Static Stability in Games

By

Igal Milchtaich

April 16, 2020

RESEARCH INSTITUTE FOR ECONOMETRICS
AND ECONOMIC THEORY

DISCUSSION PAPER NO. 1-20

Research Institute for Econometrics
and Economic Theory
המכון לאקונומטריקה ותיאוריה כלכלית

DEPARTMENT OF ECONOMICS
BAR-ILAN UNIVERSITY
RAMAT GAN 5290002, ISRAEL

Static Stability in Games

Igal Milchtaich*

April 2020

Static stability in games differs from dynamic stability in only considering the players' incentives to change their strategies. It does not rely on any assumptions about the players' reactions to these incentives, so it is not necessarily linked with any particular dynamics. This paper introduces a general notion of static stability of strategies, which is applicable to any symmetric game and population game, and strategic stability of strategy profiles, which is applicable to any asymmetric game. It examines several important, large classes of games, with strategy spaces or payoff functions that have special structures (such as unidimensional strategy spaces or multilinear payoff functions), where this general notion takes a simple, concrete form. In particular, evolutionarily stable strategy (ESS) and continuously stable strategy (CSS) are shown to be essentially special cases of the general static stability concept for symmetric games. As an application, the paper identifies a connection between static stability and comparative statics of altruism. In general, increasing internalization of the aggregate payoff or some other kind of social payoff by all players may paradoxically result in a decrease of that payoff. But this is never so if static stability holds for the equilibria involved. *JEL Classification:* C72.

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

1 Introduction

A system is at an equilibrium state if there is no (net) force pushing it towards a different state. Stability differs from equilibrium 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. More precisely, this description concerns *static* stability, which unlike dynamic stability does not rely on a particular law of motion that specifies how forces translate into actual movement of the system. 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.

In game theory, where forces may be equated with incentives, a (Nash) equilibrium point is a strategy profile where there is no incentive for any player to change his strategy unilaterally. A *strict* equilibrium has the flavor of a stable point, as it is defined by the condition that any player who deviates to another strategy would gain from reverting to his original strategy.

* Department of Economics, Bar-Ilan University, Ramat Gan 5290002, Israel
igal.milchtaich@biu.ac.il http://faculty.biu.ac.il/~milchti

However, consideration of deviations by only one player may be too narrow a perspective. For example, in the finite (that is, with only pure strategies considered) two-player game

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

both (T, L) and (B, R) are strict equilibria but arguably only the latter should be considered stable.[1] The sense in which (T, L) fails to be stable is that if one player deviates from it, his incentive to return is weaker than the other player's incentive to also deviate. Or, from a somewhat different perspective, if the players start at the "opposite" strategy profile (B, R), their overall incentive to go to (T, L) is negative, as the loss for the first mover outweighs the second mover's gain.

In this example, the sum of the players' individual gains or losses when they change one-by-one from their strategies in (B, R) to those in (T, L) does not depend on the order of moves. But in general, it may depend. For example, this would be so if the diagonal payoff vectors were both changed to $(1,3)$. This change makes going from (B, R) to (T, L) relatively "easy" if the row player moves first, so the intermediate strategy profile is (T, R), but "hard" if the path goes through (B, L). A simple way of integrating the players' incentives to change strategies in different orders is averaging them. Consider then a strategy profile x that differs from a given strategy profile y in $k \geq 2$ coordinates, so that going from y to x requires k unilateral moves, which can be performed in $k!$ different orders. The players' overall incentive to move is quantified by the average over all $k!$ permutations (which specify the order in which the changes of strategies occur) of the sum of the k individual payoff increments for each permutation. The strategy profile y may be considered globally stable or weakly stable if this average is negative or nonpositive, respectively, for all $x \neq y$. (The modifier 'globally' refers to the fact that the requirement is not limited to x close to y; omitting it implies this limitation.) In the first case, y is necessarily a strict equilibrium, and in the second case, it is an equilibrium. However, as the example above shows, these necessary conditions are not sufficient for global stability or weak stability, which are thus stronger requirements than the corresponding equilibrium ones.

There is an equivalent definition of stability, which also suggests an alternative interpretation of that concept. Suppose that some players deviate from a strategy profile y by playing according to a specified alternative x. The average payoff of these players depends of their number and (unless the game is symmetric) their identities. Define the payoff of x players when playing against y players as (i) the sum over all numbers from 1 to the total number of players N (the sum could be replaced by the average, as the two differ only by the constant factor N) of (ii) the average over all sets of players S with this many members of (iii) the average payoff of these members if they play according to x and everyone outside S plays according to y. It is shown below that the incentive to move from

[1] This assessment diverges from that of Kohlberg and Mertens (1986). When discussing their notion of "strategic stability", these authors have nothing to say about the distinction between the two equilibria, which they maintain has to do with the pre-play bargaining game, and hence with cooperative theory, rather than with the game itself.

y to x can be expressed as the difference between the payoff of x players when playing against y players and the payoff of y players when playing against x players. Global (weak) stability of y is therefore equivalent to the condition that the first payoff is (respectively, weakly) lower than the second payoff for all $x \neq y$.

The notion of static stability outlined above does not assume or require sophisticated players. There is no foresight or an attempt to manipulate the other players, and no cooperation or coordination – including the incidental kind that might result from a restriction on the times at which players can move – which effectively rules out simultaneous deviations. The above definitions do however involve a strong, crucial assumption, namely that even in an asymmetric game, the players' utilities (more specifically, the changes of payoffs resulting from unilateral moves) are comparable. The comparability assumption reflects an interpretation of these payoff changes as a determinant of how readily or willingly the moves are made. This interpretation goes against an entrenched line of thought in economics and game theory, which views utilities merely as a convenient representation of preferences. For example, von Neumann-Morgenstern utilities are unique only up to arbitrary increasing affine transformations. Here, by contrast, only shifts by additive constants and scaling of all players' payoffs by a common factor would be inconsequential. Nevertheless, comparability of utilities is not a totally radical idea. When people talk about a "strong" or "weak" incentive to do something, their audience is likely to have an intuitive understanding of the meaning of these words. It may not be a huge leap to go from such qualitative understanding to a common quantitative measure of the incentive to act. In fact, under normal circumstances, monetary incentives may provide just such a common scale.

Another strong assumption underlying the definitions is that moves that decrease the mover's payoff are not impossible, although they are less likely to occur, or are less readily made, than payoff-increasing moves. The definition of (static) stability does not however rely on any concrete assumptions about when and where such moves, or any other ones, take place.

Any of the above assumptions may be challenged, and it is not this paper's goal to try to defend them on any abstract grounds. Rather, the subsequent sections collectively present a functional, practical justification for the ideas outlined above. As they show, these simple ideas lead to formal definitions that are widely applicable and take concrete, varied forms when applied to individual families of games. Significantly, some of these special forms coincide with established "native" kinds of static stability, such as evolutionarily stable strategy (in symmetric $n \times n$ games) and continuously stable strategy (in games with a unidimensional strategy space). In other families of games, the general definitions take a specific form that is different from, and is not equivalent to, any if the previous notions of static stability, which themselves may not be mutually equivalent. These findings suggest that the general notion of static stability presented in this paper may be viewed as a unifying theme, which lies behind a number of specific notions of stability whose usual justifications may be quite different, for example, in relying on dynamics. They also underline the desirability of deriving such notions from general, universal principles rather than from game-specific considerations. The proposed notion of stability is such a universal concept, as it does not depend on structures or properties that only certain kinds of games possess, such as multilinearity of the payoff functions or unidimensionality of the strategy spaces. Moreover, this essentially single notion of static stability applies, with the required adaptations, to asymmetric, symmetric and population games.

In symmetric and population games (Section 2), static stability refers to strategies. Stability of a strategy y in a symmetric N-player game may be interpreted as an expected loss of payoff for a player switching to any nearby alternative strategy x under the belief that, with equal probabilities, $0, 1, \dots$ or $N-1$ of the other players play x and the rest play y. (The assumption of equal probabilities is generalized in Section 2.4.) In a population game, which represents the limiting case of a large population of identical, individually insignificant players, the assumption of uniform distribution is analogously applied to the *proportion* of the population that has already switched from y to the alternative x.

In asymmetric N-player games (Section 3), static stability refers to strategy profiles. Stability of a strategy profile y takes into consideration all orders of individual changes of strategies that lead from y to an alternative strategy profile x, as outlined above. As indicated, it may alternatively be interpreted as the requirement that, for any strategy profile x close to y, the players playing according to x get on average a lower payoff than those adhering to y. The average involves all divisions of the players into complementary sets of x and y players.

The simplest kind of games where static stability is applicable are finite symmetric games (Section 4), in particular, two-player games with only two strategies, a and b. If both strategies are equilibrium strategies, then normally one of the two symmetric equilibria, (a, a) or (b, b), is risk dominant. It is not difficult to see that risk dominance is equivalent to (global) stability of the corresponding strategy, a or b.

The strategy spaces in finite games by definition include only pure strategies. When mixed strategies are allowed, the result is the mixed extensions of these games (Section 5). In particular, the mixed extensions of finite symmetric two-player games are the much-studied symmetric $n \times n$ games. As shown in Section 5.1, in these games, stability of a strategy y has a familiar meaning: it is equivalent to evolutionary stability, that is, to y being an ESS. However, the same is not true in the N-player case (Section 5.2). Here, stability of a strategy is a sufficient condition for ESS and is a necessary condition for ESS with uniform invasion barrier, but in both cases, the reverse implication does not generally hold for $N > 2$. Stability has a simple meaning also in asymmetric games (Section 5.3). However, in this case, it is an even stronger condition. A strategy profile in the mixed extension of a finite asymmetric N-player game is stable if and only if it is a strict equilibrium.

