = -x_1^2 + 3x_1x_2 \text{ and } h_2(x_1, x_2) = -\frac{1}{2}x_2^2, \tag{5}$$

$P_{(0,0)}(x_1, x_2) = -x_1^2 + 3x_1x_2/2 - x_2^2/2$. This expression is positive for every $(x_1, x_2) \neq (0,0)$ that is a multiple of $(2,3)$, and so the strict equilibrium $(0,0)$ is not even weakly stable.

2.1 Symmetric games

Symmetric N-player games differ from the *asymmetric* games considered above is that the players share a single strategy space X and a single payoff function $g: X^N \longrightarrow \mathbb{R}$ that is invariant to permutations of its second through Nth arguments. If one player uses strategy x and the other players use $y, z, \dots, w$, in any order, the first player's payoff is $g(x, y, z, \dots, w)$. A strategy y is a (symmetric Nash) *equilibrium strategy* if it is a best response to itself, that is,

$$g(y, y, \dots, y) \geq g(x, y, \dots, y) \tag{6}$$

for every strategy $x \neq y$. It is a *strict equilibrium strategy* if these inequalities are all strict.

The definition of rooted potential in symmetric games is conceptually similar to that in asymmetric games; it is the overall incentive to move from the root y to an alternative point x. However, x and y here are strategies rather than strategy profiles. The initial and finite strategies profiles are the *symmetric* ones in which everyone uses x or y, and so the overall incentive to move is given by

$$F_y(x) \coloneqq \sum_{j=1}^{N} \left(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right). \tag{7}$$

Definition 3 For a strategy y in a symmetric game (with payoff function[6]) g, the *y-rooted potential* is the function $F_y: X \longrightarrow \mathbb{R}$ defined by (7).

As in asymmetric games, the rooted potential in symmetric games can be put into an alternative form, which also suggests a somewhat different interpretation of this concept. For strategies x and y, define the *payoff of x-players when playing against y-players* as the quantity

$$\mathcal{G}(x, y) \coloneqq \sum_{j=1}^{N} g(\underbrace{x, \dots, x}_{j \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}),$$

which is N times the expected payoff for an x-player, when the size of the (nonempty) set of x-players is chosen at random, with uniform distribution.

Lemma 2 For strategies x and y, $F_y(x) = \mathcal{G}(x, y) - \mathcal{G}(y, x)$.

Proof. $\mathcal{G}(x, y)$ is equal to (7) with only the minuend in the parentheses considered. Considering instead only the subtrahend, changing the summation variable from j to $N - j + 1$, and using the invariance of g to permutations of its second through Nth arguments give $\mathcal{G}(y, x)$. ■

The definition of stability in symmetric games is a straightforward adaptation of Definition 2. A strategy y is stable if, when the players move one-by-one to any nearby strategy x, their moves on average harm them. In view of Lemma 2, stability can also be interpreted as the condition that, if players only choose either x or y, those doing the former tend to fare worse .

Definition 4 A strategy y in a symmetric N-player game g is *stable*, *weakly stable* or *definitely unstable* if the y-rooted potential $F_y(x)$ is negative, nonpositive or positive, respectively, for all $x \neq y$ in some neighborhood of y. If a similar condition holds for *all* $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

In some classes of symmetric games, stability of a strategy automatically implies that it is an equilibrium strategy. Specifically, Section 4.3 below shows that this is so for the mixed extensions of finite games, and in particular for symmetric $n \times n$ games (which is the two-player case). In some other kinds of games, the reverse implication holds. In particular, an equilibrium strategy is automatically globally stable in every symmetric game with $N \geq 2$ players that satisfies the *symmetric substitutability* condition (see Milchtaich 2012, Section 6): for all strategies $x, y, z, \dots, w$ with $x \neq y$,

---

[6] Here and below, the same symbol is used for the payoff function in a symmetric game and for the game itself.

$$g(x, x, z, \dots, w) - g(y, x, z, \dots, w) < g(x, y, z, \dots, w) - g(y, y, z, \dots, w). \tag{8}$$

The condition implies that the summand in (7) strictly decreases as j increases from 1 to N, and so if y is an equilibrium strategy, which by (6) means that the summand is nonpositive for $j = 1$, then the whole sum is negative, which means that y is globally stable. *Weak symmetric substitutability*, which differs in that the strict inequality in (8) is replaced by a weak inequality, similarly implies that every equilibrium strategy is globally weakly stable.

In general, however, the equilibrium condition and the stability condition are incomparable: neither of them implies the other. An equilibrium strategy is not necessarily even weakly stable, and even a globally stable strategy is not necessarily an equilibrium (or even local equilibrium[7]) strategy. A *stable equilibrium strategy* is a strategy that satisfies both conditions. It is not difficult to see that in the special case of symmetric two-player games, where the equilibrium condition is the bivariate version of (6) and the stability condition $F_y(x) < 0$ can be rearranged to read

$$g(x, x) - g(y, x) + g(x, y) - g(y, y) < 0, \tag{9}$$

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

$$pg(x, x) + (1 - p)g(x, y) < pg(y, x) + (1 - p)g(y, y) \tag{10}$$

holds for all $0 < p \leq 1/2$. This condition means that the alternative strategy 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.

2.1.1 Symmetrization

Definitions 3 and 4 differ from 1 and 2 in that they apply to symmetric rather than asymmetric games and concern strategies rather than strategy profiles. Nevertheless, the two pairs of definitions are conceptually very similar and, as shown below, the latter can be formally derived from the former. The link between them is provided by the (standard) notion of symmetrization of an asymmetric game (Milchtaich 2012).

An asymmetric N-player game h is *symmetrized* by letting the players switch roles, with all possible permutations considered. This gives a symmetric N-player game g where the players' common strategy space is the space X of all strategy profiles in h. For a player in g, a strategy $x = (x_1, x_2, \dots, x_N) \in X$ specifies the strategy x_i the player will use when called to assume the role of any player i in h. In that role, the payoff is according to i's payoff function h_i. The player's payoff in g is the average of his payoff over all $N!$ possible assignments of players in g to roles in h. Thus, for any N strategies in X, $x^1 = (x_1^1, x_2^1, \dots, x_N^1)$, $x^2 = (x_1^2, x_2^2, \dots, x_N^2)$, …, $x^N = (x_1^N, x_2^N, \dots, x_N^N)$,

$$g(x^1, x^2, \dots, x^N) = \frac{1}{N!} \sum_{\pi \in \Pi} h_{\pi(1)}(x_1^{\pi^{-1}(1)}, x_2^{\pi^{-1}(2)}, \dots, x_N^{\pi^{-1}(N)}) \tag{11}$$

[7] A strategy y is a *local equilibrium strategy* if it has a neighborhood where every strategy x satisfies (6). In the symmetric two-player game where the strategy space is $[0,1]$ and the payoff function is $g(x, y) = x^2 - 3xy$, the origin 0 is globally stable but is not a local equilibrium strategy as $g(0,0) < g(x, 0)$ for all $x \neq 0$.

$$= \frac{1}{N!} \sum_{\rho \in \Pi} h_{\rho^{-1}(1)}(x_1^{\rho(1)}, x_2^{\rho(2)}, \dots, x_N^{\rho(N)})$$

(where Π is the set of all permutation of $(1,2, \dots, N)$). For $\pi \in \Pi$, player j in g is assigned to role $\pi(j)$ in h, and so the player $\rho(i)$ assigned to role i is $\pi^{-1}(i)$. Note that superscripts in (11) index players' strategies in the symmetric game g while subscripts refer to roles in the asymmetric game h.

Symmetrization preserves the game's equilibria. That is, a strategy profile is an equilibrium in h if and only if it is an equilibrium strategy in g, and in this case, the equilibrium payoff in g is equal to the players' average equilibrium payoff in h. To see this, note first that a strategy profile y is an equilibrium strategy in g if and only if setting $x^2 = x^3 = \cdots = x^N = y$ in (11) gives an expression that is maximized by also choosing $x^1 = y$. An alternative, simpler form of that expression is obtained by partitioning the set of permutations Π into N parts, each with cardinality $(N-1)!$, according to the value i of $\pi(1)$, which gives

$$\frac{1}{N} \sum_{i=1}^{N} h_i(y \mid x_i^1).$$

Clearly, choosing $x^1 = y$ maximizes the last expression if and only if, for each i, the ith term in the sum is maximized by choosing $x_i^1 = y_i$. The latter is also the condition for y to be an equilibrium in h. If the condition holds, then the maximum (obtained by setting $x^1 = y$) is both the equilibrium payoff in g and the players' average equilibrium payoff in h.

The next proposition shows that symmetrization also preserves the rooted potential. An immediate corollary is that it provides a means for expressing the stability of a strategy profile in an asymmetric game as stability in a symmetric game.

Proposition 1 The rooted potential in an asymmetric N-player game h is equal to the rooted potential in the symmetric game g obtained by symmetrizing h, that is,

$$P_y(x) = F_y(x), \qquad x, y \in X.$$

Proof. If follows from (7) and (11) that the rooted potential in g is given by

$$F_y(x) = \sum_{j=1}^{N} \frac{1}{N!} \sum_{\pi \in \Pi} \left(h_{\pi(1)}(y \mid x_{\{\pi(1),\pi(2),\dots,\pi(j)\}}) - h_{\pi(1)}(x \mid y_{\{\pi(1),\pi(j+1),\pi(j+n),\dots,\pi(N)\}}) \right). \quad (12)$$

Since the inner sum is over the set of all permutations, it is left unchanged by replacing the summation variable π with $\pi \circ \pi_j$, for any (j-specific) permutation π_j. For π_j that is the transposition switching 1 and j, this replacement transforms the expression on the right-hand side of (12) into that in (1). ∎

Theorem 1 A strategy profile y in an asymmetric N-player game h is stable, weakly stable or definitely unstable if and only if y has the same property as a strategy in the game g obtained by symmetrizing h.

2.2 Essentially symmetric games

In symmetric games, a stable strategy is not always a local equilibrium strategy (see footnote 7). This contrasts with the situation for asymmetric games, where, as indicated, a stable strategy profile is always a strict local equilibrium. The difference suggests that, in some sense, stability is a weaker requirement in symmetric games than in asymmetric games.

A direct comparison between the concepts of rooted potential and stability in symmetric and in asymmetric games is provided by the essentially symmetric games. An asymmetric N-player game h is *essentially symmetric* if the players share a common strategy space X and for every strategy profile $(x_1, x_2, \ldots, x_N) \in X^N$ and permutation π of $(1,2, \ldots, N)$

$$h_i(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(N)}) = h_{\pi(i)}(x_1, x_2, \ldots, x_N), \qquad i = 1,2, \ldots, N. \tag{13}$$

Thus, if the players' strategies are shuffled, such that each player i takes the strategy of some player $\pi(i)$, then the latter's old payoff becomes player i's new payoff. In other words, the rules of the game are indifferent to the players' identities and are therefore completely specified by the payoff function of any single player, and in particular by h_1. The latter may be viewed as the payoff function in a symmetric game. In fact, for fixed strategy space X and number of players N, the mapping $h \mapsto h_1$ is a one-to-one correspondence between the set of essentially symmetric games and the set of symmetric games. This fact may suggest that there is little difference between the two concepts other than that, in the former, the players are distinguished as player 1, player 2, etc. And, indeed, essentially symmetric games are usually referred to simply as symmetric games (von Neumann and Morgenstern 1953). However, there is in fact a substantive, non-technical difference between describing a strategic situation as a symmetric game and describing it as an essentially symmetric game, with each alternative corresponding to a different interpretation of the situation. This fact is well recognized in the biological game theory literature, where essential symmetry is referred to by other names such as uncorrelated asymmetry (Maynard Smith and Parker 1976; the correlation referred to is that between the players' traits and their payoff functions) and inessential asymmetry (Eshel 2005). A symmetric pairwise contest with identical contestants, such as two equal-size males seeking to obtain a newly vacated territory, is best modeled as a symmetric game such as Chicken, or the Hawk–Dove game. Precedence or other perceivable asymmetries between the contestants, which do not by themselves change the payoffs (i.e., the stakes or the opponents' fighting abilities), make the contest an essentially symmetric one and, in reality, may significantly affect the contestants' behavior (Maynard Smith 1982; Riechert 1998).

The difference between an essentially symmetric game and the corresponding symmetric game is reflected by the different notions of stability: stability of a strategy profile in the first case and stability of a strategy in the second case. The first notion is more general, in that it is applicable both to symmetric strategy profiles, in which all players use the same strategy, and to asymmetric ones. However, even in the case of a symmetric strategy profile $\vec{y} = (y, y, \ldots, y)$, stability of $\vec{y}$ in the essentially symmetric game and stability of strategy y in the symmetric game are not the same thing. As the next proposition shows, the first requirement is stronger.[8] The reason is that it takes into consideration a larger set of alternatives than the second requirement does. An alternative to y is another (nearby) strategy x, to which all the players switch. The alternatives to $\vec{y}$ include (nearby) strategy profiles that are not symmetric, which means that only some of the players may move to x while the others may move to other strategies or stick with y.

---

[8] For essentially symmetric bimatrix games, a related difference holds for the index and degree of a symmetric equilibrium, which may depend on whether it is viewed as an equilibrium in an asymmetric game or in the corresponding symmetric $n \times n$ one (Demichelis and Germano 2000).

Proposition 2 The rooted potential in an essentially symmetric N-player game h relates to the rooted potential in the corresponding symmetric game g $(= h_1)$ by

$$P_{\vec{y}}(\vec{x}) = F_y(x), \qquad x, y \in X,$$

where $\vec{x} = (x, x, \dots, x)$ and $\vec{y} = (y, y, \dots, y)$. Stability of a symmetric strategy profile $\vec{y}$ in h implies stability of strategy y in g but the reverse implication does not hold even if y is an equilibrium strategy and $N = 2$.

Note that y is an equilibrium strategy in g if and only if $\vec{y}$ is an equilibrium in h . This is because, in a symmetric strategy profile in an essentially symmetric game, a player may gain from a unilateral change of strategy if and only if player 1 would gain from making the same move.

Proof of Proposition 2. For a player i and a set of players S with $i \in S$, let π be a permutation of $(1,2, \dots, N)$ that maps $1,2,3, \dots, |S|$ to the elements of S and, in particular, maps 1 to i (that is, $\pi(1) = i$). By the essential symmetry condition (13),

$$h_i(\vec{y} \mid \vec{x}_S) = h_1(\underbrace{x, \dots, x}_{|S| \text{ times}}, \underbrace{y, \dots, y}_{|S^c| \text{ times}}), \qquad x, y \in X.$$

By this identity, Lemmas 1 and 2, Eq. (2), and the fact that $h_1 = g$,

$$P_{\vec{y}}(\vec{x}) = \mathcal{H}(\vec{x}, \vec{y}) - \mathcal{H}(\vec{y}, \vec{x}) = \sum_{j=1}^{N} \left(h_1(\underbrace{x, \dots, x}_{j \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - h_1(\underbrace{y, \dots, y}_{j \text{ times}}, \underbrace{x, \dots, x}_{N-j \text{ times}}) \right)$$
$$= \mathcal{G}(x, y) - \mathcal{G}(y, x) = F_y(x).$$

If $\vec{y}$ is stable in h, then in particular $P_{\vec{y}}(\vec{x})$ is negative for every $x \neq y$ in some neighborhood of y, which, by the above equalities, implies the same for $F_y(x)$.

