

**Evolutionary game theory for symmetric 2-player games:
fundamental results on dynamic and static stability properties**

Douglas G. Kelly

July 2009

Contents

0. Introduction
1. Notation and basic definitions
2. Best replies
3. Nash equilibrium
4. Evolutionary stability
5. Invasion barriers
6. Local superiority
7. Neutral stability
8. Replicator dynamics
9. Stationary points and Nash equilibria
10. Dynamic stability
11. Proving stability: Lyapunov's Theorem and linearization
12. Evolutionary stability and dynamic stability
13. Invariance of the dynamics under matrix operations
14. The 2x2 case: dynamics
15. The 2x2 case: static stability properties
16. Results and examples on evolutionary and neutral stability
17. Proving asymptotic stability using linearization
18. Bibliography

0. Introduction

This is a sketch, at the upper undergraduate level, of some of the basic results in evolutionary game theory for symmetric 2-player games. We introduce the *static* properties of Nash equilibrium, neutral stability, and evolutionary stability; then we introduce the replicator dynamics and the *dynamic* properties of stationarity, Lyapunov stability, and asymptotic stability.

Our primary goal is to consider the implications relating these six properties and to establish precisely which implications do and do not hold. The main results are summarized in Table 12.6; later sections provide examples and the surprising number of results needed to establish counterexamples to the implications that do not hold.

Familiarity with linear algebra is essential to reading these notes, and some exposure to differential equations will be helpful. Only a tiny amount of probability is needed;

probability mentioned only in (1.1) and in connection with Jensen's Inequality in Section 12.

It would be helpful also to have some knowledge of the most basic ideas of game theory. If the early going seems sketchy, the reader might consult the author's earlier survey, referred to in the Bibliography (Section 18 below). Also mentioned in the Bibliography are some good introductions to game theory: the book by Osborne, or, at a higher level, the one by Osborne and Rubinstein. Gintis's book is also interesting and illuminating.

These notes are based largely on material in Jorgen Weibull's book. We have also used material from Hofbauer and Sigmund's book. All the books mentioned here are listed in the Bibliography.

1. Notation and basic definitions

A **symmetric 2-player game with k actions** is determined by a $k \times k$ matrix $A = [a_{ij}]$; the entry a_{ij} represents the payoff to a player who takes action i when her opponent takes action j ($i, j = 1, \dots, k$).

A (**mixed**) **strategy** for a game with k actions is a vector $x = [x_1 \ x_2 \ \dots \ x_k]^T$ in the unit simplex in \mathbb{R}^k , which we denote by Δ_k or, when no confusion is possible, by Δ .

Formally, $\Delta = \{x \in \mathbb{R}^k : \sum_{i=1}^k x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i\}$.

All vectors named x , y , and so on will be column vectors.

We can interpret a strategy $x = [x_1 \ x_2 \ \dots \ x_k]^T$ in any of three ways:

1. An **x -player** plays action i with probability x_i ($i = 1, 2, \dots, k$).
2. An **x -population** is either
 - a. A large population of x -players, or
 - b. A **population with mix x** , which is a large group of players, each of whom always takes the same action, in which x_i is the proportion who always take action i .

The unit vectors $e^i = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0]^T$ of Δ represent the **pure strategies**. The pure strategy e^i consists of always playing action i . An individual who always plays action i is called an **i -player**.

If we denote by $u(x; y)$ the expected average payoff per play to a random member of an x -population playing against a random member of a y -population, then

$$u(x; y) = x^T A y, \text{ which for convenience we will denote } x \cdot A y. \quad (1.1)$$

This is a bilinear function on $\Delta \times \Delta$, and there is no other restriction: any bilinear function $u(x; y)$ on $\Delta \times \Delta$ (i.e., any $k \times k$ matrix A) uniquely determines a symmetric 2-player game with k actions.

The **support** of $x \in \Delta$ is the nonempty set $S(x) = \{i : x_i > 0\}$.

The (closed) face spanned by the vectors e^i for $i \in S(x)$ is the smallest face containing x , denoted $F(x)$.

2. Best replies

Strategy y is a **best reply** to strategy x if

$$u(y; x) \geq u(z; x) \text{ for all } z \in \Delta.$$

2.1 Proposition. Any linear combination in Δ of best replies to x is a best reply to x .

Proof. Suppose $y = \sum_{j=1}^m c_j y^j$ is in Δ , and each $y^j = [y_1^j \cdots y_k^j]$ is a best reply to x .

First we observe that if $y \in \Delta$, then $\sum_1^m c_j$ must equal 1 (although $0 \leq c_j \leq 1$ is not necessary). To see this, note that

$$y = \sum_{j=1}^m c_j \sum_{i=1}^k y_i^j e^i = \sum_{i=1}^k \left(\sum_{j=1}^m c_j y_i^j \right) e^i,$$

and since $y \in \Delta$, its coefficients must sum to 1:

$$1 = \sum_{i=1}^k \left(\sum_{j=1}^m c_j y_i^j \right) = \sum_{j=1}^m c_j \sum_{i=1}^k y_i^j = \sum_{j=1}^m c_j \cdot 1.$$

Now $u(y_j; x) \geq u(z; x)$ for all $z \in \Delta$, and therefore for any $z \in \Delta$ we have

$$u(y; x) = \sum_{j=1}^m c_j u(y^j; x) \geq \sum_{j=1}^m c_j u(z; x) = u(z; x). \quad \square$$

2.2 Proposition. If y is a best reply to x and $i \in S(y)$, then e^i is a best reply to x .

Proof. If any e^i is not a best reply to x , then

$$u(y; x) = \sum_{i \in S(y)} y_i u(e^i; x) < \sum_{i \in S(y)} y_i u(y; x) = u(y; x),$$

a contradiction. □

2.3 Proposition. If y is a best reply to x , then all the strategies in $F(y)$ (the smallest closed face containing y) are best replies to x .