A particularly simple characterization of stability applies to potential games (Section 6). A strategy profile y in an asymmetric N-player potential game is stable if and only if it is a strict local maximum point of the potential, and it is globally stable if and only if it is a strict global maximum point. The latter is a stronger condition than y being a strict equilibrium, which only requires that all single-player deviations from y decrease the potential. Global stability in potential games can therefore be viewed as a refinement of the strict equilibrium condition, and global weak stability is similarly a refinement of Nash equilibrium. Similar results hold for strategies in symmetric and population games that are potential games. (The concept of potential for a population game is presented in Section 6.1.)

In games where strategies are real numbers or vectors (Section 7), stability can also be formulated in a differential form, that is, as a condition involving partial derivatives of the payoff function(s). In the special case of symmetric two-player and population games with a unidimensional strategy space (Section 7.1), this differential condition coincides with that for continuous stability, so an essentially necessary and sufficient condition for an equilibrium strategy to be stable is that it is a CSS. Geometrically, the 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 7.2), the differential condition for stability of an equilibrium y is negative definiteness of a particular matrix $H(y)$ whose entries are second-order partial derivatives of the payoff functions.

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 (Section 8). 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 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 that is 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, which specifies precisely how the players' behavior changes in response to these incentives.

The reliance of static stability only on incentives makes it particularly suitable for comparative statics analysis, in particular, study of the welfare effects of altruism and spite (Section 9). 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 ,2020). If these are stable, 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 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 Symmetric and population games

A symmetric N-player game g (where $N \geq 1$) is specified by the players' common strategy space X and a 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*, with the equilibrium payoff $g(y, y, \dots, y)$, if it is a best response to itself, that is, for every strategy $x \neq y$

$$g(y, y, \dots, y) \geq g(x, y, \dots, y). \tag{1}$$

It is a *strict equilibrium strategy* if all the inequalities are strict.

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[2] population of identical individuals who are "playing the field". This means that an individual's payoff $g(x, y)$ depends only on his own strategy x and on the suitable 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 each individual. Alternatively, it may describe the *distribution* of strategies in the population (Bomze and Pötscher 1989). In the latter case, X consists of *mixed strategies*, that is, probability measures on some underlying space of allowable actions or (pure[3]) 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.[4]

In population games, the equilibrium condition

$$g(y, y) = \max_{x \in X} g(x, y), \tag{2}$$

which is formally obtained from (1) (with $N = 2$), 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 g that is linear in the first argument, Eq. (2) may express the condition that (almost) everyone in the population is using a strategy that best responds to the population strategy y. In other words, the possibly *polymorphic* population is in an equilibrium state.

[2] 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.

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

[4] The special case of this interpretation in which the set of pure strategies is finite is often the one referred to by '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.

Examples of population games are nonatomic congestion games with a continuum of identical users, and public good games with an infinite population of identical players who have to decide whether to contribute their private good for the production of some public good (Milchtaich 2012, 2020). Another, important example is the following one (Bomze and Weibull 1995, Broom et al. 1997).

Example 1. *Random matching in a symmetric* N-player *game with a multilinear payoff function*. The N players are picked up independently and randomly from an infinite population of potential players, whose 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 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 $\bar{g}(x, y)$ is given by

$$\bar{g}(x, y) = g(x, y, \dots, y). \tag{3}$$

This equation defines a population game with strategy space X and payoff function $\bar{g}: X^2 \longrightarrow \mathbb{R}$. Clearly, 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.

2.1 Previous notions of static stability

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

Definition 1. 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[5] $\epsilon > 0$ the inequality

$$g(y, \epsilon x + (1 - \epsilon)y) > g(x, \epsilon x + (1 - \epsilon)y) \tag{4}$$

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. That is, for sufficiently small $\epsilon > 0$ (which cannot vary with x), (4) or a similar weak inequality, respectively, holds for all $x \neq y$.

Continuous stability (Eshel and Motro 1981, Eshel 2005) is another kind of static stability, which is applicable to games with a unidimensional strategy space.

Definition 2. 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

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

$$g((1-\epsilon)x+\epsilon y, x) > g(x, x) \tag{5}$$

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[6] 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.

Yet another notion of static stability in symmetric and population games is local superiority (or strong uninvadability; Bomze 1991).

Definition 3. 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 $x \neq y$,

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

Local superiority is applicable to any symmetric or population game where the strategy space X is a topological space, so the notion of neighborhood is well defined.[7] It does not rely on any other properties of the strategy space or the payoff function – unlike ESS and CSS, whose definitions above would not be meaningful without a linear structure. It is well known (see Section 5.1) that in the special case of symmetric $n \times n$ games local superiority is in fact equivalent to the ESS condition. However, for games with a unidimensional strategy space (Section 7.1) local superiority and CSS are not equivalent.

The next two subsections present a universal notion of static stability, that is, one that is applicable to all symmetric and population games. It (essentially) gives ESS and CSS as special cases when applied to specific, suitable classes of games.

2.2 Stability in symmetric games

Inequality (1) in the equilibrium condition and inequality (6) in the definition of local superiority both concern a player's lack of incentive to use a particular alternative x to his strategy y. In the equilibrium condition, all the other players are using y, and in local superiority, they use x. Stability, as defined below, differs from both concepts in considering not only the incentives to be first or last to move from y to x but also all the intermediate cases. In the simplest version, which is described by the followings definition (and is extended in Section 2.4), the same weight is attached to all cases. Put differently, stability of a strategy y means that, if the players move one-by-one from y to any nearby strategy x, their moves *on average* harm them.

Definition 4. A strategy y in a symmetric N-player game g is *stable*, *weakly stable* or *definitely unstable* if it has a neighborhood where, for every strategy $x \neq y$, the expression

[6] 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.

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

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

is negative, nonpositive or positive, respectively. If a similar condition holds for *all* strategies $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

Stability, as defined here, is a local concept. It refers to neighborhood systems of strategies, in other words, to a topology on the strategy space X. In principal, the topology needs to be explicitly specified, but in practice, it can often be understood from the context, as there is a unique natural topology.[8] In a game with a finite number of strategies, it may seem natural to consider the discrete topology, that is, to view strategies as isolated. However, as discussed in Section 4 below, 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 y with respect to the trivial topology automatically implies the same with respect to any topology, and coincides with the global version of the property. Note that there can be at most one globally stable strategy.

The inequality defining stability and instability can be put into an alternative form, which also suggests a somewhat different interpretation of these concepts. For strategies x and y, define the *payoff of x players when playing against y players* as the quantity

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

Lemma 1. Expression (7) is equal to $\mathcal{G}(x,y)-\mathcal{G}(y,x)$.

The proof of Lemma 1 is immediate, and uses a change of the summation variable and the fact that g is invariant to permutations of its second through Nth arguments. The lemma shows that stability of y means that if players only choose either this strategy or some alternative nearby strategy x, those doing the latter tend to fare worse than those doing the former. For definite instability, the opposite is true.

In some classes of symmetric games, stability of a strategy automatically implies that it is an equilibrium strategy (see Section 5). In other games, the reverse implication holds. In particular, an equilibrium strategy is automatically globally stable in every symmetric game g 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 $x \neq y$,

$$g(x,x,z,\dots,w)-g(y,x,z,\dots,w)<g(x,y,z,\dots,w)-g(y,y,z,\dots,w).$$

[8] The latter applies when it is natural to view X 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 strategies are simply (real) numbers, a set of strategies is a neighborhood of a strategy y if and only if, for some $\varepsilon > 0$, every $x \in X$ with $|x-y| < \varepsilon$ is in the set.

The condition implies that the summand in (7) strictly decreases as j increases from 1 to N, so if y is an equilibrium strategy, which by (1) means that the summand is nonpositive for $j = 1$, then the whole sum is negative, which means that y is globally stable. In general, however, the equilibrium and the stability conditions 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 a 'local equilibrium' strategy.[9] 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 can be written as (2) and the condition that (7) is negative can be rearranged to read

$$\frac{1}{2}\big(g(x,x) - g(y,x) + g(x,y) - g(y,y)\big) < 0, \tag{8}$$

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

$$pg(x,x) + \ (1-p)g(x,y) < pg(y,x) + (1-p)g(y,y) \tag{9}$$

holds for all $0 < p \leq 1/2$. The 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.

Local superiority is similar to stability in being a local condition. Moreover, for equilibrium strategies in symmetric two-player games, local superiority implies stability, since inequality (8) can be obtained by averaging (the two-player versions of) inequalities (6) and (1). The same implication also holds for certain kinds of symmetric games with more than two players (see Theorem 2 in Section 5). The reverse implication does not generally hold even for equilibrium strategies in symmetric two-player games (see Section 7.1).

2.3 Stability in population games

Stability in population games is defined by a variant of Definition 4 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 with an integral.

Definition 5. A strategy y in a population game g is *stable*, *weakly stable* or *definitely unstable* if it has a neighborhood where, for every strategy $x \neq y$, the integral

$$\int_0^1 \big(g(x, px + (1-p)y) - g(y, px + (1-p)y)\big)\, dp \tag{10}$$

is negative, nonpositive or positive, respectively. If a similar condition holds for *all* strategies $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

---

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

Stability is automatically implied by local superiority if the payoff function g is linear in the first or second argument. First, it is not difficult to see that, if a strategy y is locally superior, then for every $x \neq y$ in some neighborhood of y the inequality

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

holds 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 (10), so the integral is negative. If g is linear in the second argument, then the condition that the integral (10) is negative reduces to (8), which as shown above implies that every locally superior equilibrium strategy y is automatically stable.

If a population game $\bar{g}$ is derived from a symmetric game g as in Example 1, then, depending on whether y is viewed as a strategy in g or $\bar{g}$, Definition 4 or 5, respectively, applies. However, the point of view turns out to be immaterial.

Proposition 1. A strategy y in a symmetric N-player game g with a strategy space that is a convex set in a linear topological space and a continuous and multilinear payoff function is stable, weakly stable or definitely unstable if and only if y has the same property in the population game $\bar{g}$ defined by Eq. (3).