The fact that the reverse implication does not hold is proved, for example, by the (unique) equilibrium strategy in the symmetric 2×2 game g with payoff matrix (for the row player)

$$\begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix}.$$

The equilibrium strategy y attaches to the first action probability $1/4$. Any other strategy x attaches to it a different probability, say $1/4 + \epsilon$. It is easy to see that $F_y(x) = -2\epsilon^2$, which shows that y is globally stable. But in the corresponding essentially symmetric bimatrix game

$$\begin{pmatrix} 0 & 1,3 \\ 3,1 & 0 \end{pmatrix},$$

the completely mixed symmetric equilibrium (y, y) is not stable, as it is not a strict local equilibrium: any unilateral deviation leaves the deviator's payoff unchanged. ∎

The counterexample used in the proof of Proposition 2 is but one instance of a general difference between symmetric $n \times n$ games and their asymmetric counterparts, the bimatrix games. As shown in Section 4.2 below, the stable strategy profiles in a bimatrix game are precisely the strict equilibria. By contrast, it is shown in Section 4.3 that the stable strategies in a symmetric $n \times n$ game are the evolutionarily stable strategies. All strict (hence, pure) equilibrium strategies are ESSs, but so are also some strategies that are not pure (such as y in the above example). This is another facet of the weaker meaning of stability in symmetric games than in asymmetric games.

2.3 λ-stability

Stability of a strategy y in a symmetric game g may be interpreted as associated with a particular belief of a player moving from y to an alternative strategy x about the total number of players who will be using x after he switches to it, with the rest using the original strategy y. Namely, the probabilities $\lambda_1, \lambda_2, \dots, \lambda_N$ that this number is $1,2,\dots,N$ are all equal. This interpretation suggests the natural generalization of considering other beliefs. The corresponding extension of the definition of rooted potential is that the equal coefficients in (7) are replaced by possibly unequal coefficients. Thus, for an N-tuple $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) > 0$ (whose entries are nonnegative numbers that are not all zero), define the *y-rooted λ-potential* as the function $F_y^\lambda : X \longrightarrow \mathbb{R}$ given by

$$F_y^\lambda(x) \coloneqq \sum_{j=1}^{N} \lambda_j \left(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right). \tag{14}$$

Definition 5 For an N-tuple $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) > 0$, a strategy y in a symmetric N-player game g is *λ-stable*, *weakly λ-stable* or *definitely λ-unstable* if the y-rooted λ-potential $F_y^\lambda(x)$ is negative, nonpositive or positive, respectively, for all $x \neq y$ in some neighborhood of y.

The special case of λ-stability where only the first coordinate λ_1 is not zero coincides with the condition that y is a strict local equilibrium strategy: any unilateral deviation to a nearby strategy x would lower the payoff. The diametrically opposite case where only λ_N is not zero coincides with the notion of local superiority of y discussed in Section 4.2 below. In an intermediate case, each of the other players switches to x with probability $0 < p < 1$ and stays with y with probability $1 - p$. Depending on whether the players' choices are perfectly correlated (i.e., identical) or independent, the λ's are given, respectively, by

$$\lambda_j = \begin{cases} 1 - p, & j = 1 \\ 0 & , 1 < j < N \\ p & , j = N \end{cases} \tag{15}$$

or

$$\lambda_j = B_{j-1,N-1}(p) \coloneqq \binom{N-1}{j-1} p^{j-1} (1-p)^{N-j}, \qquad j = 1,2,\dots,N, \tag{16}$$

the Bernstein polynomials of degree $N - 1$. In the "mid-point" case $p = 1/2$, for both (15) and (16)

$$\lambda_j = \lambda_{N-j+1}, \qquad j = 1,2,\dots,N. \tag{17}$$

These equalities express a belief that the number of other players using strategy x and the number using y have the same distribution. Put differently, the joint distribution of the two numbers is symmetric. A strategy y is *dependently-* or *independently-stable* if it is λ-stable with $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ given by (15) or (16), respectively, for *all* $0 < p < 1$, and it is *symmetrically-stable* if it is λ-stable for *all* $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) > 0$ satisfying (17). The last requirement obviously implies stability.

For single-player games ($N = 1$), stability and λ-stability of a strategy y mean the same thing, namely, strict local optimality. For $N = 2$, stability does not generally imply λ-stability (or vice versa) but the implication does partially hold (specifically, holds whenever $\lambda_1 \geq \lambda_2 > 0$) if y is an equilibrium strategy (see (10)). A full appreciation of the differences between stability in the sense of Definition 4 and the varieties based on λ-stability requires looking at multiplayer games. An important class of such games is examined in Section 4.3.1.

2.4 Population games

A (symmetric) *population game*, as defined in this paper, is formally a symmetric two-player game such that the strategy space X is a convex set in a (Hausdorff real) linear topological space (for example, the unit simplex in a Euclidean space $\mathbb{R}^n$) and the payoff function $g(x, y)$ is continuous in the second argument y for all $x \in X$. However, a population game is interpreted not as representing an interaction between two specific players but as one involving an (effectively) infinite population of identical individuals who are "playing the field".[9] This means that an individual's payoff $g(x, y)$ depends only on his own strategy x and on the suitably defined *population strategy* y. The latter may be, for example, the population's *mean* strategy with respect to some nonatomic (population) measure, which attaches zero mass to every individual. Alternatively, it may describe the *distribution* of strategies in the population (Bomze and Pötscher 1989), and in this case, X consists of *mixed strategies*, that is, probability measures on some underlying space of allowable actions or (pure[10]) strategies, and $g(x, y)$ is linear in x and expresses the expected payoff for an individual whose choice of action is random with distribution x.[11]

In a population game, the equilibrium condition

$$g(y, y) = \max_{x \in X} g(x, y), \tag{18}$$

which is formally obtained from (6) for $N = 2$, also admits more than one interpretation. It may mean that, in a *monomorphic* population where everyone plays strategy y, single individuals cannot increase their payoff by choosing any other strategy x. Alternatively, for y that describes the population's mean strategy or distribution of strategies, and a payoff function g that is linear in the first argument, Eq. (18) may express the condition that (almost) everyone in the population is using a strategy that is a best response to the population strategy y. In other words, the possibly *polymorphic* population is in an equilibrium state.

Examples of population games are nonatomic congestion games with a continuum of identical users, and public good games with an infinite population of identical agents who have to decide whether to contribute their private good for the production of some public good (Milchtaich 2012, 2021). Another important example is the following one (Bomze and Weibull 1995; Broom et al. 1997).

Example 2 *Random matching in a symmetric N-player game with a multilinear payoff function*. N players are picked up independently and according to the same distribution (i.i.d.) from an infinite population of potential players, whose individual probability of being selected is zero. The strategy space X is a convex set in a linear topological space, and the payoff function g is continuous and is linear in each of the N arguments. (This assumption

---

[9] An infinite population may represent the limiting case of an increasingly large population, with the effect of each player's action on each of the other players correspondingly decreasing. Alternatively, it may represent all possible characteristics of players, or *potential* players, when the number of *actual* players is finite.

[10] "Pure" and "mixed" are relative terms. In particular, a pure strategy may itself be a probability vector.

[11] The special case of this interpretation in which the set of pure strategies is finite is the one often referred to as 'population game' (Sandholm 2015). The meaning of the term in this paper is more general, both in terms of the formal setup and in the variety of possible interpretations.

may be relaxed by dropping the linearity requirement for the first argument.) Because of the multilinearity of g, a player's expected payoff only depends on his own strategy x and on the population's mean strategy y. Specifically, the expected payoff is given by

$$\bar{g}(x,y) \coloneqq g(x,y,\dots,y). \tag{19}$$

This equation defines a population game $\bar{g}$ with the strategy space X. It is easy to see that a strategy y is an equilibrium strategy in $\bar{g}$ if and only if it is an equilibrium strategy in the underlying symmetric N-player game g.

Rooted potential in population games is defined by a variant of Definition 3 that replaces the numbers of players using strategies x and y with the sizes of the subpopulations to which each strategy applies, p and $1-p$ respectively. Correspondingly, the sum in (7) is replaced by the integral:

$$\Phi_y(x) \coloneqq \int_0^1 \big(g(x, px+(1-p)y) - g(y, px+(1-p)y)\big)\, dp. \tag{20}$$

Definition 6 For a strategy y in a population game g, the *y-rooted potential* is the function $\Phi_y\colon X \longrightarrow \mathbb{R}$ defined by (20).

Stability is defined similarly to stability in symmetric games (Definition 4).

Definition 7 A strategy y in a population game g is *stable*, *weakly stable* or *definitely unstable* if the y-rooted potential $\Phi_y(x)$ is negative, nonpositive or positive, respectively, for all $x \neq y$ in some neighborhood of y. If a similar condition holds for *all* strategies $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

For a payoff function g that is linear in the second argument, the integral in (20) is equal to $1/2$ times the sum in (7) (with $N = 2$). The equality means that rooted potential and stability have essentially the same meanings whether g is viewed as a two-player symmetric game or as a population game. The example in Footnote 7 thus shows that, even in this special kind of population games, a stable, and even globally stable, strategy need not be an equilibrium strategy.

In a population game $\bar{g}$ that is derived from a symmetric N-player game g as in Example 2, the payoff function is not linear in the second argument if $N \geq 3$. Nevertheless, the next proposition shows that in this case, too, expressions (20) and (7) differ only by a positive multiplicative constant, which means that the stability of a strategy y does not depend on whether it is viewed as a strategy in g or in $\bar{g}$.

Proposition 3 The rooted potential in a symmetric N-player game g where a the strategy space X is a convex set in a linear topological space and the payoff function is continuous and multilinear is equal to N times the rooted potential in the population game $\bar{g}$ defined by Eq. (19), that is,

$$F_y(x) = N\,\Phi_y(x), \qquad x, y \in X.$$

Proof. It follows from the invariance of the payoff function g to permutations of its second through Nth arguments and its linearity in these arguments that, for every pair of strategies x and y and every $0 \leq p \leq 1$, the strategy $x_p = px + (1-p)y$ satisfies

$$\bar{g}(x,x_p) - \bar{g}(y,x_p) = g(x,x_p,\dots,x_p) - g(y,x_p,\dots,x_p)$$
$$= \sum_{j=1}^{N} B_{j-1,N-1}(p)\left(g(x,\underbrace{x,\dots,x}_{j-1\text{ times}},\underbrace{y,\dots,y}_{N-j\text{ times}}) - g(y,\underbrace{x,\dots,x}_{j-1\text{ times}},\underbrace{y,\dots,y}_{N-j\text{ times}})\right), \tag{21}$$

where the coefficients in the last sum are the Bernstein polynomials defined in (16). These polynomials satisfy the equalities

$$\int_0^1 B_{j-1,N-1}(p)\,dp = \frac{1}{N}, \qquad j = 1,2,\dots,N. \tag{22}$$

It therefore follows from (21) by integration that the expression obtained from (20) by replacing g with $\bar{g}$ is equal to $1/N$ times the sum in (7). ∎

3 Potential games

An asymmetric N-player game h is a *potential game* (Monderer and Shapley 1996) if it has an (*exact*) *potential*: a real-valued function $P: X \longrightarrow \mathbb{R}$, defined on the set of strategy profiles, such that whenever a single player i changes his strategy, the resulting change in i's payoff is equal to the change in P. Thus, for all i,

$$h_i(x) - h_i(x \mid x_i') = P(x) - P(x \mid x_i'), \qquad x \in X, x_i' \in X_i. \tag{23}$$

Fixing x_i' and rearranging (23) to read

$$h_i(x) = P(x) + \big(h_i(x \mid x_i') - P(x \mid x_i')\big), \qquad x \in X$$

gives that, in a potential game, the payoff of each player i is equal to the sum of the potential and some function of the other players' strategies. An immediate corollary is that the potential is unique up to an additive constant.

The potential can also be characterized in terms of the rooted potential, and vice versa.

Proposition 4. For an asymmetric N-player game h, a function $P: X \longrightarrow \mathbb{R}$ is a potential if and only if

$$P_y(x) = P(x) - P(y), \qquad x, y \in X. \tag{24}$$

A necessarily and sufficient condition for the rooted potential to be a potential is that it changes only by an additive constant whenever the root changes.

Proof. The sufficiency of condition (24) follows from (23) being the special case $y = x \mid x_i'$. Necessity follows from (1) and the definition of potential, which give

$$P_y(x) = \frac{1}{N!}\sum_{\pi\in\Pi}\sum_{j=1}^{N}\Big(P(y \mid x_{\{\pi(1),\pi(2),\dots,\pi(j)\}}) - P(x \mid y_{\{\pi(j),\pi(j+1),\dots,\pi(N)\}})\Big) = P(x) - P(y).$$

If the rooted potential only changes by an additive constant whenever the root changes, then, for any fixed root z, the identify (24) holds with $P = P_z$. ∎

Proposition 4 shows that, in potential games, the potential and rooted potential essentially coincide. This finding immediate gives the following result, which characterizes stability and definite instability in such games in terms of the extremum points of the potential.

Theorem 2 A strategy profile y in an asymmetric game with a potential P is stable, weakly stable or definitely unstable if and only if y is, respectively, a strict local maximum, local maximum or strict local minimum point of P. A *global* maximum point of P is both globally weakly stable (and if it is a strict global maximum point, globally stable) and an equilibrium.

3.1 Potential in symmetric games

Potential in symmetric N-player games has essentially the same meaning as in asymmetric games. The only difference is that, here, the potential is necessarily a symmetric function (meaning that it is invariant under permutations of its N arguments). Thus, for a symmetric game g with strategy space X, a symmetric function $F: X^N \longrightarrow \mathbb{R}$ is a potential if, for any $N+1$ strategies $x, x', y, z, \dots, w$,

$$g(x, y, z, \dots, w) - g(x', y, z, \dots, w) = F(x, y, z, \dots, w) - F(x', y, z, \dots, w).$$

The potential F of a symmetric N-player potential game g may itself be viewed as the payoff function in a symmetric N-player game, with the same strategy space X. This symmetric game is moreover *doubly symmetric* in the sense that its payoff function is symmetric. In other words, it is a symmetric common-payoff game. Its set of equilibria coincides with that of g.

It follows immediately from the definition that a necessary condition for the existence of a potential in a symmetric game g is that the total change of payoff of any two players who change their strategies one after the other does not depend on the order of their moves. That is, for any $N+2$ strategies $x, x', y, y', z, \dots, w$,

$$\begin{aligned} g(x, y, z, \dots, w) - g(x', y, z, \dots, w) + g(y, x', z, \dots, w) - g(y', x', z, \dots, w) \\ = g(y, x, z, \dots, w) - g(y', x, z, \dots, w) + g(x, y', z, \dots, w) - g(x', y', z, \dots, w). \end{aligned}$$

It is not difficult to show that this condition is also sufficient (see Monderer and Shapley 1996, Theorem 2.8, which however refers to asymmetric games). Moreover, if a symmetric N-player game is the mixed extension of a finite game, then it is a potential game if and only if the above condition holds for all $N+2$ *pure* strategies. In this case, the potential, like the payoff function, is multilinear (see Monderer and Shapley 1996, Lemma 2.10).

Example 3 *Symmetric* 2×2 *games.* Every symmetric 2×2 game g, with pure strategies 1 and 2, is a potential game, since it is easy to see that it satisfies the above condition for pure strategies. It is moreover not difficult to check that the following bilinear function (whose arguments are mixed strategies $x = (x_1, x_2)$ and $y = (y_1, y_2)$, with $x_1 + x_2 = y_1 + y_2 = 1$) is a potential:

$$F(x, y) = \big(g(1,1) - g(2,1)\big)x_1 y_1 + \big(g(2,2) - g(1,2)\big)x_2 y_2. \tag{25}$$

As in asymmetric games, the rooted potential in symmetric potential games is closely related to the potential. Indeed, as the following proposition shows, it is essentially equal to the potential "along the diagonal". This, again, gives a very simple characterization of stability and instability in terms of the potential.