Proof. This follows immediately from the two preceding propositions. □

2.4 Proposition. The set of all best replies to x is a face of Δ .

Proof. Every best reply is a convex combination of vertices that are best replies, by Proposition 2.2. And any convex combination of best replies is a best reply, by Proposition 2.1. So the set of best replies is the set of convex combinations of some set of vertices (namely, the union of the supports of all best replies). □

3. Nash equilibrium

A strategy x is a (*symmetric*) **Nash equilibrium** (NE) if x is a best reply to itself; that is, if

$$u(x; x) \geq u(y; x) \text{ for all } y \in \Delta. \tag{3.1}$$

A strategy x is a **strict Nash equilibrium** if the inequality in (3.1) holds strictly for all $y \neq x$.

We will frequently take advantage of linearity to write (3.1) as $u(x - y; x) \geq 0$, even though $x - y$ is not in Δ .

We state here without proof John Nash's famous theorem. (Nash, of course, did not use the term “Nash equilibrium”.)

3.2 Theorem (Nash, 1950). Any finite game has at least one Nash equilibrium.

The proof uses Kakutani's fixed-point theorem. If we let $\beta(x)$ denote the set of best replies to x , then by Proposition 2.4, $\beta(x)$ is a face of Δ , so it is nonempty, closed, and convex, for any x . One checks that the point-to-set map β is upper hemi-continuous, and it follows by Kakutani's theorem that there is at least one x such that $x \in \beta(x)$.

3.3 Proposition. x is a Nash equilibrium if and only if

$$u(e^i; x) \begin{cases} = u(x; x) & \text{if } i \in S(x) \\ \leq u(x; x) & \text{if } i \notin S(x). \end{cases} \quad (3.4)$$

Proof. Suppose x is a Nash equilibrium. Obviously then $u(e^i; x) \leq u(x; x)$ for all i . And if $i \in S(x)$, then by Proposition 2.2 e^i is a best reply to x , so $u(e^i; x) = u(x; x)$.

Conversely, if x satisfies (3.4), then for any $y = [y_1, \dots, y_n] \in \Delta$ we have

$$u(y; x) = \sum_{i=1}^k y_i u(e^i; x) \leq \sum_{i=1}^n y_i u(x; x) = u(x; x),$$

so x is a Nash equilibrium. □

3.5 Corollary. A strict Nash equilibrium is a pure strategy.

Proof. If x is a Nash equilibrium, then all e^i with $i \in S(x)$ are best replies also; but if x is a strict Nash equilibrium, then there are no best replies other than x . So $S(x)$ must contain only one index i , and x must equal e^i . □

4. Evolutionary stability

Strategy x is *evolutionarily stable* if for every $y \neq x$ there exists $\epsilon_y > 0$ such that

$$u(x; (1 - \epsilon)x + \epsilon y) > u(y; (1 - \epsilon)x + \epsilon y) \quad \text{for } 0 < \epsilon < \epsilon_y. \quad (4.1)$$

We refer to evolutionarily stable strategies as *ESS*.

If we replace the strict inequality sign $>$ in (4.1) by \geq , we get the definition of a *neutrally stable* strategy (an *NSS*).

Where we are headed in this and the next two sections. We are going to establish, among other things, that each of the following three conditions is equivalent to the evolutionary stability of a strategy x .

(Theorem 4.3) x is a Nash equilibrium (i.e., $u(x, x) \geq u(y, x)$ for all y) that satisfies in addition

$$\text{if } u(x; x) = u(y; x) \text{ for some } y \neq x, \text{ then } u(x; y) > u(y; y).$$

(Theorem 5.5) The strict inequality in (4.1) holds uniformly (x has a **uniform invasion barrier**); that is, there exists ϵ_x such that for every $y \neq x$

$$u(x; (1 - \epsilon)x + \epsilon y) > u(y; (1 - \epsilon)x + \epsilon y) \text{ for all } \epsilon < \epsilon_x. \quad (4.2)$$

(Theorem 6.1) x is **locally superior**; that is, there is a neighborhood U of x such that

$$u(x; y) > u(y; y) \text{ for all } y \neq x \text{ in } \Delta \cap U.$$

4.3 Theorem. Strategy x is evolutionarily [resp. neutrally] stable if and only if x satisfies

$$\begin{aligned} &u(y; x) \leq u(x; x) \text{ for all } y \in \Delta \quad (\text{i.e., } x \text{ is a NE}), \text{ and} \\ &\text{if } u(y; x) = u(x; x) \text{ for some } y \neq x, \text{ then } u(y, y) < [\text{resp. } \leq] u(x, y). \end{aligned} \quad (4.4)$$

Proof. The inequality in (4.1) can be rewritten

$$(1 - \epsilon)(u(x; x) - u(y; x)) + \epsilon(u(x; y) - u(y; y)) > 0,$$

and the result follows. \square

John Maynard Keynes gave 4.4 as the definition of evolutionary stability.

4.5 Corollary. An ESS is a NSS, and a NSS is a NE. \square

5. Invasion barriers

If x is an ESS and $y \neq x$, the **invasion barrier of x against y** , denoted $b_x(y)$, is the smaller of $\sup\{\epsilon_y : (4.1) \text{ holds}\}$ and 1. This is a positive number, and we have

$$u(x - y; (1 - \epsilon)x + \epsilon y) > 0 \text{ for } 0 < \epsilon < b_x(y). \quad (5.1)$$

It will be useful to define

$$\begin{aligned} f(\epsilon; x, y) &= u(x - y; (1 - \epsilon)x + \epsilon y) \\ &= u(x - y; x) + \epsilon(u(x - y; y - x)). \end{aligned}$$

This is a straight-line function of ϵ .

Notice that the definition of $f(\epsilon; x, y)$ makes sense for all $\epsilon \in \mathbb{R}$ and all x and $y \in \mathbb{R}^k$, although it has no meaningful interpretation as a difference of payoffs unless x , y , and $(1 - \epsilon)x + \epsilon y$ are all in Δ .