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 $0 \leq p \leq 1$ and strategies x, y and $x_p = px + (1-p)y$,

$$\begin{aligned}\bar{g}(x, x_p) - \bar{g}(y, x_p) &= g(x, x_p, \dots, x_p) - g(y, x_p, \dots, x_p) \\ &= \sum_{j=1}^{N} B_{j-1,N-1}(p)\Big(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}})\Big),\end{aligned} \tag{11}$$

where

$$B_{j-1,N-1}(p) = \binom{N-1}{j-1} p^{j-1}(1-p)^{N-j}, \qquad j = 1,2, \dots, N \tag{12}$$

are the Bernstein polynomials. These polynomials satisfy the equalities

$$\int_0^1 B_{j-1,N-1}(p)\, dp = \frac{1}{N}, \qquad j = 1,2, \dots, N. \tag{13}$$

It therefore follows from (11) by integration that the expression obtained by replacing g with $\bar{g}$ in (10) is equal to expression (7). ∎

The difference between stability in the sense of Definition 5 and in the sense of ESS (Definition 1) 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 if another strategy x replaces y in a small subpopulation, of size ϵ. By contrast, in Definition 5, the subpopulation to which x applies need not be small, but x itself is assumed close to y. The significance of this difference is examined in Section 5.

In that and other sections below, Definitions 4 and 5 are applied, or restricted, to a number of specific classes of symmetric and population games and the results are compared with certain "native" notions of stability for these games. The rest of the present section is concerned with an extension of the above framework, which facilitates capturing some additional native notions of stability.

2.4 $\bar{p}$-stability

Stability as defined above in a sense occupies the exact midpoint between equilibrium and local superiority. It takes into consideration a player's incentive to be the first or last to switch to a particular alternative strategy, but attaches to these extreme cases the same weight it attaches to each of the intermediate ones. This uniform distribution of weights may be interpreted as expressing a particular belief of the player about the total number of players who will be using the alternative strategy x after he switches to it, with the rest using the original strategy y. Namely, the probabilities $p_1, p_2, \dots, p_N$ that this number is $1,2, \dots, N$ are all equal,

$$p_j = \frac{1}{N}, \qquad j = 1,2, \dots, N. \tag{14}$$

Thus, unlike local superiority, in which the gain from switching from y to x is computed under the belief that all the other players are using x, in stability the expected gain is with respect to the probabilities (14), which give rise to expression (7). A straightforward generalization of both concepts is to consider any beliefs.

Definition 6. For a probability vector $\bar{p} = (p_1, p_2, \dots, p_N)$, a strategy y in a symmetric N-player game g is *$\bar{p}$-stable*, *weakly $\bar{p}$-stable* or *definitely $\bar{p}$-unstable* if it has a neighborhood where, for every strategy $x \neq y$, the expression

$$\sum_{j=1}^{N} p_j \left(g(x, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) - g(y, \underbrace{x, \dots, x}_{j-1 \text{ times}}, \underbrace{y, \dots, y}_{N-j \text{ times}}) \right) \tag{15}$$

is negative, nonpositive or positive, respectively.

If each of the other players switches to x with probability p and stays with y with probability $1 - p$, then

$$p_j = \begin{cases} 1-p, & j = 1 \\ 0 & , 1 < j < N \\ p & , j = N \end{cases} \tag{16}$$

or

$$p_j = \binom{N-1}{j-1} p^{j-1}(1-p)^{N-j} = B_{j-1,N-1}(p), \qquad j = 1,2, \dots, N, \tag{17}$$

depending on whether the choices are, respectively, perfectly correlated (i.e., identical) or independent. A strategy y is *dependently-* or *independently-stable* if it is $\bar{p}$-stable with $\bar{p} = (p_1, p_2, \dots, p_N)$ given by (16) or (17), respectively, for *all* $0 < p < 1$.

The number of other players using strategy x and the number using y have a symmetric joint distribution if the two numbers are equally distributed (hence, have an equal expectation of

$(N-1)/2$), that is,

$$p_j = p_{N-j+1}, \qquad j = 1,2,\dots,N. \tag{18}$$

A strategy y is *symmetrically-stable* if it is $\bar{p}$-stable for all $\bar{p} = (p_1, p_2, \dots, p_N)$ satisfying (18).

For single-player games ($N = 1$), stability and $\bar{p}$-stability of a strategy y mean the same thing, namely, strict local optimality: switching to any nearby alternative strategy x reduces the payoff. For $N = 2$, stability does not generally imply $\bar{p}$-stability (or vice versa) but the implication does partially hold (specifically, holds whenever $0 < p_2 \leq 1/2$) if y is an equilibrium strategy (see (9)). A full appreciation of the differences between stability in the sense of Definition 4 or and the varieties based on $\bar{p}$-stability requires looking at multiplayer games. One important class of such games is examined in Section 5.2.

3 Asymmetric games

In an asymmetric N-player game h, each player i has a strategy space X_i, which is a topological space, and a payoff function $h_i: X \longrightarrow \mathbb{R}$, where $X = \prod_i X_i$ is the space of all strategy profiles (endowed with the product topology). In such a game, consider for any 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 $(1,2,\dots,N)$ 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 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

$$\frac{1}{N!}\sum_{\pi\in\Pi}\sum_{j=1}^{N}\Big(h_{\pi(j)}(y \mid x_{\{\pi(1),\pi(2),\dots,\pi(j)\}}) - h_{\pi(j)}(x \mid y_{\{\pi(j),\pi(j+1),\dots,\pi(N)\}})\Big), \tag{19}$$

where $y \mid x_S$ denotes the strategy profile where the players in a set S play according to the strategy profile x and those outside S play according to y (and similarly with x and y interchanged). Expression (19) quantifies the overall incentive to move from y to x. The incentive to move in the opposite direction, from x to y, is given by the negative of (19).

Definition 7. A strategy profile y in an asymmetric N-player game h is *stable*, *weakly stable* or *definitely unstable* if it has a neighborhood where, for every strategy profile $x \neq y$, expression (19) is negative, nonpositive or positive, respectively. If a similar condition holds for *all* strategy profiles $x \neq y$, then y is *globally* stable, weakly stable or definitely unstable, respectively.

The somewhat unwieldy expression (19) can be put into simpler forms, which may in addition suggest alternative interpretations of the stability condition. In particular, if $N = 2$, then the condition that (19) is negative can be written in a form rather similar to (8), the stability condition in the symmetric case:

$$\frac{1}{2}\Big(\big(h_1(x_1,x_2) - h_1(y_1,x_2) + h_1(x_1,y_2) - h_1(y_1,y_2)\big) \\ + \big(h_2(x_1,x_2) - h_2(x_1,y_2) + h_2(y_1,x_2) - h_2(y_1,y_2)\big)\Big) < 0.$$

Another alternative form of (19) is given by the next lemma. This form shows that, as in the symmetric case, stability roughly means that when players only play according to x or according to y, those doing the former fare worse on average. For strategy profiles x and y, define the *payoff of x players when playing against y players* as the quantity

$$\mathcal{H}(x,y)=\sum_{j=1}^{N}\left[\frac{1}{\binom{N}{j}}\sum_{\substack{S\\|S|=j}}\bar{h}_S(y\mid x_S)\right]=\sum_{S}\frac{1}{\binom{N}{|S|}}\bar{h}_S(y\mid x_S)=\sum_{S}\frac{1}{\binom{N}{|S|}}\bar{h}_{S^{\mathrm{C}}}(x\mid y_S),\qquad(20)$$

where, for a set of players $S\subseteq\{1,2,\dots,N\}$, $|S|$ is the number of players in 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 (20) is obtained by replacing the summation variable S with the complementary set S^{C} and using the identity $x\mid y_S=y\mid x_{S^{\mathrm{C}}}$.) Note that the expression in square brackets is the average of $\bar{h}_S(y\mid x_S)$ over all sets 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.

Lemma 2. Expression (19) is equal to $\mathcal{H}(x,y)-\mathcal{H}(y,x)$.

Proof. Each of the payoffs in (19) has 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 in the first case by the equation $S=\{\pi(1),\pi(2),\dots,\pi(j)\}$ and in the second case by $S=\{\pi(j),\pi(j+1),\dots,\pi(N)\}$. In both cases, 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, (19) is equal to

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

■

It follows immediately from Definition 7 (by considering strategy profiles that differ from y in only one coordinate) that every globally weakly stable strategy profile is an equilibrium and every globally stable strategy profile is a strict equilibrium. However, the reverse implications do not hold. In fact, even in a two-player game, a strict equilibrium is not necessarily even weakly stable.

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

$$h_1(x_1,x_2)=-x_1^2+3x_1x_2\;\text{ and }\;h_2(x_1,x_2)=-\frac{1}{2}x_2^2-x_1x_2.\qquad(21)$$

It is not difficult to see that the origin is the unique (Nash) equilibrium, and it is moreover a strict equilibrium. In moving 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 $-x_2^2/2-x_1^2+3x_1x_2$. The average of the two sums is $-x_1^2+x_1x_2-x_2^2/2=-(x_1-x_2/2)^2-x_2^2/4$. This expression is negative for all $(x_1, x_2) \neq (0,0)$, which proves that the equilibrium is globally stable. However, in the game obtained by dropping the second term in h_2, where the payoff functions are

$$h_1(x_1, x_2) = -x_1^2 + 3x_1x_2 \text{ and } h_2(x_1, x_2) = -\frac{1}{2}x_2^2, \tag{22}$$

the corresponding average is $-x_1^2 + 3x_1x_2/2 - x_2^2/2$. As this expression is positive for every $(x_1, x_2) \neq (0,0)$ that is a multiple of (2,3), the strict equilibrium (0,0) is not even weakly stable.

A strategy profile y that is stable but not globally so is not necessarily a strict equilibrium, or even an equilibrium. However, it is still a "local strict equilibrium" in the sense that for each player i there is a neighborhood of y_i where, for every $x_i \neq y_i$,

$$h_i(y) - h_i(y \mid x_i) > 0, \tag{23}$$

where $y \mid x_i$ denotes the strategy profile that differs from y only in that player i uses strategy x_i. This facet of stability may be interpreted as the requirement that, when the players move one by one from y to any nearby strategy profile x, the *first* move harms the mover. This requirement is weaker than stability, which considers all the steps from y to x rather than only the first step. By contrast, the requirement that the *last* mover loses, at least on average, turns out to be a stronger condition than stability. This requirement is formalized and analyzed by the next definition and proposition.