Proposition 5. For a symmetric N-player game g with a potential F,

$$F_y(x) = F(x, x, \dots x) - F(y, y, \dots y), \qquad x, y \in X.$$

Proof. By (7) and the definition of potential,

$$F_y(x) = \sum_{j=1}^{N} \left(F(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - F(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right)$$
$$= \sum_{j=1}^{N} \left(F(\underbrace{x, \dots, x}_{j \text{ times}}, y, \dots, y) - F(\underbrace{x, \dots, x}_{j-1 \text{ times}}, y, \dots, y) \right) = F(x, x, \dots x) - F(y, y, \dots y),$$

where the second equality uses the symmetry of F. ■

Theorem 3 A strategy y in a symmetric N-player game with a potential F is stable, weakly stable or definitely unstable if and only if y is, respectively, a strict local maximum, local maximum or strict local minimum point of the function $x \mapsto F(x, x, \dots x)$. If $(y, y, \dots, y)$ is a global maximum point of F itself, then y is in addition an equilibrium strategy.

The similarity between Theorem 3 and Theorem 2 is related to the fact, established by the next proposition, that symmetrization (Section 2.1.1) maps asymmetric potential games to symmetric potential games. Indeed, it essentially maps potentials to potentials.

Proposition 6 An asymmetric N-player game h has a potential if and only if the symmetric game g obtained by symmetrizing h has a potential.

Proof. If the game h, where the set of all strategy profiles is X, has a potential P, then the symmetric function $F: X^N \longrightarrow \mathbb{R}$ defined by

$$F(x^1, x^2, \dots, x^N) = \frac{1}{N!} \sum_{\rho \in \Pi} P(x_1^{\rho(1)}, x_2^{\rho(2)}, \dots, x_N^{\rho(N)})$$

is a potential for the game g (where X is the strategy space). This is because, by (11), for $x^1, x^2, \dots, x^N$ and y in X,

$$\begin{aligned} & g(x^1, x^2, \dots, x^N) - g(y, x^2, \dots, x^N) = \\ & = \frac{1}{N!} \sum_{\rho \in \Pi} \left(h_{\rho^{-1}(1)}(x_1^{\rho(1)}, \dots, x_{\rho^{-1}(1)}^1, \dots, x_N^{\rho(N)}) - h_{\rho^{-1}(1)}(x_1^{\rho(1)}, \dots, y_{\rho^{-1}(1)}, \dots, x_N^{\rho(N)}) \right) \\ & = \frac{1}{N!} \sum_{\rho \in \Pi} \left(P(x_1^{\rho(1)}, \dots, x_{\rho^{-1}(1)}^1, \dots, x_N^{\rho(N)}) - P(x_1^{\rho(1)}, \dots, y_{\rho^{-1}(1)}, \dots, x_N^{\rho(N)}) \right) \\ & = F(x^1, x^2, \dots, x^N) - F(y, x^2, \dots, x^N). \end{aligned}$$

Conversely, if g has a potential F, then it follows from Propositions 1, 4 and 5 that the function P defined by

$$P(x) = F(x, x, \dots, x)$$

is a potential for h. ■

3.2 Potential in population games

For population games, which represent interactions involving many identical players whose individual actions have negligible effects on the other players, the definition of potential may be naturally adapted by replacing the increment of the potential with a derivative.

Definition 8 For a population game g with strategy space X, a continuous function $\Phi: X \longrightarrow \mathbb{R}$ is a *potential* if, for all $x, y \in X$ and $0 < p < 1$, the derivative on the left-hand side of the following equality exists and the equality holds:

$$\frac{d}{dp}\Phi(px+(1-p)y) = g(x, px+(1-p)y) - g(y, px+(1-p)y). \tag{26}$$

It is easy to see that the potential is unique up to an additive constant.

Example 4 *Symmetric* 2×2 *games, viewed as population games.* It is easy to check that, for every such game, with pure strategies 1 and 2, the function Φ defined (for a mixed strategy $x = (x_1, x_2)$, with $x_1 + x_2 = 1$) by

$$\Phi(x) = \frac{1}{2}\big(g(1,1) - g(2,1)\big)x_1^2 + \frac{1}{2}\big(g(2,2) - g(1,2)\big)x_2^2 \tag{27}$$

satisfies (26) and is thus a potential. Note that, unlike the bivariate function F defined in (25), Φ has a single strategy as an argument. The two functions are however connected by the identity $\Phi(x) = 1/2\ F(x,x)$.

The term potential is borrowed from physics, where it refers to a scalar field whose gradient gives the force field. Force is analogous to incentive here. The analogy can be taken one step further by presenting the payoff function g as the differential of the potential Φ. This requires Φ to be defined not only on the strategy space X (which by definition is a convex set in a linear topological space) but on its *cone* $\hat{X}$, which is the collection of all space elements that can be written as a strategy x multiplied by a positive number t. For example, if strategies are probability measures, Φ needs to be defined for (or extended to) all positive, non-zero finite measures. The *differential* of the potential is then defined as its directional derivative, that is, as the function $d\Phi: \hat{X}^2 \longrightarrow \mathbb{R}$ given by

$$d\Phi(\hat{x}, \hat{y}) = \left.\frac{d}{dt}\right|_{t=0^+} \Phi(t\hat{x} + \hat{y}) \tag{28}$$

(where the direction is specified by the first argument $\hat{x}$). The differential exists if the (right) derivative in (28) exists for all $\hat{x}, \hat{y} \in \hat{X}$.

Lemma 3 Let $\Phi: \hat{X} \longrightarrow \mathbb{R}$ be a continuous function on the cone of a convex set X in a linear topological space. If the differential $d\Phi: \hat{X}^2 \longrightarrow \mathbb{R}$ exists and is continuous in the second argument, then it is necessarily linear in the first argument and satisfies

$$\frac{d}{dp}\Phi(px+(1-p)y) = d\Phi(x, px+(1-p)y) - d\Phi(y, px+(1-p)y), \qquad x, y \in X, 0 < p < 1. \tag{29}$$

Proof (an outline). Using elementary arguments, the following can be established.

FACT. A continuous real-valued function defined on an open real interval is continuously differentiable if and only if it has a continuous right derivative.

Suppose that $d\Phi$ satisfies the condition specified by the lemma. Replacing $\hat{y}$ in (28) with $p\hat{x} + \hat{y}$ gives

$$d\Phi(\hat{x}, p\hat{x} + \hat{y}) = \left.\frac{d}{dt}\right|_{t=p^+} \Phi(t\hat{x} + \hat{y}), \qquad \hat{x}, \hat{y} \in \hat{X}, p \geq 0. \tag{30}$$

By the above Fact and the continuity properties of Φ and $d\Phi$, for $0 < p < 1$ the right derivative in (30) is actually a two-sided derivative and it depends continuously on $\hat{y}$. Therefore, the right-hand side of (29) is equal to the expression

$$\left.\frac{d}{dt}\right|_{t=p} \Phi(tx + (1-p)y) - \left.\frac{d}{dt}\right|_{t=1-p} \Phi(px + ty),$$

which by the chain rule is equal to the derivative on the left-hand side. Hence, (29) holds.

The fact that the right derivative in (30) is actually a two-sided derivative also implies that, for $t \geq 0$,

$$\int_0^t d\Phi(\hat{x}, p\hat{x} + \hat{y})\, dp = \Phi(t\hat{x} + \hat{y}) - \Phi(\hat{y}), \qquad \hat{x}, \hat{y} \in \hat{X}.$$

This result, used twice, gives that for $\lambda, t > 0$

$$\begin{aligned}\int_0^{\lambda t} \big(d\Phi(\hat{z}, p\hat{z} + \lambda t\hat{x} + \hat{y}) + d\Phi(\hat{x}, p\hat{x} + \hat{y})\big)\, dp \\ = \Phi(\lambda t\hat{z} + \lambda t\hat{x} + \hat{y}) - \Phi(\lambda t\hat{x} + \hat{y}) + \Phi(\lambda t\hat{x} + \hat{y}) - \Phi(\hat{y}) \\ = \Phi(t(\lambda\hat{z} + \lambda\hat{x}) + \hat{y}) - \Phi(\hat{y}), \qquad \hat{x}, \hat{y}, \hat{z} \in \hat{X}.\end{aligned}$$

Dividing the right- and left-hand sides by t and taking the limit $t \to 0$ gives the identity

$$\lambda\, d\Phi(\hat{z}, \hat{y}) + \lambda\, d\Phi(\hat{x}, \hat{y}) = d\Phi(\lambda\hat{z} + \lambda\hat{x}, \hat{y}), \qquad \hat{x}, \hat{y}, \hat{z} \in \hat{X}.$$

Since the identity holds for all $\lambda > 0$, it proves that $d\Phi$ is linear in the first argument. ∎

Lemma 3 immediately gives the following proposition, which may also be interpreted as an alternative definition of potential in population games.[12] While this definition is not as general as Definition 8, it is more familiar. In particular, the definition of potential in Sandholm (2015) is a special case, pertaining to games where the strategies are mixed strategies over a finite set of actions, so that the strategy space X is the unit simplex in a Euclidean space.

Proposition 7 For a population game g, let $\Phi: \hat{X} \longrightarrow \mathbb{R}$ be a continuous function on the cone of the strategy space X such that the differential $d\Phi: \hat{X}^2 \longrightarrow \mathbb{R}$ exists and is continuous in the second argument. If

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

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

As for symmetric and asymmetric potential games, potential in population games is closely linked with the rooted potential. In fact, the next proposition shows that the two are essentially equal. This coincidence means that stability and instability (here, in the sense of Definition 7) of a strategy y are equivalent to y being a local extremum point of the potential.

Proposition 8. For a population game g with a potential Φ,

$$\Phi_y(x) = \Phi(x) - \Phi(y), \qquad x, y \in X.$$

Proof. Follows from (20) and (26) by integration. ∎

[12] The proposition may also be looked at from the opposite perspective. As it shows, any suitable real-valued function Φ on a suitable set X is a potential for some population game. The payoff function in that game is the restriction of $d\Phi$ to X^2. By Lemma 3, it is necessarily linear in the first argument.

Theorem 4 A strategy y in a population game with a potential Φ is stable, weakly stable or definitely unstable if and only if y is, respectively, a strict local maximum, local maximum or strict local minimum point of Φ. In the first two cases, y is in addition an equilibrium strategy. If the potential Φ is strictly concave, then an equilibrium strategy is necessarily a strict *global* maximum point of Φ, is *globally* stable, and is therefore the game's unique stable strategy.

Proof. The first part of the theorem follows immediately from Proposition 8.

It follows from (26), in the limit $p \to 0$, that the payoff function g satisfies

$$\left.\frac{d}{dp}\right|_{p=0^+} \Phi(px + (1-p)y) = g(x, y) - g(y, y) \tag{31}$$

for all x and y. If y is a local maximum point of Φ, then the left-hand side of (31) is nonpositive for all x, which proves that y is an equilibrium strategy.

To prove the last part of the theorem, consider an equilibrium strategy y and any other strategy x. The right-, and therefore also the left-, hand side of (31) is nonpositive. If Φ is strictly concave, this conclusion implies that the left-hand side of (26) is negative for all $0 < p < 1$, which proves that y is a strict global maximum point of Φ. It then follows from the first part of the theorem that y is globally stable. ■

Since a potential is by definition a continuous function, an immediate corollary of Theorem 4 is the following result, which concerns the existence of strategies that are at least weakly stable.

Corollary 1 If a population game with a potential Φ has a compact strategy space, then it has at least one weakly stable strategy. If in addition the number of such strategies is finite, they are all stable.

4 Comparison with earlier notions of static stability

The analogy between incentives and physical forces alluded to in the last subsection can be extended to a general perspective on rooted potential and stability. Incentives in a game are specified by the payoff functions, which at every strategy profile y tell how the payoff of each player i would change as a result of a unilateral deviation: from x_i to any alternative strategy x_i'. In the special case of a potential game, Eq. (23) expresses these incentives as corresponding changes in a single function, the potential P. This is analogous to a conservative force, whose field can be written as the (negative) gradient of a potential. Rooted potential corresponds to the general case of a possibly non-conservative force, where the work done by the force may be path dependent. This generalization of potential still provides useful information about the players' incentives at and around the root. The root is defined as stable if the incentives to move away from it tend to be negative. This kind of stability is *static* in the sense that it does not involve any assumptions about the players' reactions to the incentives. These reactions are not specified by the game itself and would therefore require extraneous assumptions concerning *dynamics*. The physical analog is the static stability of a system that is at an equilibrium state. Being at equilibrium means that there is no (net) force pushing the system towards a different state. Stability differs in also considering the forces acting at states that are (usually, only slightly) different from the one under consideration and, roughly speaking, requiring that these forces push the system in the direction of that state. For example, a ball at the bottom of a pit is stable but one at the top of a hill is not. In both cases, the net force acting on the ball vanishes, but any displacement would result in a non-zero force that is directed towards the equilibrium point in the first case and away from it in the second case. This description is static rather than kinetic and therefore does not require invoking such concepts as inertia and Newton's second law, which concern dynamics.

Static stability in games may be novel as a general concept but is hardly new in content. In fact, as shown below, a number of well-established concepts in game theory fall into this category. These concepts are not all as generally applicable as stability in the sense of this paper is, in that they refer to structures on strategy spaces or payoff functions that exist only in certain classes of games. This raises the question of the relations between these "native" notion of static stability and the general one proposed here. Specifically, the question is whether the former are equivalent to the stability concept obtained by restricting to the class of games under consideration the general Definition 2, 4 or 7 (for asymmetric, symmetric or population games, respectively). To the extent that they exist, such equivalences allow viewing these earlier notions of static stability as special cases of this paper's stability, thus elevating the later to the status of a unifying notion which turns static stability from a general idea to a concrete universal concept.

4.1 Risk dominance

In every symmetric or population game, every isolated strategy is trivially stable. Therefore, if the strategy space X has the discrete topology, that is, all singletons are open sets, then all strategies are stable. The definition of stability is therefore of interest only for games with non-discrete strategy spaces. This category includes games with a finite number of strategies where the topology on X is the trivial one, so that stability and definite instability mean *global* stability and definite instability. The simplest nontrivial such game is a finite symmetric two-player game with only two strategies, for example, the game with payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} a & b \end{array} \\ \begin{array}{c} a \\ b \end{array} & \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix} \end{array}.$$

Here, both strategies are equilibrium strategies. Strategy a is globally stable and strategy b is globally definitely unstable, because (using the form (9) of the stability condition)

$$5 - 4 + 1 - 3 < 0 < 3 - 1 + 4 - 5.$$

The two inequalities, which are obviously just rearrangements of one another, have an additional meaning. Namely, they express the fact that (a, a) is the *risk dominant* equilibrium (Harsanyi and Selten 1988). It is not difficult to see that this coincidence of global stability and risk dominance holds in general – it is not a special property of the payoffs in this example.

Proposition 9. In a finite symmetric two-player game with two strategies, an equilibrium strategy y is globally stable if and only if the equilibrium (y, y) is risk dominant.

The risk dominant equilibrium strategy is the (unique) globally stable strategy not only in the finite game itself but also in its mixed extension, where mixed strategies are allowed. This is because the necessary and sufficient condition for a pure strategy y to be a globally stable equilibrium strategy is the same in both games. The condition is that inequality (10) holds for all $x \neq y$ and $0 < p \leq 1/2$ (with g denoting both the payoff function in the finite game and that in the mixed extension, which is the bilinear extension of the first function). In the finite game, the only strategy $x \neq y$ is the other pure strategy, z. In the mixed extension, x can be any convex combination of y and z other than z itself. However, it follows easily from the bilinearity of g that such a convex combination satisfies (10) for all $0 < p \leq 1/2$ if and only if z does so.

Where a finite symmetric two-player game differ from its the mixed extension is in the natural topology on the strategy space X. In the mixed extension, X is in effect the unit interval, for which the natural topology is the standard topology, not the trivial one. Stability with respect to the standard topology is a weaker condition than global stability. In particular, it follows from Theorem 6 below that if the two pure strategies are strict equilibrium strategies (as they are in the above example), then in the mixed extension they are both stable.

4.2 Local superiority

Another notion of static stability in symmetric and population games, which is similar to stability in being a local condition, is local superiority (or strong uninvadability; Bomze 1991).