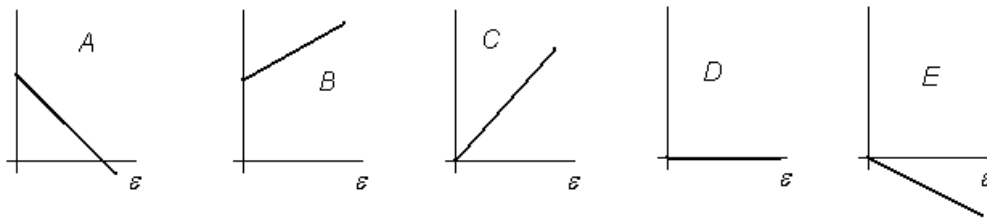
Using this notation, we can say that

x is a NE if and only if $f(0; x, y) \geq 0$ for all y ;

x is an ESS (by definition) if and only if for each $y \neq x$ there is ϵ_y such that $f(\epsilon; x, y) > 0$ for $0 < \epsilon < \epsilon_y$; and

x is an ESS (by Theorem 4.3) if and only if x is a Nash equilibrium and in addition, if $f(0; x, y) = 0$ for some $y \neq x$, then $f(1; x, y) > 0$.

We see that if x is a NE and $y \neq x$, then the graph of $f(\epsilon; x, y)$ as a function of ϵ looks like one of the following.



And an NE is an ESS if and only if $f(\epsilon; x, y)$ is of the form A, B, or C for every $y \neq x$. Obviously, then, $b_x(y) = 1$ in cases B and C; and in case A, $b_x(y)$ is either the zero of $f(\epsilon; x, y)$ or 1, whichever is smaller.

The main result of this section is Theorem 5.5 below: an ESS has a **uniform invasion barrier**. That is, there is an ϵ_x independent of y such that (4.2) holds. Equivalently, b_x , defined to be $\inf_{y \in \Delta} b_x(y)$, is positive.

It is easy to see that $b_x(y)$ is a continuous function of y . And if x is an ESS, then by definition $b_x(y)$ is positive for all $y \neq x$. We would like to argue that since Δ is compact, $b_x(y)$ attains its infimum at some $y \in \Delta$ and therefore the infimum must be positive. However, $b_x(y)$ is not defined on all of Δ , but only on $\Delta - \{x\}$, which is not compact. We have to proceed as follows.

5.2 Proposition. If x is an ESS, $y \neq x$, and z is any point on the line segment joining x and y , then $b_x(z) \geq b_x(y)$. (That is, the farther y is from x , the smaller the invasion barrier of x against y .)

Proof. The assertion is equivalent to

$$f(\epsilon; x, z) > 0 \text{ for } \epsilon < b_x(y). \quad (5.3)$$

But z is of the form $(1 - \alpha)x + \alpha y$ for some $\alpha \in (0, 1)$, so

$$x - z = \alpha(x - y),$$

and for any $\epsilon > 0$

$$\begin{aligned} (1 - \epsilon)x + \epsilon z &= (1 - \epsilon)x + \epsilon(1 - \alpha)x + \epsilon\alpha y \\ &= (1 - \epsilon\alpha)x + \epsilon\alpha y. \end{aligned}$$

Therefore

$$\begin{aligned} f(\epsilon; x, z) &= u(x - z, (1 - \epsilon)x + \epsilon z) \\ &= u(\alpha(x - y), (1 - \epsilon\alpha)x + \epsilon\alpha y) \\ &= \alpha f(\epsilon\alpha; x, y). \end{aligned}$$

If $\epsilon < b_x(y)$, then $\epsilon\alpha < b_x(y)$, so $f(\epsilon\alpha; x, z) > 0$. Also $\alpha > 0$, and it follows that (5.2) holds. \square

5.4 Corollary. If x is an ESS, then $\inf_{y \in \Delta} b_x(y) = \inf_{y \in F_x} b_x(y)$, where F_x denotes the union of all closed faces of Δ that do not contain x .

Proof. For any $y \neq x$, the line from x through y , continued, intersects the boundary of Δ in a point y^* of F_x . By Proposition 5.2, $b_x(y^*) \leq b_x(y)$. \square

5.5 Theorem. If x is an ESS, then $\inf_{y \in \Delta} b_x(y) > 0$. That is, there exists $\epsilon_x > 0$ such that

$$u(x - y, (1 - \epsilon)x + \epsilon y) > 0 \text{ for all } y \neq x \text{ and } 0 < \epsilon < \epsilon_x.$$

Proof. $b_x(y)$ is continuous on the compact set F_x , so there is $y_0 \in F_x$ such that $b_x(y_0) = \inf_{y \in F_x} b_x(y) = \inf_{y \in \Delta} b_x(y)$. But because x is an ESS, $b_x(y_0) > 0$. \square

If x is an ESS, the number $\inf_{y \in \Delta} b_x(y)$ is denoted b_x and called the *invasion barrier* of x

6. Local superiority

Strategy x is *locally superior* if there is a neighborhood U of x such that

$$u(x; y) > u(y; y) \text{ for all } y \neq x \text{ in } \Delta \cap U.$$

6.1 Theorem. A strategy x is an ESS if and only if it is locally superior.

Proof. First we show that an ESS is locally superior. An ESS x has a uniform invasion barrier b_x , so that

$$u(x - y; (1 - \epsilon)x + \epsilon y) > 0 \text{ for all } y \neq x \text{ and } 0 < \epsilon < b_x, \quad (6.2)$$

and we want to prove that there is a neighborhood U of x such that

$$u(x - y; y) > 0 \text{ for all } y \in U \cap \Delta.$$

As in the previous section, let F_x be the union of all closed faces of Δ not containing x . Let

$$V = \{(1 - \epsilon)x + \epsilon y_0 : y_0 \in F_x \text{ and } 0 \leq \epsilon < b_x\}.$$

Then V contains $U \cap \Delta$ for some neighborhood of x . (This is because F_x is compact and does not contain x , and therefore the distances of points in F_x from x are bounded away from zero.) We show that $u(x - y; y) > 0$ for $y \neq x$ in this neighborhood.