Definition 8. A strategy profile y in an asymmetric N-player game h is *locally superior* if it has a neighborhood where, for every strategy profile $x \neq y$,

$$\frac{1}{N}\sum_{i=1}^{N}\left(h_i(x) - h_i(x \mid y_i)\right) < 0. \tag{24}$$

Proposition 2. Every locally superior strategy profile is stable, but not conversely.

Proof. A locally superior strategy profile y has a rectangular neighborhood where inequality (24) holds for every $x \neq y$. 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. Multiplication by -1, division by $\binom{N-1}{|S|-1}$ and summation over all nonempty sets S give

$$\begin{aligned}
0 &> \sum_{S \neq \emptyset} \frac{1}{\binom{N-1}{|S|-1}} \frac{1}{N} \sum_{i \in S} \left(h_i(y \mid x_S) - h_i(y \mid x_{S\setminus\{i\}})\right) \\
&= \sum_{S \neq \emptyset} \frac{1}{\binom{N}{|S|}} \frac{1}{|S|} \sum_{i \in S} h_i(y \mid x_S) - \sum_i \sum_{\substack{S \\ i \in S}} \frac{1}{\binom{N-1}{|S\setminus\{i\}|}} \frac{1}{N} h_i(y \mid x_{S\setminus\{i\}}) \\
&= \sum_S \frac{1}{\binom{N}{|S|}} \bar{h}_S(y \mid x_S) - \sum_i \sum_{\substack{S \\ i \notin S}} \frac{1}{\binom{N-1}{|S|} N} h_i(y \mid x_S) = \mathcal{H}(x, y) - \sum_S \frac{1}{\binom{N}{|S|}} \bar{h}_{S^{\mathrm{C}}}(y \mid x_S) \\
&= \mathcal{H}(x, y) - \mathcal{H}(y, x),
\end{aligned}$$

which by Lemma 2 proves that y is stable.

Example 2 shows that even global stability is not a sufficient condition for local superiority. In the two-player game with payoff functions (21), the equilibrium $y = (0,0)$ is globally stable but it is not locally superior because the left-hand side of (24) is equal to the expression $-x_1^2/2 + x_1x_2 - x_2^2/4$, which is positive if $x_1 = x_2 \neq 0$. ■

3.1 Comparison with stability in symmetric games

Definitions 4 and 7 differ in that one of them concerns strategies (in a symmetric game) while the other concerns strategy profiles (in an asymmetric game). Nevertheless, the two definitions are conceptually very similar, and in a sense, the latter can be derived from the former. The link between the two definitions is provided by the (standard) concept 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 he will use when called to assume the role of any player i in h. In that role, his 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, $x^1 = (x_1^1, x_2^1, \dots, x_N^1)$, $x^2 = (x_1^2, x_2^2, \dots, x_N^2)$, …, $x^N = (x_1^N, x_2^N, \dots, x_N^N)$,

$$
\begin{aligned}
g(x^1, x^2, \dots, x^N) &= \frac{1}{N!}\sum_{\pi\in\Pi} h_{\pi(1)}\left(x_1^{\pi^{-1}(1)}, x_2^{\pi^{-1}(2)}, \dots, x_N^{\pi^{-1}(N)}\right) \\
&= \frac{1}{N!}\sum_{\rho\in\Pi} h_{\rho^{-1}(1)}\left(x_1^{\rho(1)}, x_2^{\rho(2)}, \dots, x_N^{\rho(N)}\right),
\end{aligned} \tag{25}
$$

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. (Equivalently, $\rho(i) = \pi^{-1}(i)$ is the player assigned to role i.) Note that superscripts in (25) index players' strategies in the symmetric game g while subscripts refer to roles in the asymmetric game h.

Proposition 3. A strategy profile y in an asymmetric N-player game h is stable, weakly stable, definitely unstable or locally superior if and only if y has the same property as a strategy in the game g obtained by symmetrizing h. 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.

Proof. To prove the assertions concerning stability and instability, it suffices to show that the sum in (7) is equal to (19). By (25), that sum can be written as

$$
\sum_{j=1}^{N}\frac{1}{N!}\sum_{\pi\in\Pi}\Big(h_{\pi(1)}\big(y \mid x_{\{\pi(1),\pi(2),\dots,\pi(j)\}}\big) - h_{\pi(1)}\big(x \mid y_{\{\pi(1),\pi(j+1),\pi(j+n),\dots,\pi(N)\}}\big)\Big). \tag{26}
$$

Since the inner sum in (26) 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 (26) into (19).

The assertion concerning local superiority follows from the fact that the difference between the right- and left-hand sides of (6) is equal to

$$\frac{1}{N!}\sum_{\pi\in\Pi}\Big(h_{\pi(1)}(x_1,\dots,x_{\pi(1)},\dots,x_N)-h_{\pi(1)}(x_1,\dots,y_{\pi(1)},\dots,x_N)\Big).$$

This expression can be simplified by partitioning the set of permutations Π into N equal-size parts, each with cardinality $(N-1)!$, according to the value i of $\pi(1)$. The simplification gives the expression on the left-hand side of (24).

A strategy profile y is an equilibrium strategy in g if and only if setting $x^2=x^3=\cdots=x^N=y$ in (25) gives an expression that is maximized by choosing $x^1=y$. Using the idea indicated above, that expression simplifies to

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

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

As indicated, a stable strategy profile in an asymmetric game is also a local strict equilibrium. This fact contrasts with the situation for symmetric games, where a stable strategy is not always a local equilibrium strategy (see footnote 9). The difference suggests that, in some sense, stability is a weaker requirement in symmetric games than in asymmetric games.

In some classes of symmetric games, a stable strategy *is* automatically an equilibrium strategy. For example, it is shown in Section 5.1 below that this is so for symmetric $n\times n$ games, where a strategy is stable if and only if it is an ESS. However, even in this case, the stability condition is in a sense weaker than the corresponding one for asymmetric games, as an ESS is not necessarily a pure strategy and therefore not necessarily a strict equilibrium strategy. This contrasts with the situation for asymmetric $m\times n$, or *bimatrix games*, as Theorem 3 in Section 5.3 shows that a strategy profile in a bimatrix game is stable if and only if it is a strict (hence, pure) equilibrium.

By the last fact and Proposition 3, the stable strategies in the symmetric game g obtained by symmetrizing a bimatrix game h are the strict equilibria in 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, such a game is similar in this respect to symmetric $n\times n$ games (although it is generally *not* an $n\times n$ game, for any n).

3.2 Essentially symmetric games

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

$$h_i(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(N)}) = h_{\pi(i)}(x_1, x_2, \dots, x_N), \qquad i = 1,2, \dots, N. \tag{27}$$

Thus, if the players' strategies are shuffled, such that each player i takes the strategy of some other 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 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 finding 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 particular situation as a symmetric game and describing it as an essentially symmetric one, 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 here 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 differences between a symmetric game and the corresponding essentially symmetric game are reflected by the differences between the corresponding notions of stability: stability of a strategy in the first case and stability of a strategy profile in the second case. The second notion is more general, in that it is applicable also to asymmetric strategy profiles, where not all players are using the same strategy. However, even in the case of a symmetric strategy profile, in which all players use the same strategy y, and even if y is an equilibrium strategy in the symmetric game, stability of y in the symmetric game and stability of the symmetric equilibrium $(y, y, \dots, y)$ in the essentially symmetric game are not the same thing. In fact, the second requirement is stronger.[10] The reason is that the second requirement takes into consideration a larger set of alternatives than the first one does. An alternative to a strategy y is another (nearby) strategy x, to which all the players switch. The alternatives to a symmetric strategy profile include (nearby) strategy profiles that are not

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

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.

As an example of this difference, consider any symmetric 2×2 game that has a completely mixed (i.e., not pure) ESS y, for example, the game with payoff matrix (for the row player)

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

As indicated, an ESS in a symmetric 2×2 game is a stable strategy. But in the corresponding essentially symmetric bimatrix game

$$(A, A^{\mathrm{T}}) = \begin{pmatrix} 0 & 2{,}1 \\ 1{,}2 & 0 \end{pmatrix},$$

the symmetric equilibrium (y, y) is not stable because it is not locally strict: any unilateral deviation leaves the deviator's payoff unchanged.

Proposition 4. If a symmetric strategy profile $\vec{y} = (y, y, \dots, y)$ in an essentially symmetric N-player game h is stable, then strategy y is stable in the corresponding symmetric game g ($= h_1$). However, the converse is false even if $N = 2$ and $\vec{y}$ is an equilibrium. The strategy profile $\vec{y}$ is an equilibrium in h if and only if y is an equilibrium strategy in g.

Proof. To prove the first assertion, consider another symmetric strategy profile $\vec{x} = (x, x, \dots, x)$, 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 (27),

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

It follows from this equality, (20) and the fact that $h_1 = g$ that

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

It therefore follows from Lemma 1 that strategy y is stable in g if and only if $\mathcal{H}(\vec{x}, \vec{y}) - \mathcal{H}(\vec{y}, \vec{x}) < 0$ for every $x \neq y$ in some neighborhood of y. By Lemma 2, a sufficient condition for this is that $\vec{y}$ is stable in h. The example preceding the proposition shows that this condition is not necessary even if $N = 2$ and $\vec{y}$ is an equilibrium.