Definition 9 A strategy y in a symmetric N-player game or population game g is *locally superior* if it has a neighborhood where for every strategy $x \neq y$

$$g(y, x, \dots, x) > g(x, x, \dots, x). \tag{32}$$

Local superiority differs from the strict equilibrium condition is that, whereas the strict version of inequality (6) expresses a disincentive to be the first player to switch from y to an alternative strategy x, inequality (32) expresses a player's disincentive to be the last to do so. In the former case, all the other players are using y, and in the latter, they all use x. Stability differs from both concepts in also considering all the intermediate cases, in which some of the other players play x and some play y. This difference means that even a locally superior strict equilibrium strategy is not necessarily stable. Consider, for example, the finite symmetric three-player game where the players have to choose between playing L, which gives 1, and playing R, which gives 2 or -2 if the number of players choosing it is odd or even, respectively. Strategy R is a strict equilibrium strategy and is locally superior, yet it is definitely unstable because $F_R(L) = (-1) + 3 + (-1) > 0$.

The general rule that local superiority does not imply stability have some notable exceptions. For equilibrium strategies in symmetric two-player games, the implication does hold, because summing the bivariate versions of inequalities (6) and (32) gives (9). Local superiority implies stability also in a large class of population games, which includes the "classical" population games mentioned in Footnote 11.

Proposition 10 In a population game where the payoff function g is linear in the first argument, every locally superior strategy y is stable.

Proof. Let U be any neighborhood of a strategy y. By basic properties of linear topological space, the mapping $(p, x) \mapsto px + (1-p)y$ is continuous. For any $0 \leq p_0 \leq 1$, it maps (p_0, y) to an element of U (namely, y). Therefore, there is a neighborhood $V(p_0)$ of p_0 in $[0,1]$ and a neighborhood $W(p_0)$ of y in the strategy space such that $px + (1-p)y \in U$ for all $p \in V(p_0)$ and $x \in W(p_0)$. As $[0,1]$ is compact, there is a finite set of points in this interval such that the corresponding V's cover the interval. The intersection of the corresponding W's is a sub-neighborhood of U where every x satisfies $px + (1-p)y \in U$ for all $0 \leq p \leq 1$. It follows from this finding that, if y is locally superior, then it has a neighborhood where for every $x \neq y$

$$\frac{1}{p}\big(g(px + (1-p)y, px + (1-p)y) - g(y, px + (1-p)y)\big) < 0$$

for all $0 < p \leq 1$. If g is linear in the first argument, then the expression on the left-hand side of the inequality is equal to the integrand in (20), and therefore the integral is negative, that is, $\Phi_y(x) < 0$. ■

In both exceptions mentioned above, where local superiority does imply stability, the converse does not generally hold. Indeed, it is shown in Section 4.4 below (Example 6 and the paragraph following it) that, in symmetric two-player games and in population games with bilinear payoff functions, even a globally stable strict equilibrium strategy is not necessarily locally superior.

Another class of games where local superiority implies stability is the mixed extensions of finite symmetric N-player games. Here, too, the converse does not generally hold. However, it does hold, and the two conditions are therefore equivalent, in the two-player case ($N = 2$), which is the symmetric $n \times n$ games. (Note that these games are also a special case of the two previous exceptions.) These results are proved in Section 4.3.

Local superiority can be naturally extended to the asymmetric case by defining a strategy profile y in an asymmetric game h as locally superior when it has that property as a strategy in the game obtained by symmetrizing h. It is, however, not difficult to see that this definition has a simple direct formulation. It is equivalent to the condition that, when the players move one-by-one from y to any nearby strategy profile x, the last mover *on average* loses.

Definition 10 A strategy profile y in an asymmetric N-player game h is *locally superior* if it has a neighborhood where for every strategy profile $x \neq y$

$$\frac{1}{N}\sum_{i=1}^{N}\big(h_i(x) - h_i(x \mid y_i)\big) < 0. \tag{33}$$

As indicated (see Section 2), a stable strategy profile in an asymmetric game is necessarily a strict local equilibrium but not conversely. Thus, the condition that the *first* mover to any nearby strategy profile is harmed by the move is weaker than stability. As it turns out, local superiority, which concerns the last mover, is a stronger condition than stability.

Proposition 11 In asymmetric N-player games, every locally superior strategy profile is stable but not conversely.

Proof. A locally superior strategy profile y has a rectangular neighborhood where inequality (33) holds for all $x \neq y$. In that neighborhood, a similar inequality holds with x replaced by $y \mid x_S$, for any set of players S such that the strategy profile $y \mid x_S$ is different from y. Division by $\binom{N-1}{|S|-1}$ and summation over all nonempty sets S give

$$0 > \sum_{S \neq \emptyset} \frac{1}{\binom{N-1}{|S|-1}} \frac{1}{N} \sum_{i \in S}\Big(h_i(y \mid x_S) - h_i(y \mid x_{S \setminus \{i\}})\Big)$$

$$= \sum_{S \neq \emptyset} \frac{1}{\binom{N}{|S|}} \frac{1}{|S|} \sum_{i \in S} h_i(y \mid x_S) - \sum_i \sum_{\substack{S \\ i \in S}} \frac{1}{\binom{N-1}{|S \setminus \{i\}|}} \frac{1}{N} h_i(y \mid x_{S \setminus \{i\}})$$

$$= \sum_S \frac{1}{\binom{N}{|S|}} \bar{h}_S(y \mid x_S) - \sum_i \sum_{\substack{S \\ i \notin S}} \frac{1}{\binom{N-1}{|S|}} \frac{1}{N} h_i(y \mid x_S) = \mathcal{H}(x,y) - \sum_S \frac{1}{\binom{N}{|S|}} \bar{h}_{S^c}(y \mid x_S)$$

$$= \mathcal{H}(x,y) - \mathcal{H}(y,x),$$

which by Lemma 1 proves that y is stable.

Example 1 shows that even global stability is not a sufficient condition for local superiority. In the two-player game with payoff functions (4), the strict equilibrium $y = (0,0)$ is globally stable but it is not locally superior because the left-hand side of (33) is equal to the expression $-x_1^2/2 + x_1 x_2 - x_2^2/4$, which is positive if $x_1 = x_2 \neq 0$. ■

An important class of games in which the converse in Proposition 11 does hold is mixed extensions of finite asymmetric games. These asymmetric N-player games are the multiplayer generalizations of bimatrix games. The strategy space X_i of each player i is the unit simplex in a Euclidean space $\mathbb{R}^{n_i}$ and the payoff function h_i is multilinear. Strategies are interpreted as *mixed* strategies, that is, as probability vectors that assign probabilities to the player's possible actions in the underlying finite game. The set of all actions that are assigned positive probability is the *support* of the strategy, and a strategy is pure or completely mixed if its support contains only one action or all actions, respectively.

Theorem 5 In an asymmetric N-player game h that is the mixed extension of a finite game, the following properties of a strategy profile are equivalent:

(i) The strategy profile is stable.
(ii) The strategy profile is locally superior.
(iii) The strategy profile is a strict equilibrium.

Proof. (i) ⟹ (iii). If a strategy profile y is stable, then for every player i inequality (3) holds for every strategy $x_i \neq y_i$ in some neighborhood of y_i. Therefore, for *all* $x_i \neq y_i$, a similar inequality in which x_i is replaced with $\epsilon x_i + (1-\epsilon) y_i$ holds for sufficiently small $\epsilon > 0$.[13] However, by the linearity of h_i in player i's own strategy, that inequality is actually equivalent to (3), which proves that y is a strict equilibrium.

(iii) ⟹ (ii). Suppose that y is a strict equilibrium, so that (3) holds for all i and $x_i \neq y_i$. For each player i, let Z_i be the collection of all strategies whose support does not include that of y_i (in other words, strategies that have at least one zero component that is nonzero in y_i). This is a compact subset of X_i that does not include y_i, and therefore the expression on the left-hand side of (3) is bounded away from zero for $x_i \in Z_i$. Thus, there is some $\delta_i > 0$ such that

$$h_i(y) - h_i(y \mid z_i) \geq \delta_i, \qquad z_i \in Z_i.$$

It follows, since Z_i is compact, that there is a neighborhood of y where for every strategy profile x

$$h_i(x) - h_i(x \mid z_i) \geq \delta_i/2, \qquad z_i \in Z_i. \tag{34}$$

[13] A condition holds for "sufficiently small" $\epsilon > 0$ if there is some $\delta > 0$ such that the condition holds for all $0 < \epsilon < \delta$.

For every strategy $x_i \neq y_i$ there is a unique $0 < \epsilon_i \leq 1$ (which depends on x_i) such that for some (indeed, a unique) $z_i \in Z_i$

$$x_i = (1 - \epsilon_i)y_i + \epsilon_i z_i.$$

By the linearly of h_i in the ith coordinate, this equation and (34) imply that $(1 - \epsilon_i)\big(h_i(x) - h_i(x \mid y_i)\big) = \epsilon_i\big(h_i(x \mid z_i) - h_i(x)\big) < 0$. The conclusion proves that there is a neighborhood of y where (33) holds for every $x \neq y$.

(ii) ⟹ (i). A special case of Proposition 11. ■

4.3 Evolutionary stability

In symmetric two-player games and population games, by far the best-known kind of static stability is evolutionary stability (Maynard Smith 1982). The following formulation is applicable to games where the strategy space is endowed with a linear structure, so that convex combinations of strategies are well defined.

Definition 11 A strategy y in a symmetric two-player game or population game g 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

$$g(y, \epsilon x + (1 - \epsilon)y) > g(x, \epsilon x + (1 - \epsilon)y) \tag{35}$$

or a similar weak inequality, respectively, holds. An ESS or NSS *with uniform invasion barrier* satisfies the stronger condition obtained by interchanging the two logical quantifiers: for sufficiently small $\epsilon > 0$ (which cannot vary with x), (35) or a similar weak inequality, respectively, holds for all $x \neq y$.

In population games, the difference between stability in the sense of ESS and in the sense of Definition 7 boils down to a different meaning of proximity between population strategies. The definition of ESS may be interpreted as reflecting the view that a population strategy is close to y when another strategy x replaces y in a small subpopulation, of size ϵ. In Definition 7, by contrast, the subpopulation to which x applies need not be small but x itself is assumed close to y.

The kind of games to which the concept of evolutionary stability is most often applied is symmetric $n \times n$ games, in which the payoff function g can be expressed by a square (payoff) matrix A of these dimensions. With strategies written as column vectors,

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

The game may be viewed either as a symmetric two-player game or as a population game. In the former case, the relevant definition of stability is Definition 4, and in the latter, Definition 7. However, as shown in Section 2.4, the linearity of g in the second argument means that the two definitions of stability actually coincide, and similarly for weak stability and definite instability. Moreover, as the next two results show, stability is also equivalent to evolutionary stability and to local superiority. It also follows from these results that, in symmetric $n \times n$ games, every (even weakly) stable strategy is an equilibrium strategy and every strict equilibrium strategy is stable.

The first result is rather well known (Bomze and Pötscher 1989; van Damme 1991, Theorem 9.2.8; Weibull 1995, Propositions 2.6 and 2.7; Bomze and Weibull 1995).

Proposition 12 For a strategy y in a symmetric $n \times n$ game g, the following conditions are equivalent:

(i) Strategy y is an ESS.

(ii) Strategy y is an ESS with uniform invasion barrier.

(iii) Local superiority: for every strategy $x \neq y$ in some neighborhood of y,

$$g(y, x) > g(x, x). \tag{36}$$

(iv) For every $x \neq y$, the (weak) inequality $g(y, y) \geq g(x, y)$ holds (which means that y is an equilibrium strategy), and if it holds as equality, then (36) also holds.

An NSS is characterized by similar equivalent conditions in which the strict inequality (36) is replaced by a weak inequality.

A completely mixed equilibrium strategy y in a symmetric $n \times n$ game g is said to be *definitely evolutionarily unstable* (Weissing 1991) if the reverse of inequality (36) holds for all $x \neq y$.

Theorem 6 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.

Proof. The inequality in condition (iv) in Proposition 12 and inequality (36) in condition (iii) together imply (9), and the same is true with the strict inequalities (36) and (9) both replaced by their weak versions. This proves that a sufficient condition for stability or weak stability of a strategy y is that it is an ESS or an NSS, respectively. For a completely mixed equilibrium strategy y, the inequality in (iv) automatically holds as equality for all x, and therefore a similar argument proves that a sufficient condition for definite instability of y is that it is definitely evolutionarily unstable.

In remains to prove necessity. For a stable strategy y, inequality (9) holds for all nearby strategies $x \neq y$. Therefore, y has the property that, for *every* strategy $x \neq y$, for sufficiently small $\varepsilon > 0$

$$\begin{aligned} g(\varepsilon x + (1-\varepsilon)y, \varepsilon x + (1-\varepsilon)y) - g(y, \varepsilon x + (1-\varepsilon)y) \\ + g(\varepsilon x + (1-\varepsilon)y, y) - g(y, y) < 0. \end{aligned} \tag{37}$$

It follows from the bilinearity of the payoff function that (37) is equivalent to

$$(2-\varepsilon)\big(g(y, y) - g(x, y)\big) + \varepsilon\big(g(y, x) - g(x, x)\big) > 0. \tag{38}$$

Therefore, the above property of y is equivalent to condition (iv) in Proposition 12, which proves that y is an ESS. Similar arguments show that a weakly stable strategy is an NSS and that a definitely unstable completely mixed equilibrium strategy is definitely evolutionarily unstable. In the first case, the proof needs to be modified only by replacing the strict inequalities in (36), (37) and (38) by weak inequalities, and in the second case (in which the first term in (38) vanishes for all x), they need to be replaced by the reverse inequalities. ∎

The theorem is illustrated by the following result.

Corollary 2 In a symmetric 2×2 game g, with pure strategies 1 and 2, a (mixed) strategy $y = (y_1, y_2)$ is an ESS or an NSS if and only if it is, respectively, a strict local maximum or local maximum point of the quadratic function Φ defined by (27).

Proof. Example 4 shows that Φ is a potential for g, when the latter is viewed as a population game. The corollary therefore follows from Theorems 4 and 6. An alternative proof uses Example 3, which presents the bivariate function F defined by (25) as a potential for g, when viewed as a symmetric two-player game. As $F(x,x) = 2\Phi(x)$, the corollary follows from Theorems 3 and 6. ∎

The corollary and its proof shed light on a qualitative difference between symmetric 2×2 games, which are potential games, and 3×3 or larger games, which are not generally so. The former always have at least one NSS, as every global maximum point of Φ is one (see also Corollary 1). For the latter, it is well known that this is not so. One counterexample is a variant of the rock–paper–scissors game where a draw yields a small positive payoff for both players (see Maynard Smith 1982, p. 20).

Comparison of Theorem 6 with Theorem 5 shows a qualitative difference between symmetric $n \times n$ games and the corresponding asymmetric games, which are the bimatrix, or $m \times n$, games. In the former, the stable strategies are the evolutionarily stable strategies. In the latter, a strategy profile is stable if and only if it is a strict (hence, pure) equilibrium. By Theorem 1 in Section 2.1.1, the strict equilibria in a bimatrix game h are also the stable strategies in the symmetric game g obtained by symmetrizing h. This conclusion is similar, and closely related, to the well-known fact that a strategy in g is an ESS if and only if it is a strict equilibrium in h (Selten 1980). The similarity reflects (indeed, it proves) the fact that, in the symmetric game obtained by symmetrizing a bimatrix game, a strategy is stable if and only if it is an ESS. Thus, in this respect such a game is similar to a symmetric $n \times n$ game, although it is generally *not* an $n \times n$ game, for any n.