Such a y equals $(1 - \epsilon)x + \epsilon y_0$ for some $y_0 \in F_x$ and some ϵ with $0 < \epsilon < \bar{\epsilon}$. Because b_x is a uniform invasion barrier for x , we have

$$u(x - y_0; (1 - \epsilon)x + \epsilon y_0) > 0. \quad (6.3)$$

But $(1 - \epsilon)x + \epsilon y_0$ is just y , and one easily checks that

$$x - y_0 = \frac{1}{\epsilon}(x - y).$$

So (6.3) says

$$\frac{1}{\epsilon}u(x - y; y) > 0.$$

Thus an ESS is locally superior.

To see that a locally superior strategy is an ESS, suppose $u(x - z; z) > 0$ for all $z \in U \cap \Delta$. Let $y \in \Delta$ be given ($y \neq x$). We need to show that there exists ϵ_y such that $u(x - y; (1 - \epsilon)x + \epsilon y) > 0$ for $0 < \epsilon < \epsilon_y$.

There certainly exists ϵ_y such that $z = (1 - \epsilon)x + \epsilon y \in U \cap \Delta$ for $0 < \epsilon < \epsilon_y$, and for such z we have $u(x - z; z) > 0$. But $x - z = \epsilon(x - y)$, so this says that $\epsilon u(x - y; (1 - \epsilon)x + \epsilon y) > 0$ for $0 < \epsilon < \epsilon_y$, as required. \square

7. Neutral stability

As defined in Section 4 above, strategy x is an NSS if for every $y \neq x$ there exists $\epsilon_y > 0$ such that

$$u(x; (1 - \epsilon)x + \epsilon y) \geq u(y; (1 - \epsilon)x + \epsilon y) \text{ for } 0 < \epsilon < \epsilon_y. \quad (7.1)$$

The three conditions equivalent to evolutionary stability (summarized at the beginning of Section 4 and proved as Theorems 4.3, 5.5, and 6.1) have analogous conditions equivalent to neutral superiority:

(Theorem 4.3) x is a Nash equilibrium (i.e., $u(x, x) \geq u(y, x)$ for all y) that satisfies in addition

$$\text{if } u(x; x) = u(y; x) \text{ for some } y \neq x, \text{ then } u(x; y) \geq u(y; y).$$

(Analog of Theorem 5.5) The strict inequality in (4.1) holds uniformly (x has a **uniform weak invasion barrier**); that is, there exists ϵ_x such that for every $y \neq x$

$$u(x; (1 - \epsilon)x + \epsilon y) \geq u(y; (1 - \epsilon)x + \epsilon y) \text{ for all } \epsilon < \epsilon_x.$$

(Analog of Theorem 6.1) x is **weakly locally superior**; that is, there is a neighborhood U of x such that

$$u(x; y) \geq u(y; y) \text{ for all } y \neq x \text{ in } \Delta \cap U.$$

On p. 48 of Weibull's book are references to papers in which the proofs of the two analogous theorems can be found.

8. Replicator dynamics

In this and later sections it is most convenient to view a mixed strategy x as the mix in a population of pure-strategy players. That is, x_i is the proportion of i -players in the population.

The **replicator dynamics** are governed by a system of differential equations that model the evolution in time of the population mix. The motivating scenario is that individuals in the population have many encounters with randomly-chosen others, each encounter consisting of one play of the game. When an i -player encounters a j -player, her payoff $u(e^i; e^j) = a_{ij}$ is added to her accumulation of "fitness points". Her accumulated

fitness points over time determine her number of offspring, all of whom grow up to be i -players.

For simplicity we also assume that the number of encounters an individual has is large, uniformly distributed in time, and roughly the same as that for other individuals. Thus we can use the average number of points per play as a determiner of the number of offspring.

Over many encounters, an i -player will have an expected average number of fitness points per play equaling $e^i \cdot Ax = u(e^i; x)$. And across the whole population, the expected average number of fitness points per play is $x \cdot Ax = u(x; x)$.

Finally, since the population is large, we make the simplifying assumption that $x = x(t)$ (for $t \geq 0$) changes continuously over time according to

$$\dot{x}_i = x_i(u(e^i; x) - u(x; x)) \quad \text{for } i = 1, 2, \dots, k. \quad (8.1)$$

These equations determine the *replicator dynamics* of the population. If i -players earn more fitness points than the population average, their population share will increase; if less, it will decrease.

Notice that the equations (8.1) make sense for any $x \in \mathbb{R}^k$, although they have meaning for us only when $x \in \Delta$. We establish first that Δ is an *invariant set* under the replicator dynamics; that is, that if $x(0) \in \Delta$, then $x(t) \in \Delta$ for all $t \in \mathbb{R}$. In fact, more is true:

8.2 Proposition. Any face of Δ is invariant under (8.1), as is the interior of any face.

Proof. This is a consequence of the following facts:

Unions, intersections, complements, interiors, and closures of invariant sets are invariant.

The orbit $\{x(t) : t \in \mathbb{R}\}$ is a continuous curve in \mathbb{R}^k .

$Z_i = \{x : x_i = 0\}$ is invariant.

Any closed face of Δ is invariant.

The first two of these properties are standard results on systems of differential equations; proofs can be found in Chapter 6 of Weibull's book. The third property is an immediate consequence of (8.1).

To prove the fourth, let F be a closed face and let S be the set of vertices of F (the support of any $x \in F$). For any x let $s_F(x) = \sum_{i \in S} x_i$, so that $s_F(x) = 1$ if $x \in F$.

But for such x ,

$$\begin{aligned}\dot{s}_F &= \sum_{i \in F} x_i u(e^i; x) - \sum_{i \in F} x_i u(x; x) \\ &= u(x; x) - u(x; x) \sum_{i \in F} x_i = 0.\end{aligned}\quad \square$$

From now on, all vectors x , y , and so on under consideration are in Δ , and although we usually use the term “point” for such an x , we think of it as a population mix.