The last assertion in the proposition follows from the fact that, 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. ∎

4 Finite games and 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 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 interesting such game is a finite symmetric two-player game with only two strategies, for example, the game with payoff matrix

$$\begin{matrix} & a & b \\ a & \multicolumn{2}{c}{\multirow{2}{*}{}} \end{matrix} \quad \begin{array}{c} \\ a \\ b \end{array}\begin{array}{c} \begin{matrix} a & b \end{matrix} \\ \begin{pmatrix} 3 & 4 \\ 1 & 5 \end{pmatrix} \end{array}.$$

In this example, both strategies are equilibrium strategies. Strategy a is globally stable and strategy b is globally definitely unstable, because (using the form (8) of the stability condition)

$$\frac{1}{2}(5-4+1-3) < 0 < \frac{1}{2}(3-1+4-5).$$

The two inequalities, which are clearly 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 5. 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.

Risk dominance of the equilibrium (y, y) is equivalent to global stability of the (pure) strategy y also when the latter is viewed as a strategy in the mixed extension g of the finite game, that is, when mixed strategies are allowed. This follows from the fact that global stability of the equilibrium strategy y in the finite game implies that inequality (9) holds for all $0 < p \leq 1/2$, where x is the other pure strategy. Since the payoff function in g is bilinear, a similar inequality holds with x replaced by any convex combination of x and y other than y itself, which proves that y is globally stable also in the mixed extension g. However, since in the latter game the strategy space X is in effect the unit interval, the natural topology on it is not the trivial topology but the usual one. Stability with respect to the latter is a weaker condition than global stability. In particular, as shown in the next section, *both* pure strategies in g are stable if (as in the above example) they are strict equilibrium strategies.

5 Mixed extensions of finite games and evolutionary stability

N-player games where a strategy space is the unit simplex in a Euclidean space and a payoff function is multilinear are the multiplayer generalization of symmetric $n \times n$ games in the case of symmetric games and of bimatrix games in the case of asymmetric games. Viewed differently, these games are the mixed extensions of finite games. Correspondingly, strategies are interpreted as *mixed* strategies, that is, as probability vectors that assign a probability to each of the player's possible actions in the underlying finite game. The *support* of a strategy is the set of all actions that are assigned positive probability, and a strategy is *pure* or *completely mixed* if its support contains only one action j (in which case the strategy itself may also be denoted by j) or all actions, respectively.

Each of the following subsections considers a particular class of games as above, examines the meaning of stability in this class, and compares this concept with other notions of static stability that are meaningful for the class. The first subsection concerns the important class of symmetric $n \times n$ games, which are the mixed extensions of finite symmetric two-player games, the second subsection examines the general, N-player case, and the last subsection looks at asymmetric games.

5.1 Symmetric $n \times n$ games

In a symmetric $n \times n$ game g, the payoff function may be expressed by a square (payoff) matrix A of these dimensions. Thus, with strategies written as column vectors,

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

The game may be viewed either as a symmetric two-player game or as a population game. In the former case, Definition 4 applies, and in the latter, Definition 5 applies. However, it follows from Proposition 1 that the two definitions of stability in fact 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 a symmetric $n \times n$ game, 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 6. 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,

$$g(y, x) > g(x, x). \tag{28}$$

(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 (28) also holds.

An NSS is characterized by similar equivalent conditions, in which the strict inequality (28) is replaced by a weak one.

A completely mixed equilibrium strategy y in a symmetric $n \times n$ game is said to be *definitely evolutionarily unstable* (Weissing 1991) if the reverse of inequality (28) holds for all $x \neq y$.

Theorem 1. 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 two inequalities in (iii) and (iv) in Proposition 6 together imply (8), and the same is true with the strict inequalities (28) and (8) 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 (8) holds for all nearby strategies $x \neq y$. Therefore, y has the property that, for *every* strategy $x \neq y$, for sufficiently small $\varepsilon > 0$

$$g(\varepsilon x + (1-\varepsilon)y, \varepsilon x + (1-\varepsilon)y) - g(y, \varepsilon x + (1-\varepsilon)y) \\ + g(\varepsilon x + (1-\varepsilon)y, y) - g(y, y) < 0. \tag{29}$$

It follows from the bilinearity of the payoff function that (29) is equivalent to

$$(2-\varepsilon)\big(g(y,y) - g(x,y)\big) + \varepsilon\big(g(y,x) - g(x,x)\big) > 0. \tag{30}$$

Therefore, the above property of y is equivalent to (iv) in Proposition 6, 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 (28), (29) and (30) by weak inequalities, and in the second case (in which the first term in (30) vanishes for all x), they need to be replaced by the reverse inequalities. ■

5.2 Mixed extensions of finite symmetric multiplayer games

Extending the results obtained in the two-player case ($n \times n$ games) to the multiplayer case requires a corresponding generalization of the definition 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 1, stability of a strategy y in such a game is equivalent to stability of y in the population game $\bar{g}$ defined by (3). Thus, a natural way to define *evolutionary* stability is to require it to have a similar property. As evolutionary stability for population games is given by Definition 1, this requirement yields the following natural extension of that definition.

Definition 9. 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

$$g(y, x_\epsilon, \dots, x_\epsilon) > g(x, x_\epsilon, \dots, x_\epsilon). \tag{31}$$

An ESS *with uniform invasion barrier* satisfies the stronger condition that, for sufficiently small $\epsilon > 0$, inequality (31) 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 all smaller ϵ.

An equivalent definition of ESS is given by a generalization of condition (iv) in Proposition 6 (Broom et al. 1997; see also the proof of Lemma 5 below).

Lemma 3. 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 6), 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 a uniform invasion barrier is that the ESS is locally superior, and this condition is in fact also necessary (Bomze and Weibull 1995, Theorem 3; Lemma 4 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 $\bar{p}$-stability (Section 2.4). In fact, two of the latter turn out to be equivalent to ESS with uniform invasion barrier.

Theorem 2. 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 hold:

symmetrically-stable $\Longrightarrow$ dependently-stable $\Longleftrightarrow$ independently-stable
$\Longleftrightarrow$ locally superior $\Longleftrightarrow$ ESS with uniform invasion barrier $\Longrightarrow$ stable $\Longrightarrow$ ESS.

Each of the three implications is actually an equivalence in the two-player case but not in general.

The proof of Theorem 2 uses the next two lemmas, which hold for every game g as in the theorem. The first lemma uses the following terminology. A strategy y is *conditionally locally superior* if it has a neighborhood where inequality (6) holds for every strategy $x \neq y$ that satisfies the reverse of inequality (1).

Lemma 4. For any $0 < p < 1$, the following properties of an equilibrium strategy y are equivalent, and imply that y is stable:

(i) local superiority,
(ii) conditional local superiority,
(iii) $\bar{p}$-stability with $\bar{p} = (p_1, p_2, \dots, p_N)$ given by (16),
(iv) $\bar{p}$-stability with $\bar{p} = (p_1, p_2, \dots, p_N)$ given by (17),
(v) ESS with uniform invasion barrier.

Proof. The implication (i) $\Longrightarrow$ (iii) is trivial: inequality (1) (from the equilibrium condition) and inequality (6) together give

$$(1-p)\big(g(x, y, \dots, y) - g(y, y, \dots, y)\big) + p\big(g(x, x, \dots, x) - g(y, x, \dots, x)\big) < 0.$$

Clearly, if the first term on the left-hand side is nonnegative, then the second term must be negative. This proves that (iii) $\Longrightarrow$ (ii).

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

$$g(y, x_k, \dots, x_k) - g(x_k, x_k, \dots, x_k) \le 0 \tag{32}$$

and

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

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

$$x_k = \alpha_k z_k + (1 - \alpha_k) w_k, \tag{33}$$

where $0 < \alpha_k \le 1$, 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.,

$$g(y, y, \dots, y) - g(w_k, y, \dots, y) = 0. \tag{34}$$

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

$$g(y, y, \dots, y) - g(z_k, y, \dots, y) > 2\delta. \tag{35}$$

By (32), (33), (34) and (35), for all k

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

As $k \to \infty$, the two expressions in parentheses tend to zero, since $x_k \to y$. Therefore, $\alpha_k \to 0$, which by (33) implies that $w_k \to y$. Since y is conditionally locally superior and (34) holds for all k, for almost all k (that is, all $k > K$, for some integer K)

$$g(y, w_k, \dots, w_k) - g(w_k, w_k, \dots, w_k) \ge 0.$$

Therefore, for almost all k, by (33) and (12)

$$\begin{aligned}
&\frac{1}{\alpha_k}\big(g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k)\big) \\
&\le \frac{1}{\alpha_k}\Big(\big(g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k)\big) + (1 - \alpha_k)^{N-1}\big(g(y, w_k, \dots, w_k) - g(w_k, w_k, \dots, w_k)\big)\Big) \\
&= \sum_{j=2}^{N} \frac{B_{j-1,N-1}(\alpha_k)}{\alpha_k}\left(g(w_k, \underbrace{z_k, \dots, z_k}_{j-1 \text{ times}}, \underbrace{w_k, \dots, w_k}_{N-j \text{ times}}) - g(y, \underbrace{z_k, \dots, z_k}_{j-1 \text{ times}}, \underbrace{w_k, \dots, w_k}_{N-j \text{ times}})\right).
\end{aligned}$$

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

$$g(w_k, x_k, \dots, x_k) - g(y, x_k, \dots, x_k) < \alpha_k \delta. \tag{36}$$

On the other hand, by (35) and since $x_k \to y$, for almost all k

$$\begin{aligned}
\alpha_k \Big(\big(g(y, y, \dots, y) - g(z_k, y, \dots, y)\big) + \big(g(z_k, y, \dots, y) - g(z_k, x_k, \dots, x_k)\big) \\
+ \big(g(w_k, x_k, \dots, x_k) - g(w_k, y, \dots, y)\big)\Big) > \alpha_k \delta.
\end{aligned}$$

By (33) and (34), the left-hand side is equal to $g(w_k, x_k, \dots, x_k) - g(x_k, x_k, \dots, x_k)$, which by

(32) is less than or equal to

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

This contradicts (36). The contradiction proves that (ii) ⟹ (i).