4.3.1 Multiplayer games

Extending the results obtained in the two-player case ($n \times n$ games) to the mixed extensions of finite symmetric N-player games requires a corresponding multiplayer generalization of ESS. That is, evolutionary stability needs to be defined for any symmetric N-player game g with a strategy space X that is the unit simplex in a Euclidean space and a multilinear payoff function. By Proposition 3, *stability* of a strategy y in such a game is equivalent to stability of y in the population game $\bar{g}$ defined by (19), and the same is easily seen to be true for local superiority. Therefore, a natural way to define *evolutionary stability* is to require it to have a similar property. This requirement yields the following natural extension of Definition 11.

Definition 12 A strategy y in the mixed extension g of a finite symmetric N-player game is an *evolutionarily stable strategy* (*ESS*) if, for every strategy $x \neq y$, for sufficiently small $\epsilon > 0$ the strategy $x_\epsilon = \epsilon x + (1 - \epsilon)y$ satisfies

$$g(y, x_\epsilon, \dots, x_\epsilon) > g(x, x_\epsilon, \dots, x_\epsilon). \tag{39}$$

An ESS *with uniform invasion barrier* satisfies the stronger condition that, for sufficiently small $\epsilon > 0$, inequality (39) holds for all $x \neq y$.

Note that for the existence of a uniform invasion barrier it suffices that the last condition holds for *some* $0 < \epsilon < 1$, as this automatically implies the same for any smaller ϵ.

An equivalent definition of ESS is given by a generalization of condition (iv) in Proposition 12.

Lemma 4 (Broom et al. 1997; see also the proof of Lemma 5 below) A strategy y in the mixed extension g of a finite symmetric N-player game is an ESS if and only if, for every $x \neq y$, at least one of the N terms in (7) is not zero and the first such term is negative. In particular, an ESS is necessarily an equilibrium strategy (as the first term in (7) must be nonpositive).

Unlike in the special two-player case (Proposition 12), in the mixed extension of a finite symmetric multiplayer game not every ESS has a uniform invasion barrier. It is easy to see that a sufficient condition for the existence of such a barrier is that the ESS is locally superior, and this condition is in fact also necessary (Bomze and Weibull 1995, Theorem 3; Lemma 6 below). This raises the question of how stability (in the sense of Definition 4) compares with the two nonequivalent versions of ESS. As the following theorem shows, it is equivalent to neither of them, and instead occupies an intermediate position: weaker than one and stronger than the other. The two ESS conditions are also comparable with the stronger stability conditions derived from λ-stability (Section 2.3). In fact, two of the latter turn out to be equivalent to ESS with uniform invasion barrier.

Theorem 7. In the mixed extension g of a finite symmetric game with $N \geq 2$ players, the following implications and equivalences among the possible properties of a strategy y hold:

symmetrically-stable ⟹ dependently-stable ⟺ independently-stable
⟺ locally superior ⟺ ESS with uniform invasion barrier ⟹ stable ⟹ ESS.

Each of the three implications is actually an equivalence in the two-player case but not in general. All the properties imply that y is an equilibrium strategy.

The proof of Theorem 7 uses the next two lemmas, which hold for every game g as in the theorem.

Lemma 5 For a nonnegative vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ with $\lambda_N > 0$, every λ-stable strategy y is an ESS.

Proof. For distinct strategies x and y and $0 < \epsilon < 1$,

$$\sum_{k=1}^{N} \lambda_k \left(g(x_\epsilon, \underbrace{x_\epsilon, \dots, x_\epsilon}_{k-1 \text{ times}}, \underbrace{y, \dots, y}_{N-k \text{ times}}) - g(y, \underbrace{x_\epsilon, \dots, x_\epsilon}_{k-1 \text{ times}}, \underbrace{y, \dots, y}_{N-k \text{ times}}) \right)$$
$$= \sum_{k=1}^{N} \lambda_k \, \epsilon \sum_{j=1}^{k} B_{j-1,k-1}(\epsilon) \left(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right)$$
$$= \sum_{j=1}^{N} \left(\sum_{k=j}^{N} \binom{k-1}{j-1} (1-\epsilon)^{k-j} \lambda_k \right) \left(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right) \epsilon^j.$$

The last sum is negative for sufficiently small $\epsilon > 0$ if and only if at least one of its N terms is not zero and the first such term (that is, the nonzero term ending with the smallest power of ϵ) is negative. Observe that the sign of each term is completely determined by the sign of the second expression in parentheses (the difference). The first expression (the inner sum) is necessarily positive, as by assumption $\lambda_N > 0$. This observation proves that if y is λ-stable, then the condition in Lemma 4 holds. Parenthetically, note that in the special case $\lambda_N = 1$ the observation also proves Lemma 4 itself. ∎

The next lemma uses the following terminology. A strategy y is *conditionally locally superior* if it has a neighborhood where inequality (32) holds for every strategy $x \neq y$ that satisfies the reverse of inequality (6).

Lemma 6 For any $0 < p < 1$, the following properties of an equilibrium strategy y are equivalent:

(i) Strategy y is locally superior.
(ii) Strategy y is conditional locally superior.
(iii) Strategy y is λ-stable with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ given by (15).
(iv) Strategy y is λ- stable with $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ given by (16).
(v) Strategy y is an ESS with uniform invasion barrier.

Proof. The implication (i) ⟹ (iii) follows from the fact that inequality (6) (from the equilibrium condition) and inequality (32) (local superiority) together imply that, for λ given by (15),

$$F_y^\lambda(x) = (1-p)\big(g(x, y, \ldots, y) - g(y, y, \ldots, y)\big) + p\big(g(x, x, \ldots, x) - g(y, x, \ldots, x)\big) < 0.$$

The implication (iii) ⟹ (ii) follows from the fact that, if the last inequality holds and the first term on its left-hand side is nonnegative, then the second term must be negative.

To prove that (ii) ⟹ (i), assume that this implication does not hold. The assumption means that strategy y is not locally superior but is conditionally locally superior, which implies that there is a sequence $(x_k)_{k\geq 1}$ of strategies converging to y such that for all k

$$g(y, x_k, \ldots, x_k) - g(x_k, x_k, \ldots, x_k) \leq 0 \tag{40}$$

and

$$g(y, y, \ldots, y) - g(x_k, y, \ldots, y) > 0.$$

The last inequality means that, when all the other players use y, strategy x_k is not a best response. Therefore, the strategy can be presented as

$$x_k = \alpha_k z_k + (1-\alpha_k) w_k, \tag{41}$$

where $0 < \alpha_k \leq 1$, z_k is a strategy whose support includes only pure strategies that are not best responses when everyone else uses the equilibrium strategy y, and w_k is a strategy that is a best response, i.e.,

$$g(y, y, \ldots, y) - g(w_k, y, \ldots, y) = 0. \tag{42}$$

Since there are only finitely many pure strategies, there is some $\delta > 0$ such that for all k

$$g(y, y, \ldots, y) - g(z_k, y, \ldots, y) > 2\delta. \tag{43}$$

By (40), (41), (42) and (43), for all k

$$\big(g(x_k, x_k, \ldots, x_k) - g(x_k, y, \ldots, y)\big) - \big(g(y, x_k, \ldots, x_k) - g(y, y, \ldots, y)\big) > 2\delta\alpha_k.$$

As $k \to \infty$, the two expressions in parentheses tend to zero, since $x_k \to y$. Therefore, $\alpha_k \to 0$, which by (41) implies that $w_k \to y$. Since y is conditionally locally superior and (42) holds for all k, for almost all k (that is, all $k > K$, for some integer K)

$$g(y, w_k, \ldots, w_k) - g(w_k, w_k, \ldots, w_k) \geq 0.$$

Therefore, by (41), for almost all k

$$\frac{1}{\alpha_k}\big(g(w_k, x_k, \ldots, x_k) - g(y, x_k, \ldots, x_k)\big)$$

$$\leq \frac{1}{\alpha_k}\Big(\big(g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k)\big) + (1-\alpha_k)^{N-1}\big(g(y, w_k, \dots, w_k) - g(w_k, w_k, \dots, w_k)\big)\Big)$$

$$= \sum_{j=2}^{N} \frac{B_{j-1,N-1}(\alpha_k)}{\alpha_k}\left(g(w_k, \underbrace{z_k, \dots, z_k}_{j-1 \text{ times}}, \underbrace{w_k, \dots, w_k}_{N-j \text{ times}}) - g(y, \underbrace{z_k, \dots, z_k}_{j-1 \text{ times}}, \underbrace{w_k, \dots, w_k}_{N-j \text{ times}}) \right),$$

where $B_{j-1,N-1}$ is defined by (16). The last sum tends to zero as $k \to \infty$, since $w_k \to y$. Therefore, for almost all k the expression on the left-hand side is less than δ, so that

$$g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k) < \alpha_k \delta. \tag{44}$$

On the other hand, by (43) and since $x_k \to y$, for almost all k

$$\alpha_k \Big(\big(g(y, y, \dots, y) - g(z_k, y, \dots, y)\big) + \big(g(z_k, y, \dots, y) - g(z_k, x_k, \dots, x_k)\big) + \big(g(w_k, x_k, \dots, x_k) - g(w_k, y, \dots, y)\big)\Big) > \alpha_k \delta.$$

By (41) and (42), the left-hand side of the inequality is equal to $g(w_k, x_k, \dots, x_k) - g(x_k, x_k, \dots, x_k)$, which by (40) is less than or equal to

$$g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k).$$

This contradicts (44). The contradiction proves that (ii) ⟹ (i).

To prove that (i) ⟹ (iv), suppose that y is locally superior, which means that it has a *convex* neighborhood U where (32) holds for every strategy $x \neq y$. The convexity of U and the linearity of g in the first argument imply that, for every $x \in U \setminus \{y\}$, the strategy $x_p = px + (1-p)y$ satisfies

$$g(y, x_p, \dots, x_p) > g(x, x_p, \dots, x_p). \tag{45}$$

By the second equality in (21), inequality (45) is equivalent to $F_y^\lambda(x) < 0$, with $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ given by (16).

The proof of the reverse implication (iv) ⟹ (i) is rather similar. As shown, y has property (iv) if and only if it has a neighborhood U such that inequality (45) holds for all strategies $x \neq y$ in U, equivalently, inequality (32) holds for all $x \neq y$ in the set

$$U_p = \{pz + (1-p)y \mid z \in U\}.$$

In this case, y is locally superior, since U_p is also a neighborhood of y. Indeed, for *any* neighborhood U of any strategy y, $\{U_\epsilon\}_{0<\epsilon<1}$ is a base for the neighborhood system of y (see Bomze and Pötscher 1989, Lemma 42; Bomze 1991, Lemma 6).

The special case $U = X$ of the last topological fact gives the equivalence (i) ⟺ (v). A strategy y has a neighborhood where (32) holds for every $x \neq y$ if and only if it has such a neighborhood of the form X_ϵ, for some $0 < \epsilon < 1$. ■

Proof of Theorem 7. By Lemma 5, a strategy that has any of the seven properties in the theorem is an ESS, hence (by Lemma 4) an equilibrium strategy. An immediate corollary of Lemma 6 is that, for an equilibrium strategy, the properties of dependent- and independent stability, local superiority, and ESS with uniform invasion barrier are all equivalent. The special case $p = 1/2$ of the same lemma (specifically, of the implication (iii) ⟹ (i)) shows that symmetric-stability implies local superiority. Finally, local superiority implies stability.

This is because it follows immediately from Definition 9 that a strategy is locally superior if and only if it has this property in the population game $\bar{g}$ defined in Example 2. Proposition 3 gives that the same is true for stability. The implication therefore follows from Proposition 10 (applied to $\bar{g}$).

With only two players ($N = 2$), stability and symmetric-stability are the same, and thus the equivalence of all the properties in the theorem follows from the first part of the proof and Proposition 12. The counterexamples in Example 5 below (where $N = 4$) complete the proof. ■

Example 5 *The mixed extension of a finite symmetric four-player game*. There are three pure strategies, so that the strategy space X consists of all probability vectors $x = (x_1, x_2, x_3)$ (with $x_1 + x_2 + x_3 = 1$). The payoff of a player using strategy x against opponents using strategies $y = (y_1, y_2, y_3)$, $z = (z_1, z_2, z_3)$ and $w = (w_1, w_2, w_3)$ is given by

$$g(x, y, z, w) = \sum_{i,j,k,l=1}^{3} g_{ijkl} x_i y_j z_k w_l.$$

It does not matter which of the other players uses which strategy, since the coefficients $(g_{ijkl})_{i,j,k,l=1}^{3}$ that define the game satisfy the symmetry condition $g_{ijkl} = g_{ij'k'l'}$, for all i and all triplets (j, k, l) and (j', k', l') that are permutations of one another. There are three versions of the game, with different coefficients, as detailed in the following table:

Coefficient	Version 1	Version 2	Version 3
g_{2211}	-2	-18	-4
g_{2221}	0	-16	-4
g_{3221}	4	4	0
g_{2331}	4	20	4
g_{2222}	3	-9	-3
g_{2332}	4	12	2
g_{3333}	-3	-15	-4
g_{2322}	4	4	0

Coefficients that are not listed in the table and cannot be deduced from it by using the symmetry condition are zero. In all three versions of the game, the strategy $y = (1,0,0)$ is an equilibrium strategy, since if all the other players use y, the payoff is zero regardless of the player's own strategy. However, the stability properties of y are different for the three versions.

CLAIM. The equilibrium strategy $y = (1,0,0)$ is an ESS in all three versions of the game, but it is stable only in versions 2 and 3, ESS with uniform invasion barrier (equivalently, locally superior, independently-stable, dependently-stable) only in version 3, and symmetrically-stable in none of them.

In view of Theorem 7, to prove the Claim it suffices to show that y is: (i) an ESS but not stable in version 1, (ii) stable but not independently-stable in version 2, and (iii) independently-stable but not symmetrically-stable in version 3.

In version 1 of the game, the condition that $F_y^\lambda(x)$ (given by (14)) is negative for a strategy $x = (x_1, x_2, x_3)$ reads

$$-2\lambda_2 x_2^2 - 4\lambda_3(x_1 x_2^2 - x_2^2 x_3 - x_2 x_3^2) \\ - 3\lambda_4(2x_1^2 x_2^2 - 4x_1 x_2^2 x_3 - 4x_1 x_2 x_3^2 - x_2^4 - 4x_2^2 x_3^2 + x_3^4 - 4x_2^3 x_3) < 0.$$

Stability corresponds to $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (1,1,1,1)$, for which the inequality simplifies to

$$\frac{7}{16} x_2^2 < (x_2 - \frac{3}{8}(1 - x_1)^2)^2.$$

There are strategies x arbitrarily close to but different from $(1,0,0)$ for which this inequality does not hold. For example, this is so whenever $x_2 = (3/8)(1 - x_1)^2 > 0$. This proves that the equilibrium strategy is not stable. To prove that it is nevertheless an ESS, consider inequality (39), which in the present case simplifies to

$$2x_2^2 < (2x_2 - \epsilon(1 - x_1)^2)^2.$$

For every (fixed) strategy $x \neq (1,0,0)$, this inequality holds for sufficiently small $\epsilon > 0$. Therefore, $(1,0,0)$ is an ESS.

In version 2 of the game, for $\lambda = (1,1,1,1)$ the condition $F_y^\lambda(x) < 0$ simplifies to

$$-\frac{1}{80} x_2^2 < (x_2 - \frac{3}{8}(1 - x_1)^2)^2.$$

This inequality holds for all strategies x other than $(1,0,0)$, and therefore the latter is stable. However, it is not independently-stable, since for $\lambda = (1/8,3/8,3/8,1/8)$ the condition $F_y^\lambda(x) < 0$ simplifies to

$$\frac{1}{10} x_2^2 < (x_2 - \frac{1}{4}(1 - x_1)^2)^2.$$

This inequality does not hold for strategies x with $x_2 = (1/4)(1 - x_1)^2 > 0$, which exist in every neighborhood of $(1,0,0)$.