9. Stationary points and Nash equilibria

A **stationary point** of a dynamical system $\dot{x} = f(x)$ is a point x such that $\dot{x}_i = 0$ for all i . So a stationary point of the replicator dynamics (8.1) is a vector x such that

$$x_i (u(e^i; x) - u(x; x)) = 0 \quad \text{for } i = 1, 2, \dots, k. \quad (9.1)$$

Thus, obviously, we have

9.2 Proposition. x is a stationary point of (8.1) if and only if

$$u(e^i; x) = u(x; x) \quad \text{for all } i \text{ such that } x_i > 0. \quad \square$$

That is, a population mix is stationary if and only if all strategies that are present have the same expected payoff.

We will refer to a stationary point of the system (8.1) simply as a stationary point.

9.3 Theorem.

- a. Every vertex of Δ is a stationary point.
- b. Every Nash equilibrium is a stationary point.
- c. Every stationary point in the interior of Δ is a Nash equilibrium.
(Thus, in the interior, x is a stationary point iff it is a Nash equilibrium.)

Proof. a. This follows immediately from Proposition 9.2.

b. If x is a Nash equilibrium, then (9.1) holds by Proposition 3.3.

c. If x is an interior stationary point, then $x_i > 0$ for all i , so by Proposition 9.2, $u(e^i; x) = u(x; x)$ for all i ; by Proposition 3.3, x is a Nash equilibrium. \square

A stationary point that is *not* a Nash equilibrium thus lies on a proper face of Δ , and has the property that $u(e^i; x) > u(x; x)$ for some $i \notin S(x)$, while $u(e^i; x) = u(x; x)$ for all

$i \in S(x)$. As we see in the next section, these are the stationary points that are not Lyapunov stable under the replicator dynamics (8.1).

9.4 Proposition. The set of interior stationary points of (8.1) is the intersection of an affine subspace of \mathbb{R}^k with Δ . It is either empty or a singleton unless the matrix on the left side of (9.6) below is singular. In particular, if there is an isolated interior stationary point, then there is exactly one interior stationary point.

Proof. An interior stationary point satisfies $x_i > 0$ and $u(e^i; x) = u(x; x)$ for $i = 1, \dots, k$, and thus

$$u(e^1; x) = u(e^2; x) = \dots = u(e^k; x), \text{ and also } x_1 + \dots + x_k = 1, \quad (9.5)$$

Using that $u(x; y) = x \cdot Ay$ for a $k \times k$ matrix $A = [a_{ij}]$, so that $u(e^i; x) = \sum_{j=1}^k a_{ij}x_j$,

we see that (9.5) amounts to the system of linear equations

$$\begin{bmatrix} a_{11} - a_{k1} & a_{12} - a_{k2} & \dots & a_{1k} - a_{kk} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k-1,1} - a_{k1} & a_{k-1,2} - a_{k2} & \dots & a_{k-1,k} - a_{kk} \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \quad (9.6)$$

So the set of interior stationary points is the intersection of the solution set of (9.6) with the positive orthant in \mathbb{R}^k . \square

9.7 Counterexamples to converse implications. The converses of the implications in Theorem 9.3 are all false in general. In particular,

- Not every stationary point is a vertex.
- Not every stationary point is a Nash equilibrium.
- Not every stationary point that is a Nash equilibrium is in the interior.

We can give examples of all of these with $k = 2$ and payoff functions $u(x; y)$ of the form

$$u(x; y) = [x_1 \quad x_2] \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = x_2(ay_1 + by_2).$$

For such a payoff function we have

$$\begin{aligned} u(e^1; x) &= 0, \\ u(e^2; x) &= ax_1 + bx_2, \text{ and} \\ u(x; x) &= x_2(ax_1 + bx_2), \end{aligned}$$

and the replicator dynamics are governed by

$$\begin{aligned}\dot{x}_1 &= -x_1x_2(ax_1 + bx_2) \\ \dot{x}_2 &= +x_1x_2(ax_1 + bx_2).\end{aligned}$$

Notice that since $x_2 = 1 - x_1$ on Δ_2 , the dynamics are fully described by

$$\dot{x}_1 = -x_1(1 - x_1)(ax_1 + b(1 - x_1)). \quad (9.5)$$

- a. We see that x is a stationary point if and only if x is a vertex ($x_1 = 0$ or 1) or $ax_1 + b(1 - x_1) = 0$. Thus, if $\frac{b}{b-a}$ is strictly between 0 and 1 — that is, if a and b are nonzero with opposite signs — then there is a stationary point that is not a vertex.
- b. The vertex e^2 is a stationary point. It is a Nash equilibrium if and only if $u(e^2; e^2) \geq u(e^2; e^1)$; that is, if and only if $b \geq 0$. So if $b < 0$, there is a stationary point that is not a Nash equilibrium.
- c. As seen in the example for b, e^2 is a Nash equilibrium if and only if $b \geq 0$; it is stationary, and it is not in the interior. \square

10. Dynamic stability

A stationary point x^0 of a dynamical system $\dot{x} = f(x)$ is called **Lyapunov stable** if every neighborhood of x^0 contains a neighborhood B of x^0 such that if $x(0) \in B$, then $x(t) \in B$ for all t . We may use “stable” to mean “Lyapunov stable”.

A stationary point x^0 is **asymptotically stable** if it is Lyapunov stable and in addition there is a neighborhood B^0 such that if $x(0) \in B^0$, then $\lim_{t \rightarrow \infty} x(t) = x^0$.

(We will see below that in the 2-dimensional situation of 9.7 above, any Lyapunov stable stationary point is asymptotically stable. This is not true for replicator dynamics in general. Later we will present counterexamples to this and a number of other implications that will have appeared.)

A stationary point that is not Lyapunov stable is called **unstable**.