To prove that (i) ⟹ (iv), assume that y is locally superior, and thus has a *convex* neighborhood U where (6) holds for every strategy $x \neq y$. By the convexity of U and the linearity of g in the first argument, for every $x \in U \setminus \{y\}$

$$g(y, x_p, \dots, x_p) > g(x, x_p, \dots, x_p), \tag{37}$$

where $x_p = px + (1 - p)y$. By the second equality in (11), inequality (37) is equivalent to the requirement that (15) is negative, with $(p_1, p_2, \dots, p_N)$ given by (17). Thus, y has property (iv).

Clearly, the above arguments also apply with p replaced by any other number in $(0,1)$. Integration of (15) over this interval, with $p_1, p_2, \dots, p_N$ given by (17), therefore gives that, for every $x \in U \setminus \{y\}$, expression (15) is negative also with each p_j given by the corresponding integral in (13). The equalities (13) therefore prove that the locally superior strategy y is stable.

The proof of the reverse implication (iv) ⟹ (i) is rather similar. As shown above, y has property (iv) if and only if it has a neighborhood U such that (37) holds for all strategies $x \neq y$ in U, equivalently, inequality (6) 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 (6) holds for every $x \neq y$ if and only if it has such a neighborhood of the form X_ϵ, for some $0 < \epsilon < 1$. ■

Lemma 5. For a probability vector $\bar{p} = (p_1, p_2, \dots, p_N)$ with $p_N > 0$, every $\bar{p}$-stable strategy y is an ESS.

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

$$\sum_{k=1}^{N} p_k \left(g(x_\epsilon, \underbrace{x_\epsilon, \dots, x_\epsilon}_{k-1 \text{ times}}, \underbrace{y, \dots, y}_{N-k \text{ times}}) - g(y, \underbrace{x_\epsilon, \dots, x_\epsilon}_{k-1 \text{ times}}, \underbrace{y, \dots, y}_{N-k \text{ times}}) \right)$$

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

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

The sum on the right-hand side 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 term (the inner sum) is necessarily positive, as by assumption $p_N > 0$. This observation proves that if y is $\bar{p}$-stable, then the condition in Lemma 3 holds. Parenthetically, note that in the special case $p_N = 1$ the observation also proves Lemma 3 itself. ■

Proof of Theorem 2. By Lemma 5, a strategy that has any of the seven properties in the theorem is an ESS, hence (by Lemma 3) an equilibrium strategy. An immediate corollary of Lemma 4 is that, for an equilibrium strategy, the properties of dependent- and independent stability, local superiority, and ESS with uniform invasion barrier are all equivalent, and imply stability. The special case $p = 1/2$ of the same lemma (specifically, of the implication (iii) $\Longrightarrow$ (i)) shows that symmetric-stability implies local superiority.

With only two players ($N = 2$), there is no difference between stability and symmetric-stability, and thus the equivalence of all the properties in the theorem follows from the first part of the proof and Proposition 6. The counterexamples in Example 3 below (where $N = 4$) complete the proof. ■

Example 3. *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 2, 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 (15) (with $y = (1,0,0)$) is negative reads

$$-2p_2x_2^2 - 4p_3(x_1x_2^2 - x_2^2x_3 - x_2x_3^2) - 3p_4(2x_1^2x_2^2 - 4x_1x_2^2x_3 - 4x_1x_2x_3^2 - x_2^4 - 4x_2^2x_3^2 + x_3^4 - 4x_2^3x_3) < 0.$$

Stability corresponds to $\bar{p} = (p_1, p_2, p_3, p_4) = (1/4,1/4,1/4,1/4)$, 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 $x = (x_1, x_2, x_3)$ 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 (31), which in the present case simplifies to

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

For every (fixed) strategy $x = (x_1, x_2, x_3) \neq (1,0,0)$, this inequality holds for sufficiently small $\epsilon > 0$. Therefore, $(1,0,0)$ is an ESS.

In version 2 of the game, for $\bar{p} = (1/4,1/4,1/4,1/4)$ the condition that (15) is negative 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 $\bar{p} = (1/8,3/8,3/8,1/8)$ the condition that (15) is negative 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 $\bar{p} = (1/8\,,3/8,3/8,1/8)$ the condition that (15) is negative 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 4 (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 probability vectors $\bar{p}$ satisfying (18) for which (15) is nonnegative for some strategies x arbitrarily close to $(1,0,0)$. For examples, for $\bar{p} = (1/20,9/20,9/20,1/20)$, the condition that (15) is negative 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.

5.3 Mixed extensions of finite asymmetric games

The mixed extensions of finite asymmetric games are the asymmetric N-player games where each player i has a strategy space X_i that is the unit simplex in a Euclidean space $\mathbb{R}^{n_i}$ and a multilinear payoff function h_i. As the following theorem shows, in this class of games stability has a simple, strong meaning.

Theorem 3. For a strategy profile in the mixed extension h of a finite asymmetric N-player game the following conditions 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 (23) 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 $\epsilon > 0$. However, by the linearity of h_i in player i's own strategy, that inequality is actually equivalent to (23), which proves that y is a strict equilibrium.

(iii) ⟹ (ii). Suppose that y is a strict equilibrium, so that (23) 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 (23) 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,

$$h_i(x) - h_i(\, x \mid z_i\,) \geq \delta_i/2, \qquad z_i \in Z_i. \tag{38}$$

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 (38) imply that $(1 - \epsilon_i)(h_i(x) - h_i(\, x \mid y_i\,)) = \epsilon_i(h_i(\, x \mid z_i\,) - h_i(x)) < 0$. The conclusion proves that there is a neighborhood of y where (24) holds for every $x \neq y$.

(ii) ⟹ (i). A special case of Proposition 2 (in which the inverse implication does hold!). ∎

6 Potential games and potential maximization

An asymmetric N-player game h is a *potential game* (Monderer and Shapley 1996) if it has an (*exact*) *potential*: a real-valued function on the set of strategy profiles, $P\colon X \longrightarrow \mathbb{R}$, 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,

$$h_i(\, y \mid x_i\,) - h_i(y) = P(\, y \mid x_i\,) - P(y), \qquad x_i \in X_i, y \in X.$$

It follows immediately from the definition that the potential is unique up to an additive constant.

For potential games, the overall incentive to move from one strategy profile to another is given by the corresponding potential difference. This observation means that static stability and instability have very simple characterizations in terms of the extremum points of the potential.

Theorem 4. A strategy profile y in an asymmetric N-player 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.

Proof. The first part of the theorem follows from the fact that, by definition of P, expression (19) can be written as

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

The special case where the topology is the trivial one and the definition of P immediately give the second part of the theorem. ∎

For symmetric N-player games, the notion of potential has essentially the same meaning as for 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 $x, x', y, z, \dots, w$,

$$g(x, y, z, \dots, w) - g(x', y, z, \dots, w) = F(x, y, z, \dots, w) - F(x', y, z, \dots, w). \tag{39}$$

It follows immediately from the definition that a necessary condition for the existence of a potential 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 g is the mixed extension of a finite symmetric game, then it is a potential game if and only if the above condition holds for any $N+2$ *pure* strategies. In this case, the potential, like the payoff function, is multilinear (see Monderer and Shapley 1996, Lemma 2.10).

Example 4. *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 $y = (y_1, y_2)$) is a potential:

$$F(x, y) = \big(g(1,1) - g(2,1)\big)x_1 y_1 + \big(g(2,2) - g(1,2)\big)x_2 y_2. \tag{40}$$

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, indeed, a doubly symmetric one,[11] with the same strategy space as the game g. It follows immediately from (39) that the two games have the same equilibrium, stable, weakly stable and definitely unstable strategies. As for asymmetric potential games, stability and instability have a strikingly simple characterization, which follows immediately from the observation that the sum in (7) is equal to the difference $F(x, x, \dots, x) - F(y, y, \dots, y)$.

Theorem 5. 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 theorem is illustrated by the following result, which also makes use of Theorem 1 and Example 4.

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

Corollary 1. 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 $\Phi: X \longrightarrow \mathbb{R}$ defined by

$$\Phi(x) = \frac{1}{2}\big(g(1,1) - g(2,1)\big)x_1^2 + \frac{1}{2}\big(g(2,2) - g(1,2)\big)x_2^2. \tag{41}$$

The next proposition shows that symmetrization of asymmetric games (see Section 3.1) maps potential games to potential games. Indeed, the proof of the proposition shows that it also essentially maps potentials to potentials.

Proposition 7. An asymmetric N-player game h has a potential if and only if this is so for the symmetric game g obtained by symmetrizing h.

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 g (where X is the strategy space). This is because, by (25), for $x^1, x^2, \dots, x^N$ and y in X,

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

Conversely, if g has a potential F, then a potential for h is the function P defined by

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

This is because, for $x, y \in X$ that differ only in the strategy of a single player i,

$$\begin{aligned}
&P(x) - P(y) = F(x, x, \dots, x) - F(y, y, \dots, y) \\
&= \sum_{j=1}^{N}\left(F(x, \underbrace{y, \dots, y}_{j-1\text{ times}}, \underbrace{x, \dots, x}_{N-j\text{ times}}) - F(y, \underbrace{y, \dots, y}_{j-1\text{ times}}, \underbrace{x, \dots, x}_{N-j\text{ times}})\right) \\
&= \sum_{j=1}^{N}\left(g(x, \underbrace{y, \dots, y}_{j-1\text{ times}}, \underbrace{x, \dots, x}_{N-j\text{ times}}) - g(y, \underbrace{y, \dots, y}_{j-1\text{ times}}, \underbrace{x, \dots, x}_{N-j\text{ times}})\right) \\
&= \sum_{j=1}^{N}\frac{1}{N!}\sum_{\substack{\rho\in\Pi \\ \rho(i)=1}}\Big(h_{\rho^{-1}(1)}(x_1, \dots, x_i, \dots, x_N) - h_{\rho^{-1}(1)}(x_1, \dots, y_i, \dots, x_N)\Big) = h_i(x) - h_i(y),
\end{aligned}$$

where the second equality uses the symmetry of the function F and the fourth equality uses the fact that $x_k = y_k$ for $k \neq i$. ∎

6.1 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 10. 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:

$$\frac{d}{dp}\Phi(px + (1-p)y) = g(x, px + (1-p)y) - g(y, px + (1-p)y). \tag{42}$$

Example 5. *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 by (41) is a potential. Note that, unlike the function F defined in (40), Φ is a function of one variable only.