Finally, in version 3 of the game, for $\lambda = (1/8\ ,3/8,3/8,1/8)$ the condition $F_y^\lambda(x) < 0$ simplifies to

$$-x_3^4 < 3(4x_2 - (x_2 + x_3)^2)^2.$$

This inequality holds for all strategies x other than $(1,0,0)$. Therefore, by Lemma 6 (which implies that, if (iv) holds for $p = 1/2$, then it holds for all $0 < p < 1$), $(1,0,0)$ is independently-stable. However, it is not symmetrically-stable. There are vectors $\lambda > 0$ satisfying (17) for which $F_y^\lambda(x) \nless 0$ for some strategies x arbitrarily close to $(1,0,0)$. For examples, for $\lambda = (1,9,9,1)$, the condition $F_y^\lambda(x) < 0$ simplifies to

$$24x_2^2 - \frac{1}{3} x_3^4 < (8x_2 - (1 - x_1)^2)^2.$$

For strategies x with $x_2 = (1/8)(1 - x_1)^2$, this inequality is equivalent to $(1 - x_1)^4 - 32(1 - x_1)^3 + 384(1 - x_1)^2 - 2048(1 - x_1) > 512$. Hence, it does not hold if x_1 is sufficiently close to 1. This completes the proof of the Claim.

The Claim has some significance beyond the present context. The fact that, in version 2 of the game, the ESS $(1,0,0)$ does not have a uniform invasion barrier and is not locally superior refutes two published results. A theorem of Crawford (1990), which is reproduced by Hammerstein and Selten (1994, Result 7), implies that every ESS in the mixed extension of a finite symmetric game has a uniform invasion barrier. However, there is a known error in the proof of that theorem (Bomze and Pötscher 1993). Theorem 2 of Bukowski and Miekisz (2004) asserts that local superiority and the ESS condition are equivalent even for $N > 2$. However, these authors employ a definition of ESS that is different from that used here (and in other papers) in that it *requires* the existence of a uniform invasion barrier.

4.4 Continuous stability

Continuous stability (Eshel and Motro 1981; Eshel 2005) is another kind of static stability in symmetric and population games, which is applicable when the strategy space is unidimensional.

Definition 13 In a symmetric two-player game or population game g with a strategy space that is a (finite or infinite) interval in the real line $\mathbb{R}$, an equilibrium strategy y is a *continuously stable strategy* (CSS) if it has a neighborhood where, for every strategy $x \neq y$, for sufficiently small $\epsilon > 0$ the inequality

$$g((1-\epsilon)x + \epsilon y, x) > g(x,x) \tag{46}$$

holds while a similar inequality where ϵ is replaced by $-\epsilon$ does not hold.

In other words, a strategy y that satisfies the global condition of being an equilibrium strategy[14] is a CSS if it also satisfies the local condition (known as *m*-stability or convergence stability; Taylor 1989; Christiansen 1991) that every nearby strategy x is *not* a best response to itself, specifically, any small deviation from x towards y, but not in the opposite direction, increases the payoff.

Depending on whether a game g as in Definition 13 is viewed as a symmetric two-player game or as a population game, stability of an equilibrium strategy y is defined either by Definition 4 or by 7. However, as the following theorem shows, in both cases stability is essentially equivalent to y being a CSS, as the two conditions share the same differential form.

Theorem 8 In a symmetric two-player game or population game g with a strategy space that is an interval in the real line, let y be an interior[15] equilibrium strategy such that at the equilibrium point (y, y) the payoff function has continuous second-order partial derivatives.[16] If

$$g_{11}(y,y) + g_{12}(y,y) < 0, \tag{47}$$

then y is stable and a CSS. If the reverse inequality holds, then y is definitely unstable and not a CSS.[17]

---

[14] The original definition of CSS differs slightly from the version given here in that it requires a stronger global condition, which is a version of ESS.

[15] An *interior* strategy is a strategy lying in the interior of the strategy space.

[16] Partial derivatives are denoted here by subscripts. Thus, g_{12} is the mixed second-order partial derivative of the payoff function.

[17] Very similar sufficient conditions for stability and definite instability hold in symmetric multiplayer games. The only difference is that, with $N \geq 2$ players, the expression on the left-hand side of (47) is replaced by $g_{11} + (N-1)g_{12}$ (computed at the equilibrium point $(y, y, \dots, y)$).

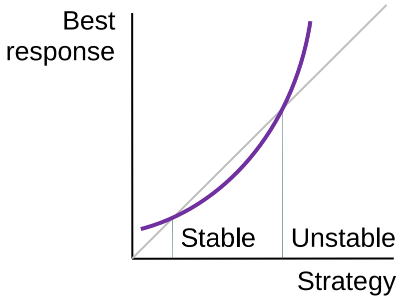

Figure 1. An equilibrium strategy is both stable and a CSS if, at the equilibrium point, the reaction curve (thick line) intersects the forty-five degree line (thin line) from above, and it is definitely unstable and not a CSS if the intersection is from below.

Proof. It is not difficult to show, using Taylor's theorem, that for x tending to y the left-hand side of (9) can be expressed as

$$2g_1(y,y)(x-y) + \big(g_{11}(y,y) + g_{12}(y,y)\big)(x-y)^2 + o((x-y)^2). \quad (48)$$

Similarly, the integral in (20) can be expressed as $1/2$ times (48). Since y is an interior equilibrium strategy, the first term in (48) is zero. Therefore, a sufficient condition for (48) to be negative or positive for every $x \neq y$ in some neighborhood of y (hence, for y to be stable or definitely unstable, respectively) is that $g_{11}(y,y) + g_{12}(y,y)$ has that sign.

Consider now the CSS condition in Definition 13. It may be possible to determine whether this condition holds by looking at the sign of

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \big(g(x,x) - g((1-\ \epsilon)x + \epsilon y, x)\big) = g_1(x,x)(x-y). \quad (49)$$

For x tending to y, the right-hand side of (49) is given by an expression similar to (48) except that it lacks the factor 2 in the first term (which, as indicated, is zero anyway). Therefore, if (47) or the reverse inequality holds, then the expression on the left-hand side of (49) is negative or positive, respectively, for every $x \neq y$ in some neighborhood of y. In the first or second case, (46) holds or does not hold, respectively, for $\epsilon > 0$ sufficiently close to 0 and the converse is true for $\epsilon < 0$. Therefore, in the first case, y is a CSS, and in the second case, it is not a CSS. ■

The differential condition (47) has a simple geometrical interpretation (Eshel 2005). It expresses the familiar condition that, at the (symmetric) equilibrium point, the graph of the best-response function, or reaction curve, intersects the forty-five degree line from above. This follows from the fact that, for an interior equilibrium strategy y, the equilibrium condition (18) implies that $g_1(y,y) = 0$ and $g_{11}(y,y) \leq 0$. If the last inequality is in fact strict, then by the implicit function theorem there is a continuously differentiable function f from some neighborhood of y to the strategy space, with $f(y) = y$, such that $g_1(f(x),x) = 0$ and $g_{11}(f(x),x) < 0$ for all strategies x in the neighborhood. Thus, strategy $f(x)$ is a local best response to x. By the chain rule, at the point y

$$f'(y) = -\frac{g_{12}(y,y)}{g_{11}(y,y)}.$$

Therefore, (47) holds (so that y is stable and a CSS) or the reverse inequality holds (y definitely unstable and not a CSS) if and only if the slope of the function f at y is less or greater than 1, respectively.[18] In the first case, the reaction curve, which is the graph of f (see Figure 1), intersects the forty-five degree line from above (which means that the (local) fixed point index is $+1$; see Dold 1980), and in the second case, the intersection is from below (and the fixed point index is -1).

As Theorem 8 only establishes the identity of the differential conditions for stability and continuous stability, it leaves some "cracks" where the coincidence between stable strategy and CSS may break down. An example of such breakdown is provided by the doubly symmetric two-player game with strategy space $X = [0,1]$ and the strictly concave payoff function $g(x, y) = -2|x - y| - (x - 1/2)^2 - (y - 1/2)^2$ (Neyman 1997). Here, all strategies are equilibrium strategies, hence best response to themselves, which implies that none of them is a CSS. But strategy $1/2$ is globally stable, as can be inferred from Theorem 3 and the fact that $(1/2,1/2)$ is the unique maximum point of the payoff function g, which, because of its symmetry, is also the game's potential. All other strategies are not stable. This fact is in agreement with Neyman's assertion that the equilibrium points (x, x) with $x \neq 1/2$ lack some form of stability. Although that assertion apparently refers to dynamic stability of one kind or the other, a concrete kind of static stability rendering it true is that of Definition 4.

The common differential condition for stability and continuous stability is not shared by a third kind of static stability, local superiority, which in the present context is also known by a different name. In a symmetric two-player or population game with a unidimensional strategy space, an equilibrium strategy that is locally superior is said to be a *neighborhood invader strategy* (*NIS*; Apaloo 1997). A sufficient and "almost" necessary condition for an interior equilibrium strategy y to be an NIS is given by the differential form of the condition in Definition 9. That condition differs from the one in Theorem 8 in that the first term in (47), $g_{11}(y, y)$, is multiplied by $1/2$. Since, as indicated above, $g_{11}(y, y) \leq 0$, this difference makes the condition generally more demanding. Thus, unlike for symmetric $n \times n$ games, where stability and local superiority are equivalent (see Section 4.3), here stability is essentially a weaker requirement.

Example 6 *Stability does not imply local superiority*. In the symmetric two-player game or population game $g(x, y) = -2x^2 + 3xy$, with a strategy space that is a finite interval whose interior includes 0, the latter is a globally stable strict equilibrium strategy but is not an NIS. This is because, for $y = 0$ and all $x \neq 0$, inequality (9) holds but $g(y, x) < g(x, x)$. The conclusion is born out by the differential conditions for stability and local superiority, as $g_{11} + g_{12} < 0 < 1/2\, g_{11} + g_{12}$.

A similar relation between stability and local superiority holds for the mixed extensions of symmetric two-player and population games with a unidimensional strategy space. A mixed strategy is any (Borel) probability measure on the strategy space X. If the payoff function g is bounded and continuous, then the game has a well-defined mixed extension where the payoff $g(\mu, \nu)$ for a player using a strategy μ against a strategy (or population strategy) ν is given by

[18] This geometric condition for static stability is weaker than the corresponding one for dynamic stability, which requires the *absolute value* of slope to be less than 1 (Fudenberg and Tirole 1995). See also Section 6.

$$g(\mu,\nu)=\int_X\int_X g(x,y)\,d\mu(x)\,d\nu(y).$$

With any suitable topology on the space of mixed strategies, the mixed extension is itself a symmetric two-player game or population game, respectively, with bilinear payoff function. As shown in Section 4.2 (Proposition 10 and the preceding paragraph), for an equilibrium strategy in that game local superiority implies stability. However, the reverse implication does not hold: even a globally stable strict equilibrium strategy is not necessarily locally superior. In particular, this is so for local superiority with respect to the topology of weak convergence of measures, a concept called *evolutionary robustness* (Oechssler and Riedel 2002; van Veelen and Spreij 2009). For example, in the mixed extension of the game in Example 6 (which is similar to Example 4 in Oechssler and Riedel 2002; see also their 2001 paper), the degenerate measure δ_0 is a strict equilibrium strategy which is not evolutionary robust, because $g(\delta_x,\delta_x)>g(\delta_0,\delta_x)$ for all $x\neq 0$. However, δ_0 is globally stable, because $g(\mu,\mu)-g(\delta_0,\mu)+g(\mu,\delta_0)-g(\delta_0,\delta_0)=-\mathrm{E}^2-4\,\mathrm{Var}$ (where the two symbols refer to the mean and variance of μ) and the last expression is negative for all $\mu\neq\delta_0$.

5 Games with differentiable payoffs

This section concerns stability in a class of asymmetric games which includes, but is much larger than, the class of mixed extensions of finite games considered in the last part of Section 4.2. Here, a strategy space is not necessarily the unit simplex but can be any subset of a Euclidean space, and a payoff function is not necessarily multilinear. Multilinearity is replaced, where needed, by an (explicit) assumption that the payoff function has continuous second-order partial derivatives at the point or points under consideration.[19]

With strategies written as column vectors, a strategy profile $x=(x_1,x_2,\dots,x_N)$ in an asymmetric N-player game h where the strategy space of each player i is a set in a Euclidean space $\mathbb{R}^{n_i}$ is represented by a column vector of dimension $n=\sum_i n_i$. The profile will be said to be *interior* if each of the strategies x_i lies in the *relative interior* of the corresponding player's strategy space X_i (which coincides with the interior of X_i if the latter has affine dimension n_i, in other words, if it is of full affine dimension in $\mathbb{R}^{n_i}$). The gradient with respect to the components of player i's strategy is denoted ∇_i and is written as an n_i-dimensional row vector (of first-order differential operators). Correspondingly, for each i and j, $\nabla_i^{\mathrm{T}}\nabla_j$ is an $n_i\times n_j$ matrix (of second-order differential operators). In particular, $\nabla_i^{\mathrm{T}}\nabla_i h_i$ is the Hessian matrix of player i's payoff function with respect to the player's own strategy. These Hessian matrices are the diagonal blocks in the $n\times n$ block matrix

$$H=\begin{pmatrix}\nabla_1^{\mathrm{T}}\nabla_1 h_1 & \cdots & \nabla_1^{\mathrm{T}}\nabla_N h_1\\ \vdots & \ddots & \vdots\\ \nabla_N^{\mathrm{T}}\nabla_1 h_N & \cdots & \nabla_N^{\mathrm{T}}\nabla_N h_N\end{pmatrix}. \tag{50}$$

The value that the matrix H attains when its entries are evaluated at a strategy profile x is denoted $H(x)$. The next result, which is an extension of Proposition 7 in Milchtaich (2012), connects this value with the stability of the strategy profile.

[19] Technically, this assumption means that each point has an open neighborhood in the underlying Euclidean space where a twice continuously differentiable extension of the payoff function exists.