In Section 11 we will consider the question of proving that a stationary point is Lyapunov stable or asymptotically stable. In this section we show that stability implies Nash equilibrium, but not conversely.

10.1 Theorem. A Lyapunov stable stationary point is a Nash equilibrium.

Proof. Suppose x^0 is a stationary point but not a Nash equilibrium. Then, as noted following the proof of Theorem 9.3, x^0 is on a proper face of Δ and $u(e^i; x^0) > u(x^0; x^0)$ for some $i \notin S(x^0)$.

Since u is continuous, there exists $\delta > 0$ and a neighborhood U of x^0 such that $u(e^i; y) - u(y; y) > \delta$ for all $y \in U \cap \Delta$. By (8.1), then, $\dot{x}_i > \delta x_i$ for any $x \in U \cap \Delta$.

So if $x(0) \in U \cap \Delta$, then $x_i(t) > x_i(0)e^{\delta t}$, for all t such that $x(t)$ remains in $U \cap \Delta$; and thus certainly $x(t)$ does not remain in any neighborhood of x^0 . Thus a stationary point that is not a Nash equilibrium cannot be Lyapunov stable. \square

10.2 Counterexample to converse. A Nash equilibrium is a stationary point, but it need not be Lyapunov stable. For an example, consider the setup of 9.3 above. The vertex e^2 is a Nash equilibrium if and only if $b \geq 0$. It is Lyapunov stable if, for every $\epsilon > 0$ there is a positive $\delta \leq \epsilon$ such that if x is within δ of e^2 then $\dot{x}_1 < 0$. This is the same as saying that $\dot{x}_1 < 0$ if x_1 is sufficiently small. But (9.2) shows that this is true if $a < 0$. So if $a < 0 \leq b$, then e^2 is a Nash equilibrium but not Lyapunov stable. \square

11. Proving stability: Lyapunov's Theorem and linearization

Here we state a couple of theorems that we will need for proving the asymptotic or Lyapunov stability of stationary points in the replicator dynamics. In Section 12 we will use Lyapunov's Theorem 11.2 to prove that an evolutionarily stable population mix is asymptotically stable in the replicator dynamics, and a neutrally stable mix is Lyapunov stable. In Section 17 we will use Theorem 11.3 on linearization to prove asymptotic stability in certain examples.

Suppose $x(t)$ follows a system of differential equations with (forward and backward) orbits in $C \subseteq \mathbb{R}^k$, with $x(0) = x^0$. A **Lyapunov function** for $x(t)$ at x^0 is a continuously differentiable function $H(x)$ on a neighborhood D of x^0 , satisfying

$$\begin{aligned} H(x^0) &= 0, \\ H(x) &> 0 \text{ if } x \neq x^0, \text{ and} \\ \dot{H}(x) &\leq 0 \text{ for all } x \in D. \end{aligned} \tag{11.1}$$

Here $\dot{H}(x)$ denotes the time derivative of $H(x)$, taken as $x = x(t)$ follows the dynamics in question. That is,

$$\dot{H}(x) = \nabla H(x) \cdot \dot{x} = \sum_{i=1}^k \frac{\partial}{\partial x_i} H(x(t)) \dot{x}_i.$$

$H(x)$ is a *strict Lyapunov function* if the following strengthening of the third requirement in (11.1) holds.

$$\dot{H}(x) < 0 \text{ for all } x \in D \text{ other than } x = x^0.$$

The following theorem provides a method — Lyapunov's direct method — for proving that a stationary point of a system of differential equations is Lyapunov stable or asymptotically stable. It is a special case of theorems found on pp. 245-249 of Weibull's book.

11.2 Theorem. Let x^0 be a stationary point of a system of differential equations.

If the system has a Lyapunov function at x^0 , then x^0 is a Lyapunov stable stationary point.

If it has a strict Lyapunov function at x^0 , then x^0 is an asymptotically stable stationary point.

11.3 Theorem (linearization). Suppose x^0 is a stationary point of the system $\dot{x} = f(x) = [f_1(x) \cdots f_k(x)]^T$. If $f(x)$ is continuously differentiable at x^0 and all the eigenvalues of the Jacobian of $f(x)$ at x^0 have negative real parts, then x^0 is an asymptotically stable stationary point.

The Jacobian of $f(x)$ is the matrix whose ij entry is $\frac{\partial f_i}{\partial x_j}$.

12. Evolutionary stability and dynamic stability

The main results here, in Theorem 12.4, are that evolutionary stability implies asymptotic stability, and that neutral stability implies Lyapunov stability. To prove these we use Lyapunov's direct method (Theorem 11.2); we also need Jensen's Inequality (12.7 below).

Let x be any population mix in Δ . The *Kullback-Leibler relative entropy* of y with respect to x is defined as

$$H_x(y) = - \sum_{i \in S(x)} x_i \log(y_i/x_i),$$

and this definition makes sense for any y in the set

$$Q_x = \{y : S(y) \supseteq S(x)\}.$$

This is the set of all y such that $y_i > 0$ whenever $x_i > 0$. It is an open set in \mathbb{R}^k , because x is in the interior of the face of Δ spanned by the e^i for $i \in S(x)$, and Q_x consists of all

y that are not on the boundary of that face. (Unless x is a vertex, in which case Q_x is the complement of $\{x\}$.)

12.1 Proposition. $H_x(y) \geq 0$ for all $y \in Q_x$, with equality if and only if $y = x$.

Proof. It is obvious that $H_x(x) = 0$. For any $y \in Q_x$ other than $y = x$, define the discrete random variable Z to have the nonzero value $z_i = y_i/x_i$ with probability x_i , and let $\varphi(z) = -\log(z)$. The function φ is concave upwards. So

$$\begin{aligned} H_x(y) &= E\varphi(Z) \\ &\geq \varphi(EZ) \quad \text{by Jensen's inequality (12.7 below)} \\ &= -\log \sum_{i \in S(x)} y_i \\ &\geq -\log \sum_{i=1}^k y_i = \log 1 = 0. \end{aligned} \tag{12.2}$$

Now the first inequality in (12.2) holds strictly unless, for some a and b , $\log(y_i/x_i) = ay_i/x_i + b$ for all $i \in S(x)$. This can happen only if $S(x)$ is a singleton. In this case x is a vertex, and since $y \neq x$, $S(y) \neq S(x)$. But then the second inequality in (12.2) holds strictly. \square

12.3 Corollary.

- If $\dot{H}_x(y) \leq 0$ for all y in a neighborhood of x , then x is a Lyapunov stable stationary point of the replicator dynamics.
- If $\dot{H}_x(y) < 0$ for all $y \neq x$ in a neighborhood of x , then x is an asymptotically stable stationary point of the replicator dynamics.