Theorem 1, Corollary 1 and Example 5 hint at the following general result.[12] As for symmetric games, stability and instability (here, in the sense of Definition 5) of a strategy y in a population game with a potential Φ is related to y being a local extremum point of the potential.

Theorem 6. A strategy y in a population game g 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. By (42), the integral (10) can be written as

$$\int_0^1 \frac{d}{dp}\Phi(px + (1-p)y)\, dp,$$

and is therefore equal to $\Phi(x) - \Phi(y)$. This proves the first part of the theorem. It also follows from (42), in the limit $p \to 0$, that for all x and y

$$\left.\frac{d}{dp}\right|_{p=0^+} \Phi(px + (1-p)y) = g(x, y) - g(y, y). \tag{43}$$

If y is a local maximum point of Φ, then the left-hand side of (43) is nonpositive, 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 (43) is nonpositive. If Φ is strictly concave, this conclusion implies that the left-hand side of (42) is negative for all $0 < p < 1$, which proves that y is a strict global maximum point of Φ. By the first part of the proof, the conclusion implies that y is globally stable. ∎

[12] Conversely, Theorem 6 below and Example 5 together provide an alternative proof for Corollary 1.

Since by definition a potential is a continuous function, an immediate corollary of Theorem 6 is the following result, which concerns the existence of strategies that are (at least) weakly stable. The result sheds light on the difference in this respect between symmetric 2×2 games and, for example, 3×3 games. The former, which as indicated are potential games, always have at least one NSS, whereas for the latter, it is well known that this is not so. (One counterexample is a variant of the rock–scissors–paper game where a draw yields a small positive payoff for both players; see Maynard Smith 1982, p. 20.)

Corollary 2. 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.

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 (or extended) for 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

$$d\Phi(\hat{x}, \hat{y}) = \frac{d}{dt}\bigg|_{t=0^+} \Phi(t\hat{x} + \hat{y}). \tag{44}$$

(Note that the direction is specified by the first argument $\hat{x}$.) The differential exists if the (right) derivative in (44) exists for all $\hat{x}, \hat{y} \in \hat{X}$.

Proposition 8. 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.

The proposition is an immediate corollary of the following result.

Lemma 6. 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

$$\frac{d}{dp}\Phi(px + (1-p)y) = d\Phi(x, px + (1-p)y) - d\Phi(y, px + (1-p)y), \tag{45}$$
$$x, y \in X, 0 < p < 1.$$

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 specified condition. Replacing $\hat{y}$ in (44) with $p\hat{x} + \hat{y}$ gives

$$d\Phi(\hat{x}, p\hat{x} + \hat{y}) = \frac{d}{dt}\bigg|_{t=p^+} \Phi(t\hat{x} + \hat{y}), \qquad \hat{x}, \hat{y} \in \hat{X}, p \geq 0. \tag{46}$$

By the above Fact and the continuity properties of Φ and $d\Phi$, for $0 < p < 1$ the right derivative in (46) is actually a two-sided derivative and it depends continuously on $\hat{y}$. Therefore, the right-hand side of the equality in (45) is equal to the expression

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

which by the chain rule is equal to the derivative on the left-hand side. Hence, (45) holds.

The fact that the right derivative in (46) 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 $\lambda, t > 0$

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

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. ■

One interpretation of Lemma 6 is that any suitable univariate real-valued function Φ on a suitable set X *defines*, by means of the differential, a potential game. The payoff function in that game (namely, the restriction of $d\Phi$ to X^2) is necessarily linear in the first argument.

7 Games with differentiable payoffs and continuous stability

This section concerns a (very large) superset of the symmetric and asymmetric games considered in Section 5. 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.[13] The first subsection examines an important special case pertaining to symmetric games, and the second subsection studies the general case for asymmetric games.

[13] 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.

7.1 Symmetric and population games with a unidimensional strategy space

In a symmetric two-player or population game where the strategy space X is an interval in the real line $\mathbb{R}$, stability or instability of an equilibrium strategy, in the sense of either Definition 4 or 5, has a simple, familiar meaning. As shown below, if the payoff function is twice continuously differentiable, the condition for an interior equilibrium strategy (that is, one lying in the interior of X) to be stable or definitely unstable is essentially that, at the (symmetric) equilibrium point, the graph of the best-response function, or reaction curve, intersects the forty-five degree line from above or from below, respectively. This geometric characterization of stability and its differential counterpart are also shared by continuous stability (Definition 2), which shows that these two notions of static stability are essentially equivalent.

Theorem 7. In a symmetric two-player game or population game g with a strategy space X that is a (finite or infinite) interval in the real line, let y be an interior equilibrium strategy such that at the equilibrium point (y, y) the payoff function has continuous second-order partial derivatives.[14] 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.

Proof. Using Taylor's theorem, it is not difficult to show that, for x tending to y, both the left-hand side of (8) and expression (10) can be expressed as

$$g_1(y, y)(x - y) + \frac{1}{2}\big(g_{11}(y, y) + g_{12}(y, y)\big)(x - y)^2 + o((x - y)^2). \tag{48}$$

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 2. It may be possible to determine whether this condition holds by looking at the sign of

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

For x tending to y, the right-hand side of (49) is given by an expression similar to (48) (where, as indicated, the first term is zero) except that it lacks the factor $1/2$. Therefore, if (47) or the reverse inequality holds, then 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, (5) 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. ∎

---

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

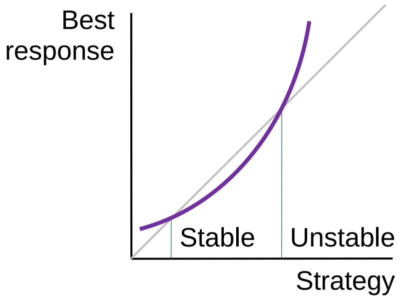

Figure 1. An equilibrium strategy is stable (and a CSS) or definitely unstable (and not a CSS) if, at the equilibrium point, the reaction curve (thick line) intersects the forty-five degree line (thin) from above or from below, respectively.

The connection between inequality (47) and the slope of the reaction curve can be established as follows (Eshel 2005). If y is an interior equilibrium strategy, then it follows from the equilibrium condition (2) 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 y is stable) or the reverse inequality holds (y definitely unstable) if and only if the slope of the function f at y is less or greater than 1, respectively.[15] 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). In the second case, the intersection is from below (and the fixed point index is -1).

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). Unlike for symmetric $n \times n$ games (Section 5), this condition is essentially stronger than stability. This can be seen most easily by working out the differential form of Definition 3 for an interior equilibrium strategy y. The sufficient and "almost" necessary condition obtained is similar to that in Theorem 7 except that the second term $g_{12}(y,y)$ in (47) is multiplied by 2. Since the first term $g_{11}(y,y)$ is necessarily nonpositive (as y is an equilibrium strategy), this makes the condition more demanding than that for stability (and CSS).

Example 6. *Stability does not imply local superiority*. In the symmetric two-player game or population game $g(x,y) = -2x^2 + 3xy$, where the strategy space X is some finite interval whose interior includes 0, the latter is a stable equilibrium strategy but is not an NIS, because $g_{11} + g_{12} < 0 < g_{11} + 2g_{12}$.

[15] 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).

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

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

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 2.2, every locally superior equilibrium strategy is automatically stable. 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). However, the reverse implication does not hold: a stable equilibrium strategy is not necessarily evolutionary robust. 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 an equilibrium strategy that is not evolutionary robust, because $g(\delta_x, \delta_x) > g(\delta_0, \delta_x)$ for all $x \neq 0$. However, it is even 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 $\mu \neq \delta_0$.

7.2 Asymmetric games with differentiable payoffs

Consider an asymmetric N-player game h where the strategy space of each player i is a set in a Euclidean space $\mathbb{R}^{n_i}$. Strategies are written as column vectors and, correspondingly, a strategy profile $x = (x_1, x_2, \ldots, x_N)$ is an n-dimensional column vector, where $n = \sum_i n_i$. It is an *interior* strategy profile if each x_i is an interior strategy in the sense that it lies in the *relative interior* of player i's strategy space (which coincides with the interior if the strategy space has affine dimension n_i, in other words, if it is of full affine dimension). 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.

Theorem 8. 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 the players' strategy spaces are convex, or if y

is an interior equilibrium, then a sufficient condition for it to be stable is that the matrix $H(y)$ is negative definite. If the strategy spaces are of full affine dimension and y is an interior equilibrium, then a necessary condition for it is to be weakly stable is that $H(y)$ is negative semidefinite.[16]

Proof. By Lemma 2, and with 1_S denoting the indicator function of a set of players S, expression (19) can be written as

$$\frac{1}{N}\sum_S \sum_i \frac{1}{\binom{N-1}{|S\setminus\{i\}|}}\Big(1_S(i)\, h_i(y \mid x_S) - 1_{S^c}(i)\, h_i(y \mid x_S)\Big). \tag{51}$$

For x tending to y, that is, $\epsilon_i = x_i - y_i \to 0$ for all i, (51) can be written as

$$\frac{1}{N}\sum_S \sum_i \frac{1}{\binom{N-1}{|S\setminus\{i\}|}}\Big(1_S(i) - 1_{S^c}(i)\Big)\left(h_i + \sum_{j\in S} \nabla_j h_i\, \epsilon_j + \frac{1}{2}\sum_{j\in S}\sum_{k\in S} \epsilon_k^{\mathrm{T}}\, \nabla_k^{\mathrm{T}} \nabla_j h_i\, \epsilon_j\right) + o(\|\epsilon\|^2), \tag{52}$$