Theorem 9 In an asymmetric N-player game h where the strategy space of each player is a set in a Euclidean space, let y be an equilibrium at which the players' payoff functions have continuous second-order partial derivatives. If y is interior or the players' strategy spaces are convex, then a sufficient condition for y to be stable is that the matrix $H(y)$ is negative definite. If y is interior and the strategy spaces are of full affine dimension, then a necessary condition for y to be weakly stable is that $H(y)$ is negative semidefinite.[20]

Proof. It easily follows from Lemma 1 that, with 1_S denoting the indicator function of a set of players S,

$$P_y(x) = \sum_S \sum_i \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \frac{1}{N} \Big(1_S(i)\, h_i(\, y \mid x_S\,) - 1_{S^\mathrm{C}}(i)\, h_i(\, y \mid x_S\,)\Big).$$

For x tending to y, that is, $\epsilon_i \coloneqq x_i - y_i \to 0$ for all i, this expression can be written as

$$\frac{1}{N}\sum_S \sum_i \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \Big(1_S(i) - 1_{S^\mathrm{C}}(i)\Big)\left(h_i + \sum_{j\in S} \nabla_j h_i\, \epsilon_j + \frac{1}{2}\sum_{j\in S}\sum_{k\in S} \epsilon_k^\mathrm{T}\, \nabla_k^\mathrm{T} \nabla_j h_i\, \epsilon_j\right) + o(\|\epsilon\|^2), \quad (51)$$

where the payoff functions h_i and their partial derivatives are evaluated at the point y and $\|\epsilon\|$ is the (Euclidean) length of the vector $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_N) = x - y$. For each player i, the coefficient of h_i in (51) is

$$\frac{1}{N}\sum_S \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \Big(1_S(i) - 1_{S^\mathrm{C}}(i)\Big) = \frac{1}{N}\sum_{\substack{S\\ i\notin S}} \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \Big[\Big(1_S(i) - 1_{S^\mathrm{C}}(i)\Big) + \Big(1_{S\cup\{i\}}(i) - 1_{(S\cup\{i\})^\mathrm{C}}(i)\Big)\Big].$$

This coefficient is equal to zero, because the condition $i \notin S$ implies that the expression in square brackets is zero. For each i and j, the coefficient of $\nabla_j h_i\, \epsilon_j$ in (51) is

$$\frac{1}{N}\sum_{\substack{S\\ j\in S}} \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \Big(1_S(i) - 1_{S^\mathrm{C}}(i)\Big),$$

which by a similar argument is zero if $j \neq i$, and is equal to

$$\frac{1}{N}\sum_{\substack{S\\ i\in S}} \frac{1}{\binom{N-1}{|S|-1}} = \frac{1}{N}\sum_{l=1}^{N} \frac{\binom{N-1}{l-1}}{\binom{N-1}{l-1}} = 1 \quad (52)$$

if $j = i$. For each i, j and k, the coefficient of $(1/2)\epsilon_k^\mathrm{T}\, \nabla_k^\mathrm{T} \nabla_j h_i\, \epsilon_j$ in (51) is

$$\frac{1}{N}\sum_{\substack{S\\ j,k\in S}} \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \Big(1_S(i) - 1_{S^\mathrm{C}}(i)\Big),$$

which again is zero if j and k are both different from i. If $j = k = i$, then, by (52), the coefficient is equal to 1, and if $k = i$ but $j \neq i$ or vice versa, then it is equal to

[20] A square matrix A is said to be negative definite or semidefinite if the symmetric matrix $(1/2)(A + A^\mathrm{T})$ has the same property, equivalently, if the latter's eigenvalues are all negative or nonpositive, respectively.

$$\frac{1}{N}\sum_{\substack{S\\ i,j\in S}}\frac{1}{\binom{N-1}{|S|-1}}=\frac{1}{N}\sum_{l=2}^{N}\frac{\binom{N-2}{l-2}}{\binom{N-1}{l-1}}=\frac{1}{N}\sum_{l=2}^{N}\frac{l-1}{N-1}=\frac{1}{2}.$$

Therefore, (51) reduces to

$$\sum_i \nabla_i h_i\,\epsilon_i+\sum_i\left(\frac{1}{4}\sum_j \epsilon_i^{\mathrm{T}}\,\nabla_i^{\mathrm{T}}\nabla_j h_i\,\epsilon_j+\frac{1}{4}\sum_k \epsilon_k^{\mathrm{T}}\,\nabla_k^{\mathrm{T}}\nabla_i h_i\,\epsilon_i\right)+o(\|\epsilon\|^2) \tag{53}$$
$$=\sum_i \nabla_i h_i\,\epsilon_i+\frac{1}{2}\epsilon^{\mathrm{T}}H(y)\epsilon+o(\|\epsilon\|^2),$$

where the equality holds because, at y, the first-order partial derivatives of h_i commute and therefore $\epsilon_k^{\mathrm{T}}\,\nabla_k^{\mathrm{T}}\nabla_i h_i\,\epsilon_i=\epsilon_k^{\mathrm{T}}\,(\nabla_i^{\mathrm{T}}\nabla_k h_i)^{\mathrm{T}}\epsilon_i=\epsilon_i^{\mathrm{T}}\,\nabla_i^{\mathrm{T}}\nabla_k h_i\,\epsilon_k$.

If player i has a convex strategy space, then every convex combination of strategies x_i and y_i is also a strategy. The one-sided limit $\lim_{\lambda\to 0^+}(1/\lambda)(h_i(y \mid \lambda x_i+(1-\lambda)y_i)-h_i(y))$ exists and is equal to $\nabla_i h_i(y)(x_i-y_i)=\nabla_i h_i\,\epsilon_i$, and since y is an equilibrium, the limit is necessarily nonpositive. The same conclusions hold if the strategy space is not necessarily convex but y_i lies in its relative interior. Moreover, in this case, the one-sided limit with $\lambda\to 0^-$ also exists, and (again, because y is an equilibrium) is necessarily nonnegative. However, the last limit, too, is equal to $\nabla_i h_i\,\epsilon_i$, and so the latter must be zero. This proves that, if the players' strategy spaces are convex or the equilibrium y is interior, then the first term in (53) is nonpositive or zero, respectively.

If $H(y)$ is negative definite, then $\epsilon^{\mathrm{T}}H(y)\epsilon\le -|\lambda_0|\|\epsilon\|^2$, where λ_0 (<0) is the eigenvalue of $(1/2)(H(y)+H(y)^{\mathrm{T}})$ having the smallest absolute value. Therefore, if in addition either of the conditions in the last paragraph holds, then (53) is negative for $\epsilon\neq 0$ sufficiently close to 0, which means that $P_y(x)<0$ for $x\neq y$ sufficiently close to y. Thus, y is stable.

If $H(y)$ is not negative semidefinite, then $(1/2)(H(y)+H(y)^{\mathrm{T}})$ has an eigenvector v with eigenvalue $\lambda>0$, so that $v^{\mathrm{T}}H(y)v$ is positive and equal to $\lambda\|v\|^2$. If in addition y is interior (which, as shown above, implies that the first term in (53) is zero) and the strategy spaces are of full affine dimension, this means that there are strategy profiles x arbitrarily close to y for which $P_y(x)>0$. Thus, y is not weakly stable. ∎

It is interesting to note that negative definiteness of H is also connected with the *uniqueness* of the equilibrium (Rosen 1965). In particular, it follows from the next proposition that an equilibrium is necessarily unique if the players' strategy spaces are convex and H is negative definite everywhere. The same is true with 'equilibrium' replaced with 'interior equilibrium' and 'everywhere' replaced with 'at every interior strategy profile'.

Proposition 13 In an asymmetric N-player game h where the strategy space of each player is a set in a Euclidean space, let X' be a convex set of strategy profiles where the players' payoff functions are twice continuously differentiable. If $H(x)$ is negative definite for all $x\in X'$, then X' includes at most one equilibrium.

Proof. As shown in the proof of Theorem 9, for every $x,y\in X'$ such that y is an equilibrium the inequality $\nabla_i h_i(y)(x_i-y_i)\le 0$ holds for all i. If x, too, is an equilibrium, then similar inequalities hold with x and y interchanged, and so

$$\begin{aligned}
0 &\leq \sum_i \big(\nabla_i h_i(x) - \nabla_i h_i(y)\big)(x_i - y_i) \\
&= \sum_i \left(\int_0^1 \frac{d}{d\lambda} \nabla_i h_i(\lambda x + (1-\lambda) y)\, d\lambda\right)(x_i - y_i) \\
&= \int_0^1 \sum_{i,j} (x_j - y_j)^{\mathrm{T}}\, \nabla_j^{\mathrm{T}} \nabla_i h_i(\lambda x + (1-\lambda) y)\,(x_i - y_i)\, d\lambda \\
&= \int_0^1 (x-y)^{\mathrm{T}}\, H(\lambda x + (1-\lambda) y)(x-y)\, d\lambda.
\end{aligned}$$

If H is negative definite at every point on the line segment connecting x and y, then the *non*negativity of the last integral implies that $x - y$ must be zero. ■

6 Comparison with dynamic stability

As explained in Section 4, static stability is based on incentives rather than motion. Dynamic stability, by contrast, also depends on explicit assumptions about the way the incentives to move translate into actual changes of strategies. For example, if the players' strategy spaces in an asymmetric N-player game as in Section 5 are unidimensional (that is, $n_i = 1$ for all i), the law of motion may take the form

$$\frac{dx_i}{dt} = d_i h_{i,i}(x_1, x_2, \dots, x_N), \qquad i = 1,2,\dots,N, \tag{54}$$

with $d_i > 0$ for all i, where the symbol $h_{i,i}$ is shorthand for the partial derivative $\partial h_i / \partial x_i$ and t is the time variable. This system of differential equations expresses the assumption that the rate of change of each strategy x_i is proportional to the corresponding marginal payoff. With these dynamics, the condition for asymptotic stability of an interior equilibrium y where the players' payoff functions are twice continuously differentiable is that, at y, the (Jacobian) matrix

$$\begin{pmatrix} d_1 h_{1,11} & \cdots & d_1 h_{1,1N} \\ \vdots & \ddots & \vdots \\ d_N h_{N,N1} & \cdots & d_N h_{N,NN} \end{pmatrix}$$

(where $h_{i,jk} \coloneqq \partial^2 h_i / \partial x_j \partial x_k = \partial^2 h_i / \partial x_k \partial x_j$) is stable, that is, all its eigenvalues have negative real parts. The requirement that this condition holds for all positive adjustment speeds $d_1, d_2, \dots, d_N$ (Dixit 1986) is known as *D-stability* of the matrix obtained by omitting the d_i's, which is the matrix $H(y)$ (where H here is the special, unidimensional case of (50), in which each block is a single entry).

Every negative definite matrix is *D*-stable but not conversely, and so *D*-stability of $H(y)$ is a weaker condition than negative definiteness. In particular, a necessary and sufficient condition for *D*-stability in the two-player case[21] is

$$h_{1,11} < 0 \text{ and } h_{2,22} \leq 0 \text{ or vice versa, and } h_{1,11} h_{2,22} > h_{1,12} h_{2,21} \tag{55}$$

(Hofbauer and Sigmund 1998), while negative definiteness is equivalent to the stronger

[21] Unlike negative definiteness, for which a number of useful characterizations are known, necessary and sufficient conditions for *D*-stability of $n \times n$ matrices are known only for small n (Impram et al. 2005) and they are reasonably simple only for $n = 2$.

condition

$$h_{1,11}, h_{2,22} < 0 \text{ and } h_{1,11}h_{2,22} > \left(\frac{h_{1,12} + h_{2,21}}{2}\right)^2. \tag{56}$$

Moreover, unlike negative definiteness, *D*-stability of $H(y)$ is not a sufficient condition for (static) stability of an equilibrium y.

Example 1 (continued) Both in the game (4), where

$$H = \begin{pmatrix} -2 & 3 \\ -1 & -1 \end{pmatrix},$$

and in (5), where

$$H = \begin{pmatrix} -2 & 3 \\ 0 & -1 \end{pmatrix},$$

the matrix H satisfies (55) and is thus *D*-stable. Therefore, in both games the equilibrium $(0,0)$ is asymptotically stable with respect to the dynamics (54). However, as shown above, in the first game the equilibrium is stable but in the second game it is not even weakly stable. Note that these facts also follow from Theorem 9, because in the first game H is negative definite, as it satisfies (56), and in the second game it is not even negative semidefinite, as one eigenvalue of $(1/2)(H + H^{\mathrm{T}})$ is positive.

While asymptotic stability with respect to the dynamics (54) is essentially a weaker condition than static stability, the same may not be true for other kinds of dynamic stability. In particular, static stability does not imply asymptotic stability with respect to another natural adjustment process, in which the two players alternate in myopically playing a best response to their opponent's strategy. As seen in Figure 2, starting from any other strategy profile, these dynamics quickly bring the players to the origin in the game (5) but take them increasingly farther away from it in (4). Thus, the equilibrium $(0,0)$ is dynamically stable in (5) but not in (4) – which is the opposite of the situation for static stability and is also different from that for the simultaneous and continuous adjustment process (54) (for which the equilibrium is asymptotically stable in both games).

These differences between the different kinds of stability can be understood by noting that, if both inequalities in the first part of (55) are strict, then the second part can be written as

$$\left(-\frac{h_{2,21}}{h_{2,22}}\right)\left(-\frac{h_{1,12}}{h_{1,11}}\right) < 1.$$

Thus, asymptotic stability of an interior equilibrium y with respect to the dynamics (54) essentially requires that, at that point, the product of the slope of player 2's reaction curve and the reciprocal of the slope of player 1's curve is less than 1.[22] (Both reaction, or best-response, curves lie in the plane where the horizontal and vertical axes correspond to the strategies of player 1 and player 2, respectively, as in Figure 2.) This condition is weaker than the condition for asymptotic stability of the equilibrium with respect to alternating best responses, which is that the *absolute value* of the product is less than 1 (Fudenberg and Tirole 1995). The latter, stronger condition, which means that player 1's reaction curve is steeper than that of player 2, is not implied by (55). The condition is also not implied by, and

[22] The (weak version of this strict) inequality is also the differential condition for Motro's (1994) notion of continuous stability of a strategy profile (CSS) in an asymmetric two-player game.

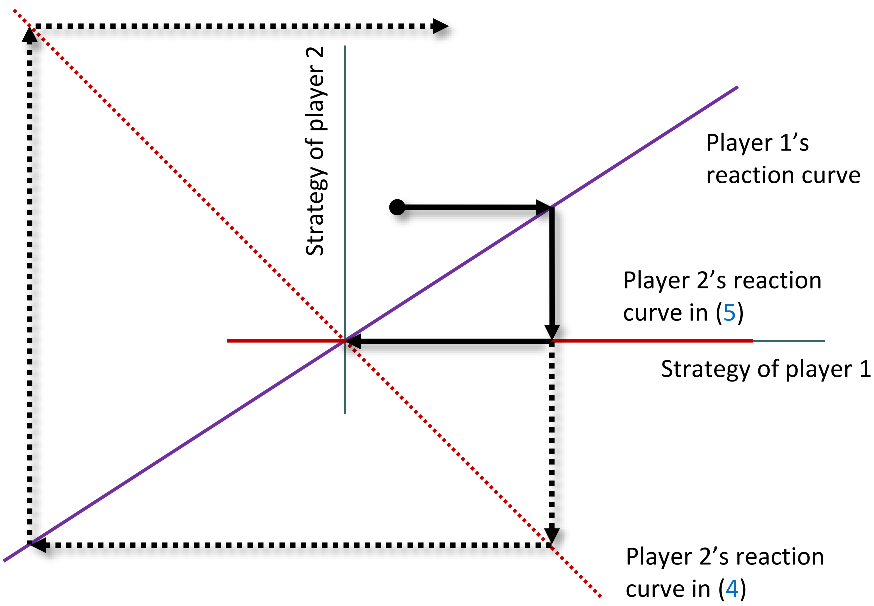

Figure 2. The players’ reaction curves in the two games in Example 1. Player 1’s reaction curve (upward sloping line) is the same in both games, but those of player 2 (horizontal and downward sloping lines) are different. The arrows show possible trajectories under the alternating-best-response dynamics, in which player 1 moves first, then player 2, then player 1 again, and so on. For the game given by (5) (solid arrows), the trajectory ends at the equilibrium point $(0, 0)$. For the game in (4) (dotted arrows), it spirals away.

it does not imply, negative definiteness of H, as demonstrated by the fact that it does not hold for the game in (4) but does hold for (5).

A general lesson that can be learned from the above analysis is that there is no single, general notion of dynamic stability with which static stability can be compared. Even for a specific, simple class of games, one kind of dynamic stability may be weaker than static stability while another may be incomparable with it.

An exception to the above general conclusion is provided by the essentially symmetric games (Section 2.2) with unidimensional strategy spaces. In these games, the matrix $H(y)$ is symmetric at any symmetric strategy profile y. A symmetric matrix is negative definite if and only if it is *D*-stable. This means that static stability of a symmetric strategy profile is essentially equivalent to asymptotic stability with respect to the dynamics (54). For example, in the two-player case, the essential symmetry condition (13) implies that, at any interior symmetric strategy profile,

$$h_{1,11} = h_{2,22} \text{ and } h_{1,12} = h_{2,21}.$$

With these equalities, both (55) and (56) are equivalent to the requirement that

$$h_{2,22} < 0 \text{ and } \left|\frac{h_{2,21}}{h_{2,22}}\right| < 1. \tag{57}$$

At any interior equilibrium, the second-order maximization condition $h_{i,ii} \leq 0$ holds automatically for $i = 1,2$, so that the first inequality in (57) only adds the requirement that the inequalities are strict. The second inequality, as indicated, means that the equilibrium is asymptotically stable with respect to alternating best responses. Thus, for an interior symmetric equilibrium, this kind of (dynamic) stability, asymptotic stability with respect to the continuous dynamics (54) and static stability are all essentially equivalent to one another and to the condition that, at the equilibrium point, the slope of player 2's reaction curve is less than 1 but greater than -1. On the other hand, as noted in footnote 18 in Section 4.4, the last pair of inequalities is stronger than the condition for static stability of an equilibrium strategy in a symmetric game, which consists of the first inequality only. This difference is another example of the more lenient nature of the (static) stability condition in symmetric games in comparison with the corresponding essentially symmetric ones (see Section 2.2).