Proof. This follows immediately from Proposition 12.1 and Theorem 11.2. \square

Remarkably, the time derivative of H_x relates directly to the local superiority of x :

12.4 Proposition. $\dot{H}_x(y) = u(y; y) - u(x; y)$ for any $y \in Q_x$.

Proof.

$$\begin{aligned}
 \dot{H}_x(y) &= -\sum_{i \in S(x)} \frac{x_i}{y_i/x_i} \cdot \frac{1}{x_i} \dot{y}_i \\
 &= -\sum_{i \in S(x)} \frac{x_i}{y_i} y_i (u(e^i; y) - u(y; y)) \\
 &= -\sum_{i \in S(x)} x_i (u(e^i; y) - u(y; y)) \\
 &= -(u(x; y) - u(y; y)). \quad \square
 \end{aligned}$$

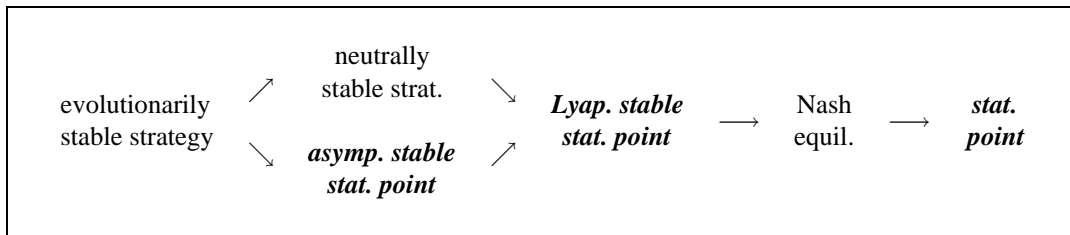
12.5 Theorem. If x is evolutionarily stable, then x is an asymptotically stable stationary point of the replicator dynamics. If x is neutrally stable, then x is a Lyapunov stable stationary point of the replicator dynamics.

Proof. By Theorem 6.1, x is evolutionarily stable if and only if it is locally superior; that is if and only if $u(x; y) > u(y; y)$ for all y in some neighborhood of x . In this case $\dot{H}_x(y) < 0$ in a neighborhood of x , so x is an asymptotically stable stationary point.

By the extension of Theorem 6.1 mentioned in Section 7, x is neutrally stable if and only if $u(x; y) \geq u(y; y)$ for all y in some neighborhood of x . In this case $\dot{H}_x(y) \leq 0$ in a neighborhood of x , so x is a Lyapunov stable stationary point. \square

At this point we have established the following implications.

Table 12.6



That evolutionary stability implies neutral stability, and that asymptotic stability implies Lyapunov stability, follow directly from the definitions. The other implications were established in Theorems 9.3, 10.1, and 12.4.

The properties in bold italics are dynamic stability properties, defined in connection with the replicator dynamics. The others are static, game-theoretic stability properties, defined solely in terms of the payoff function $u(x; y)$.

We will use the following abbreviations for the properties listed in Table 12.6:

SP: stationary point
 NE: Nash equilibrium
 LSSP: Lyapunov stable stationary point
 ASSP: asymptotically stable stationary point
 NSS: neutrally stable strategy
 ESS: evolutionarily stable strategy

As we will see, no implication not entailed by Table 12.6 holds in general. To confirm this, we will need examples of the following.

A SP that is not a NE
 A NE that is not a LSSP
 A NSS that is not an ASSP
 (this will also be an LSSP that is not an ASSP, and a NSS that is not an ESS)
 An ASSP that is not a NSS
 (this will also be an LSSP that is not a NSS, and an ASSP that is not an ESS)

We have already seen in (9.7) a SP that is not a NE. Much of what follows will be taken up with giving examples of the others.

We used the following in the proof of Theorem 12.1.

12.7 Jensen's Inequality. If Z is a random variable and $\varphi(z)$ is concave upward on an interval containing the set of possible values of Z , and if EZ and $E\varphi(Z)$ are both finite, then

$$\varphi(EZ) \leq E\varphi(Z). \quad (12.8)$$

Moreover, the inequality (12.8) is strict unless $\varphi(Z) = aZ + b$ with probability 1 for some constants a and b .

Proof. Let $y = az + b$ be a supporting line for the convex curve $y = \varphi(z)$ at the point $(z = EZ, y = \varphi(EZ))$. That is,

$$aEZ + b = \varphi(EZ) \quad (12.9)$$

and $az + b \leq \varphi(z)$ for $x \in I$. So

$$aZ + b \leq \varphi(Z) \text{ with probability 1,}$$

and therefore

$$aEZ + b \leq E\varphi(Z). \quad (12.10)$$

This inequality holds strictly unless $aZ + b = \varphi(Z)$ with probability 1. Combining (12.9) and (12.10) gives the theorem. \square

13. Invariance of the dynamics under matrix operations

Before proceeding it will be helpful to establish some reductions that simplify the analysis of the population dynamics for specific games. In particular, we show that if the payoff matrix is multiplied by a positive constant, or if any constant is added to any column of the payoff matrix, the six static and dynamic stability properties in Table 12.6 are unchanged. In addition, the orbits of the replicator dynamics are unchanged; all that changes is the rate at which the population evolves.