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 (52) is

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

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 (52) is

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

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

$$\frac{1}{N}\sum_{\substack{S\\ i\in S}} \frac{1}{\binom{N-1}{|S|-1}} = \frac{1}{N}\sum_{l=1}^{N} \frac{\binom{N-1}{l-1}}{\binom{N-1}{l-1}} = 1 \tag{53}$$

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 (52) is

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

which again is zero if j and k are both different from i. If $j = k = i$, then, by (53), the coefficient is equal to 1, and if $k = i$ but $j \neq i$ or vice versa, then it is equal to

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

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

Therefore, (52) reduces to

$$\sum_i \nabla_i h_i\, \epsilon_i + \sum_i \left(\frac{1}{4} \sum_j \epsilon_i^{\mathrm{T}}\, \nabla_i^{\mathrm{T}} \nabla_j h_i\, \epsilon_j + \frac{1}{4} \sum_k \epsilon_k^{\mathrm{T}}\, \nabla_k^{\mathrm{T}} \nabla_i h_i\, \epsilon_i \right) + o(\|\epsilon\|^2) \tag{54}$$
$$= \sum_i \nabla_i h_i\, \epsilon_i + \frac{1}{2} \epsilon^{\mathrm{T}} H(y) \epsilon + o(\|\epsilon\|^2),$$

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 the strategy space of a player i is convex, 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) it is necessarily nonnegative. However, the last limit, too, is equal to $\nabla_i h_i\, \epsilon_i$, so the latter must be zero. 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 closest to 0 of $(1/2)(H(y) + H(y)^{\mathrm{T}})$, and therefore (54) is negative for $\epsilon \neq 0$ sufficiently close to 0, which proves that (19) is negative 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 $v^{\mathrm{T}} H(y) v$ is positive and equal to $\lambda \|v\|^2$. If in addition y is an interior equilibrium (which, as shown above, implies that the first term in (54) 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 (19) is positive. Thus, y is not weakly stable. ■

Interestingly, 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.

Proposition 9. 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 8, for $x, y \in X'$, if y is an equilibrium then 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, so

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

$$= \int_0^1 (x-y)^{\mathrm{T}} H(\lambda x + (1-\lambda)y)(x-y)\, d\lambda.$$

If H is negative definite at all points on the line segment connecting x and y, then it follows from the nonnegativity of the last integral that $x - y$ must be zero. ■

8 Comparison with dynamic stability

The notion of static stability, as defined in this paper, is based on incentives rather than motion. Dynamic stability, by contrast, is based 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 7.2 are unidimensional (i.e., $n_i = 1$ for all i), the law of motion may take the form

$$\frac{dx_i}{dt} = d_i h_{i,i}(x_1, x_2, \dots, x_N), \qquad i = 1,2,\dots,N, \tag{55}$$

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 $h_{i,jk} = \partial^2 h_i / \partial x_j \partial x_k = \partial^2 h_i / \partial x_k \partial x_j$) 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 setting $d_1 = d_2 = \cdots = d_N = 1$, which is $H(y)$. (H here is the "unidimensional" case of (50), in which each block is a single entry.)

Every negative definite matrix is *D*-stable but not conversely, so *D*-stability of $H(y)$ is a weaker condition than negative definiteness. For example, a necessary and sufficient condition for *D*-stability in the two-player case ($N = 2$) is

$$h_{1,11} < 0 \text{ and } h_{2,22} \leq 0\text{, or vice versa, and } h_{1,11}h_{2,22} > h_{1,12}h_{2,21} \tag{56}$$

(Hofbauer and Sigmund 1998),[17] whereas negative definiteness is equivalent to the stronger condition

$$h_{1,11}, h_{2,22} < 0 \text{ and } h_{1,11}h_{2,22} > \left(\frac{h_{1,12} + h_{2,21}}{2}\right)^2. \tag{57}$$

[17] 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 $n = 2$.

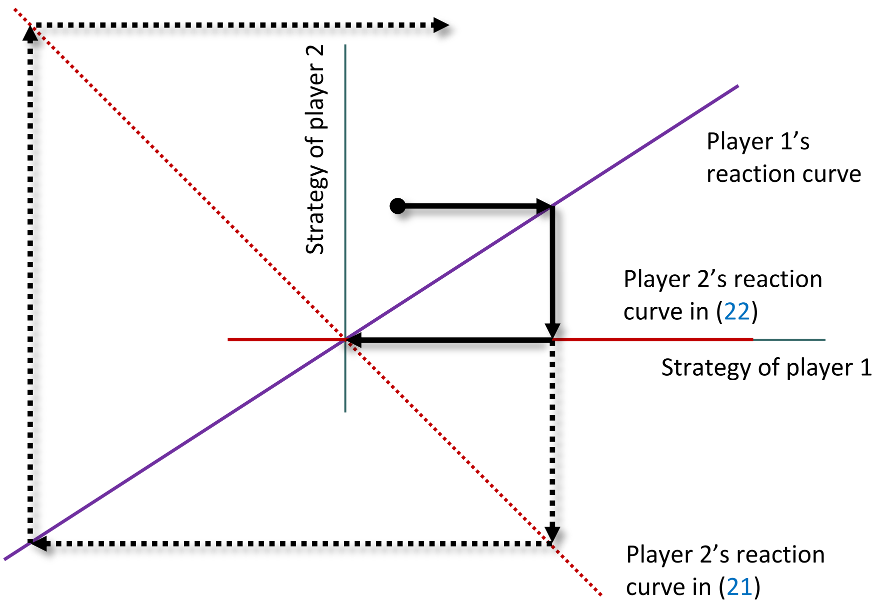

Figure 2. The players' reaction curves in the two games in Example 2. Player 1's reaction curve (upward sloping line) is the same in both games, but those of player 2 (horizontal and downward sloping lines) are different. The arrows show possible trajectories under the alternating-best-response dynamics, in which player 1 moves first, then player 2, then player 1 again, and so on. For the game given by (22) (solid arrows), the trajectory ends at the equilibrium point $(0, 0)$. For the game in (21) (dotted arrows), it spirals away.

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

Example 2 (continued). Both in the game (21), where

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

and in (22), where

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

the matrix H satisfies (56), and is thus *D*-stable. Therefore, the equilibrium $(0,0)$ is asymptotically stable with respect to the dynamics (55) in both games. However, as shown, in the first game the equilibrium is (statically) stable but in the second game it is not even weakly stable. Note that these facts also follow from Theorem 8, because in the first game H is negative definite, as it satisfies (57), 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 (55) 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 (22) but take them increasingly farther away from it in (21). Thus, the equilibrium $(0,0)$ is (dynamically) stable in (22) but not in (21), which is the opposite of the situation for static stability and is also different from that for the simultaneous and continuous adjustment process (55), 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 (56) 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 (55) 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.[18] (The two reaction, or best-response, curves lie in the space 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 (56). The condition is also not implied by, and it does not imply, negative definiteness of H, as demonstrated by the fact that it does not hold for the game in (21) but does hold for (22).

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 one may be incomparable with it.

An exception to the above general conclusion is provided by the essentially symmetric games (Section 3.2) with unidimensional strategy spaces. In such 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 (55). For example, in the two-player case ($N = 2$), the essential symmetry condition (27) 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 (56) and (57) are equivalent to

$$h_{2,22} < 0 \text{ and } \left|\frac{h_{2,21}}{h_{2,22}}\right| < 1. \tag{58}$$

At any interior equilibrium, the second-order maximization condition $h_{i,ii} \leq 0$ holds automatically for $i = 1,2$, so the first inequality in (58) 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

[18] 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.

symmetric equilibrium, this kind of (dynamic) stability, asymptotic stability with respect to the continuous dynamics (55) 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, 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 (see Section 7.1, and footnote 15). 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 3.1).

9 Static 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, 2020). 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 contributions of the other individual's material utility and the person's own material utility to the latter's perceived utility. 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*

$$h_i^r = (1-r)h_i + rf,$$

where $f = \sum_i h_i$ is the aggregate payoff. More generally, for altruism coefficient $r \leq 1$, the above formula associates a *modified game* h^r with any given *social payoff function* $f: X \longrightarrow \mathbb{R}$, which specifies a value for every strategy profile. In this general framework, the altruism coefficient r expresses the players' common degree of internalization of the social payoff, whose value depends on all players' actions. The basic question of comparative statics is whether a higher r necessarily entails a higher level of f.

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 chosen measure of social welfare in a symmetric manner. With altruism coefficient $r \leq 1$, the modified payoff, which is the payoff function in the modified game g^r, is given by

$$g^r = (1-r)g + rf.$$

In population games, f in the last formula is replaced by the *differential* $d\varphi$ of a 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 6.1). This difference from symmetric games reflects the 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, 2020). 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.[19] 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 (effectively, the other player's personal payoff) as the social payoff, the level of that payoff (and therefore also 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 9 (Milchtaich 2020, 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 10. 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:

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

[19] This description reflects the following identify, which follows from (45) (with φ substituted for Φ):

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

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

(The first equality uses the symmetry of the function f.) The identity implies that the difference $f(x,x,\dots,x) - f(y,y,\dots,y)$ is nonpositive or negative if the first term on the left-hand side is nonpositive or negative, respectively, and the second term is nonpositive. By Lemma 1, 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 11 (Milchtaich 2020, 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 strategies or strategy profiles or the coefficients are close or that it is possible to connect them 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 12 (Milchtaich 2012, Theorem 8). For an asymmetric game h and a social payoff function f such that the payoff functions and social payoff function are Borel measurable,[20] and altruism coefficients r_0 and r_1 with $r_0 < r_1 \leq 1$, suppose that there is a continuous and

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

finitely-many-to-one[21] 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

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

is absolutely continuous.[22] 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 13 (Milchtaich 2012, Theorem 1). For a symmetric game g and a social payoff function f such that the payoff function and social payoff function 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 14 (Milchtaich 2012, Theorem 2). For a population game g and a social payoff function φ such that 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.[23] 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

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

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

[23] 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).

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.

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., 2020. Internalization of social cost in congestion games. Forthcoming in *Econ. Theory*.

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.

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.