7 Stability and altruism

In both symmetric and asymmetric games, static stability is closely linked with the comparative statics of altruism and spite, or more generally, of the degree of internalization of social welfare (Milchtaich 2012, 2021). This general connection may take different forms, as detailed below.

Altruism or spite is the willingness to bear a cost in order to benefit or harm, respectively, another individual. It may be quantified by the *altruism coefficient* r, which is the ratio between the marginal contribution of the other individual's material utility and the marginal contribution of the person's own material utility to the latter's perceived payoff. Positive, negative or zero r expresses altruism, spite or complete selfishness, respectively. In particular, if all players in an asymmetric game h are equally altruistic or spiteful towards one another, the expression that each player i seeks to maximize is not i's own, *personal payoff* h_i but the *modified payoff* $h_i + r\sum_{j\neq i} h_j$, which can also be written as

$$h_i^r \coloneqq (1-r)h_i + rf, \tag{58}$$

with $f = \sum_j h_j$ denoting the aggregate payoff. More generally, for any given *social payoff function* $f: X \longrightarrow \mathbb{R}$, which specifies a metric of social welfare that is determined by the players' strategy profile (either through the personal payoffs or directly), formula (58) for the modified payoff associates a *modified game* h^r with every $r \leq 1$. In this general framework, the altruism coefficient r expresses the players' common degree of internalization of the social payoff. Thus, higher r means higher rate of substitution between social and personal payoff. Comparative statics concern the effect of varying r on the actual level of f, which reflects the players' choice of actions. In particular, a basic question here is whether higher (common) degree of altruism necessarily entails higher social welfare.

A very similar setting and an identical question apply to symmetric games. The only difference is that, for a symmetric game g, the social payoff function f is assumed a symmetric function, which means that the players' actions or their personal payoffs affect the measure of social welfare under examination in a symmetric manner. With altruism coefficient $r \leq 1$, the modified game is the symmetric game that differs from g in that the payoff function is

$$g^r \coloneqq (1-r)g + rf.$$

In population games, f in the last formula is replaced by the *differential* $d\varphi$ of a specified social payoff function φ, which in this context is any univariate real-valued function on the cone of the strategy space with a differential that is continuous in the second argument (see Section 3.2). This difference from symmetric games reflects the assumed insignificance of single individuals in a large population, where the social payoff $\varphi(y)$ depends only on the population strategy y. Correspondingly, consideration for social welfare is interpreted as internalization of the *marginal* effect of one's action x on φ, which is given by $d\varphi(x, y)$ (Chen and Kempe 2008; Milchtaich 2012, 2021). Thus, an individual's concern is not with the effect of a unilateral adoption of x (which is null) but with the effect that adoption by a small but significant (and representative) proportion p of the population would have,[23] so that the modified payoff is given by

$$g^r(x, y) = (1 - r)g(x, y) + r\, d\varphi(x, y).$$

The question, again, is whether an increase in the weight r attached to these concerns actually results in a higher level of social payoff.

In general, the answer to the above questions is No (Milchtaich 2006, 2012). For example, even in a symmetric 3×3 game g, and with the aggregate payoff as the social payoff, the level of that payoff (and therefore also of both players' personal, material payoffs) at the unique, symmetric equilibrium in the modified game g^r may actually be lower when the players are mildly altruistic ($r = 0.25$, say) than when they are completely selfish ($r = 0$). However, as the next three theorems show, such a paradoxical effect of altruism on social welfare necessarily involves equilibria or equilibrium strategies that are not globally stable.

Theorem 10 (Milchtaich 2021, Theorem 7) For an asymmetric N-player game h, a social payoff function f, and altruism coefficients r and s with $r < s \leq 1$, if two distinct strategy profiles y^r and y^s are globally weakly stable in the modified games h^r and h^s, respectively, then

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

If moreover y^s is globally stable, then the inequality is strict. A strategy profile that is globally weakly stable or globally stable in h^1 is a maximum or strict maximum point, respectively, of f in the set of all strategy profiles.

Theorem 11 For a symmetric N-player game g, a social payoff function f, and altruism coefficients r and s with $r < s \leq 1$, if two distinct strategies y^r and y^s are globally weakly stable in the modified games g^r and g^s, respectively, then

$$f(y^r, y^r, \dots, y^r) \leq f(y^s, y^s, \dots, y^s).$$

If moreover y^s is globally stable, then the inequality is strict. A strategy that is globally weakly stable or globally stable in g^1 is a maximum or strict maximum point, respectively, of the function $x \mapsto f(x, x, \dots, x)$ in the set of all strategies.

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

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

[23] This description reflects the following identify, which follows from (29) (with φ instead of Φ):

$$d\varphi(x, y) - d\varphi(y, y) = \frac{d}{dp}\bigg|_{p=0^+} \varphi(px + (1 - p)y).$$

$$
\begin{aligned}
&= (1-r)(1-s)\sum_{j=1}^{N}\left(g(\underbrace{x,\dots,x}_{j\text{ times}},y,\dots,y) - g(\underbrace{y,\dots,y}_{j\text{ times}},x,\dots,x)\right) \\
&\qquad + (1-s)(1-r)\sum_{j=1}^{N}\left(g(\underbrace{y,\dots,y}_{j\text{ times}},x,\dots,x) - g(\underbrace{x,\dots,x}_{j\text{ times}},y,\dots,y)\right) \\
&\qquad + (1-r)s\left(f(x,x,\dots,x) - f(y,y,\dots,y)\right) \\
&\qquad + (1-s)r\left(f(y,y,\dots,y) - f(x,x,\dots,x)\right) \\
&= (s-r)\big(f(x,x,\dots,x) - f(y,y,\dots,y)\big).
\end{aligned}
$$

(The first equality uses the symmetry of the function f.) The identity implies that a sufficient condition for the difference $f(x,x,\dots,x) - f(y,y,\dots,y)$ to be nonpositive or negative is that the first term on the left-hand side is nonpositive or negative, respectively, and the second term is nonpositive. By Lemma 2, this condition holds with $x = y^r$ and $y = y^s$ if the latter strategy is globally weakly stable or globally stable, respectively, in g^s and the former is globally weakly stable in g^r. For $s = 1$, the condition also holds with any other $x \neq y^s$. ■

Theorem 12 (Milchtaich 2021, Theorem 8) For a population game g, a social payoff function φ, and altruism coefficients r and s with $r < s \leq 1$, if two distinct strategies y^r and y^s are globally weakly stable in the modified games g^r and g^s, respectively, then

$$\varphi(y^r) \leq \varphi(y^s).$$

If moreover y^s is globally stable, then the inequality is strict. A strategy that is globally weakly stable or globally stable in g^1 is a maximum or strict maximum point, respectively, of φ in the set of all strategies.

The reference in these theorems to global stability corresponds to the fact that they concern *global* comparative statics (Milchtaich 2012, Sections 6 and 7.1). That is, the comparison is between two strategies or strategy profiles in two modified games corresponding to different altruism coefficients r and s, without assuming that the coefficients are close or that the strategies or strategy profiles are close or can be connected in a continuous manner. However, as stability is fundamentally a local concept, it is relevant also to *local* comparative statics, which involve small, continuous changes to the altruism coefficient r and the corresponding strategies or strategy profiles, and may be thought of as tracing the players' evolving behavior as they respond to the changing r. The next three theorems are the local counterparts of those above. As they show, (local) stability is associated with "normal", positive local comparative statics, whereby a continuous increase in the altruism coefficient increases social welfare, and definite instability is associated with negative local comparative statics, in which the opposite relation holds.

Theorem 13 (Milchtaich 2012, Theorem 8) For an asymmetric game h and a social payoff function f such that both the payoff functions and f are Borel measurable,[24] and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leq 1$, suppose that there is a continuous and finitely-many-to-one[25] function assigning to each $r_0 \leq r \leq r_1$ a strategy profile y^r such that the function $\pi: [r_0, r_1] \longrightarrow \mathbb{R}$ defined by

[24] A sufficient condition for Borel measurability of a function is that it is continuous.

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

$$\pi(r) = f(y^r)$$

is absolutely continuous.[26] If the strategy profile y^r is stable, weakly stable or definitely unstable in the modified game h^r for every $r_0 < r < r_1$, then π is strictly increasing, nondecreasing or strictly decreasing, respectively.

Theorem 14 (Milchtaich 2012, Theorem 1) For a symmetric game g and a social payoff function f such that both the payoff function and f are Borel measurable, and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leq 1$, suppose that there is a continuous and finitely-many-to-one function assigning to each $r_0 \leq r \leq r_1$ a strategy y^r such that the function $\pi: [r_0, r_1] \longrightarrow \mathbb{R}$ defined by

$$\pi(r) = f(y^r, y^r, \dots, y^r)$$

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

Theorem 15 (Milchtaich 2012, Theorem 2) For a population game g and a social payoff function φ such that both the payoff function and $d\varphi$ are Borel measurable, and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leq 1$, suppose that there is a continuous and finitely-many-to-one function assigning to each $r_0 \leq r \leq r_1$ a strategy y^r such that the function $\pi: [r_0, r_1] \longrightarrow \mathbb{R}$ defined by

$$\pi(r) = \varphi(y^r)$$

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

The very general connection between static stability and comparative statics established by the above theorems is hardly intuitively obvious. Whereas stability concerns a comparison between different strategies or strategy profiles in a single, given game, comparative statics compare corresponding strategies or strategy profiles in different (modified) games.[27] Significantly, a similar connection does not generally hold for dynamic stability. Specifically, this is so in the class of symmetric $n \times n$ games, where a prominent notion of dynamic stability is asymptotic stability under the continuous-time replicator dynamics (Hofbauer and Sigmund 1998). As shown in Milchtaich (2012), even in a symmetric 3×3 game and with the aggregate payoff as the social payoff, continuously increasing the altruism coefficient may actually lower the players' identical (personal, material) payoffs when the equilibrium strategies involved are dynamically stable and raise them when the strategies are unstable. Thus, unlike static stability, dynamic stability does not preclude negative local comparative statics and instability does not preclude positive local comparative statics.

---

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

[27] This connection is somewhat reminiscent of that between the (local) degree of an equilibrium (or of a connected component of equilibria) and its index in several classes of games (Govindan and Wilson 1997; Demichelis and Germano 2000). The index of an equilibrium is connected with its asymptotic stability or instability with respect to a large class of natural dynamics, which determine how strategies *in the game* change over time. The degree, by contrast, expresses a topological property of the same equilibrium when viewed as a point on a manifold that includes the various equilibria of *different games* (Ritzberger 2002).

References

Apaloo, J., 1997. Revisiting strategic models of evolution: The concept of neighborhood invader strategies. *Theoretical Popul. Biol.* 52, 71–77.

Bomze, I.M., 1991. Cross entropy minimization in uninvadable states of complex populations. *J. Math. Biol.* 30, 73–87.

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

Bomze, I.M., Pötscher, B.M., 1993. "On the definition of an evolutionarily stable strategy in the playing the field model" by V. P. Crawford. *J. Theoretical Biol.* 161, 405.

Bomze, I.M., Weibull, J.W., 1995. Does neutral stability imply Lyapunov stability? *Games Econ. Behav.* 11, 173–192.

Broom, M., Cannings, C., Vickers, G.T., 1997. Multi-player matrix games. *Bull. Math. Biol.* 59, 931–952.

Bukowski, M., Miekisz, J., 2004. Evolutionary and asymptotic stability in symmetric multi-player games. *Int. J. Game Theory* 33, 41–54.

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.

Christiansen, F.B., 1991. On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* 138, 37–50.

Crawford, V.P., 1990. On the definition of an evolutionarily stable strategy in the "playing the field" model. *J. Theoretical Biol.* 143, 269–273.

Demichelis, S., Germano, F., 2000. On the indices of zeros of Nash fields. *J. Econ. Theory* 94, 192–217. (An extended version: Discussion paper, CORE, Louvain-la-Neuve, 2000.)

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

Dold, A., 1980. *Lectures on Algebraic Topology*, Second Edition. Springer-Verlag, Berlin.

Eshel, I., 2005. Asymmetric population games and the legacy of Maynard Smith: From evolution to game theory and back? *Theoretical Popul. Biol.* 68, 11–17.

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

Fudenberg, D., Tirole, J.,1995. *Game Theory.* MIT Press, Cambridge, MA.

Govindan, S., Wilson, R., 1997. Equivalence and invariance of the index and the degree of Nash equilibria. *Games Econ. Behav.* 21, 56–61.

Hammerstein, P., Selten, R., 1994. Game theory and evolutionary biology, in: Aumann, R.J., Hart, S. (Eds.), *Handbook of Game Theory with Economic Applications*, vol. 2. North-Holland, Amsterdam, pp. 929–993.

Harsanyi, J., Selten, R., 1988. *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge, MA.

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

Impram, S.T., Johnson, R., Pavani, R., 2005. The *D*-stability problem for 4 × 4 real matrices. *Arch. Math.* 41, 439–450.

Kohlberg, E., Mertens, J.-F., 1986. On the strategic stability of equilibria. *Econometrica* 54, 1003–1037.

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

Maynard Smith, J., Parker, G.R., 1976. The logic of asymmetric contests. *Anim. Behav.* 24, 159–175.

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

Milchtaich, I., 2012. Comparative statics of altruism and spite. *Games Econ. Behav.* 75, 809–831.

Milchtaich, I., 2021. Internalization of social cost in congestion games. *Econ. Theory* 71, 717–760.

Monderer, D., Shapley, L.S., 1996. Potential games. *Games Econ. Behav.* 14, 124–143.

Motro, U., 1994. Evolutionary and continuous stability in asymmetric games with continuous strategy sets: The parental investment conflict as an example. *Am. Nat.* 144, 229–241.

Neyman, A., 1997. Correlated equilibrium and potential games. *Int. J. Game Theory* , 223–227.

Oechssler, J., Riedel, F., 2001. Evolutionary dynamics on infinite strategy spaces. *Econ. Theory* 17, 141–162.

Oechssler, J., Riedel, F., 2002. On the dynamic foundation of evolutionary stability in continuous models. *J. Econ. Theory* 107, 223–252.

Riechert, S.E., 1998. Game theory and animal contests, in: Dugatkin, L.A., Reeve, H.K. (Eds.), *Game Theory and Animal Behavior*. Oxford University Press, Oxford, pp. 64–93.

Ritzberger, K., 2002. *Foundations of Non-Cooperative Game Theory*. Oxford University Press, Oxford.

Rosen, J.B., 1965. Existence and uniqueness of equilibrium points for concave N-person games. *Econometrica* 33, 520–534.

Samuelson, P.A., 1983. *Foundations of Economic Analysis*, Enlarged Edition. Harvard University Press, Cambridge, MA.

Sandholm, W.H., 2015. Population games and deterministic evolutionary dynamics, in: Young, H.P., Zamir, S. (Eds.), *Handbook of Game Theory*, vol. 4. North-Holland, Amsterdam, pp. 703–778.

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

Taylor, P.D., 1989. Evolutionary stability in one-parameter models under weak selection. *Theoretical Popul. Biol.* 36, 125–143.

van Damme, E., 1991. *Stability and Perfection of Nash Equilibria*, Second Edition. Springer-Verlag, Berlin.

van Veelen, M., Spreij, P., 2009. Evolution in games with a continuous action space. *Econ. Theory* 39, 355–376.

von Neumann, J., Morgenstern, O., 1953. *Theory of Games and Economic Behavior*, Third Edition. Princeton University Press, Princeton, NJ.

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*. Springer-Verlag, Berlin, pp. 29–97.