For a symmetric 2-player game with k actions and payoff function $u(x; y) = x \cdot Ay$, consider the game with payoff function $v(x; y)$ determined by the matrix $B = aA + bE^j$, where E^j is the matrix with 1's in the j^{th} column and zeros elsewhere. Thus B is the result of multiplying A by a constant and adding a constant to one of the columns.

13.2 Proposition. If A is any $k \times k$ matrix and $B = aA + bE^j$ where a and b are any real numbers, we have, for any x and z in Δ ,

$$x \cdot Bz = ax \cdot Az + bz_j. \quad (13.2)$$

Consequently, for any x, y , and z in Δ ,

$$(x - y) \cdot Bz = a(x - y) \cdot Az. \quad (13.3)$$

Proof.

$$\begin{aligned} x \cdot Bz &= x \cdot (aA + bE^j)z \\ &= ax \cdot Az + bz_j \sum_1^k x_j \\ &= ax \cdot Az + bz_j. \end{aligned}$$

And (13.3) follows:

$$(x - y) \cdot Bz = ax \cdot Az - ay \cdot Az + bz_j - bz_j = a(x - y) \cdot Az. \quad \square$$

Recall the following definitions, from Sections 3 and 4, for an arbitrary payoff matrix A and $x \in \Delta$.

x is a Nash equilibrium if and only if $(x - y) \cdot Ax \geq 0$ for all $y \in \Delta$.

x is neutrally stable if and only if $(x - y) \cdot A((1 - \epsilon)x + \epsilon y) \geq 0$ for all y and all sufficiently small ϵ .

x is asymptotically stable if and only if $(x - y) \cdot A((1 - \epsilon)x + \epsilon y) > 0$ for all y and all sufficiently small ϵ .

13.4 Theorem. The Nash equilibria, neutrally stable strategies, and evolutionarily stable strategies for $B = aA + bE^j$ are the same as for A , as long as a is positive.

Proof. This follows from (13.3) and the definitions repeated just above. \square

13.5 Proposition. If $a > 0$, then the replicator dynamics for $B = aA + bE^j$ are

$$\dot{x}_i = ax_i((e^i - x) \cdot Ax). \quad (13.6)$$

Proof. The replicator dynamics for B are

$$\dot{x}_i = x_i((e^i - x) \cdot Bx),$$

and the result follows immediately from (13.3). \square

13.7 Theorem. The orbits of the replicator dynamics for B are the same as those for A . As a consequence, the stationary points for B are the same as those for A , and, if a is positive, so are the Lyapunov stable stationary points and the asymptotically stable stationary points.

Proof. This follows from (13.6); the replicator dynamics for A and B differ only in that those for B have the right side multiplied by the positive constant a . \square

Again we note that the replicator dynamics are not identical for A and B ; however, they differ only in that the *rate* of evolution for B is a times that for A .

14. The 2x2 case: dynamics

Here we characterize the replicator dynamics of any two-player symmetric game with only two actions. We identify all the stationary points and establish conditions under which they are Lyapunov stable and asymptotically stable. In the next section we will investigate the static equilibrium properties of these points (NE, NSS, ESS).

The replicator dynamics are

$$\begin{aligned} \dot{x}_1 &= x_1((e^1 \cdot Ax - x \cdot Ax)) \quad \text{and} \\ \dot{x}_2 &= x_2((e^2 \cdot Ax - x \cdot Ax)). \end{aligned} \quad (14.1)$$

Notice first that for any $x = [x_1 \ x_2] \in \Delta_2$, we have $x_2 = 1 - x_1$ and therefore $\dot{x}_2 = -\dot{x}_1$, and so the replicator dynamics are described completely by

$$\dot{x}_1 = x_1((e^1 \cdot Ax - x \cdot Ax)). \quad (14.2)$$

From Theorem 13.7 we see that we need to study the dynamics only for matrices of the form

$$A = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \tag{14.3}$$

for real numbers a and b , because any 2×2 matrix can be reduced to this form by subtracting appropriate numbers from the columns (in this case, subtracting the first row from both rows). For A of this form, (14.2) becomes

$$\dot{x}_1 = x_1 x_2 (a x_1 + b x_2) = x_1 (1 - x_1) ((b - a)x_1 - b). \tag{14.4}$$

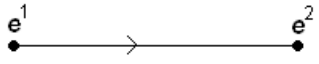
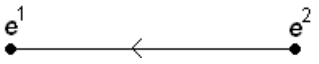
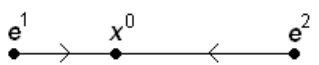
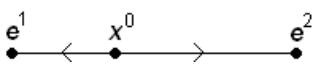
It is clear from (14.4) that $x = [x_1 \ x_2]$ is a stationary point if $x_1 = 0$, if $x_1 = 1$, or if $(b - a)x_1 - b = 0$. The following is immediate.

14.5 Proposition. In the trivial case $a = b = 0$, all $x \in \Delta_2$ are stationary points. Otherwise, there are either two or three stationary points in Δ_2 : e^1 and e^2 are stationary points, and, if $0 < \frac{b}{b-a} < 1$, $x^0 = [\frac{b}{b-a}, \frac{-a}{b-a}]^T$ is a stationary point. □

In the rest of this section we will determine, under all possible conditions on a and b , what the orbits of the dynamics are, and thus classify the stationary points as unstable (USP), Lyapunov stable (LSSP), or asymptotically stable (ASSP).

[In the next section we will examine the stationary points for their static stability properties, classifying them as not equilibria, Nash equilibria (NE), neutrally stable strategies (NSS), or evolutionarily stable strategies (ESS).]

It is easy to see at the outset that except in the trivial case $a = b = 0$ (Case 0 below), the orbits must be as in one of the following four pictures, and the dynamic stability of the stationary points e^1, e^2 , and x^0 must be as indicated.

	orbits	e^1	e^2	x^0
A		USP	ASSP	—
B		ASSP	USP	—
C		USP	USP	ASSP
D		ASSP	ASSP	USP

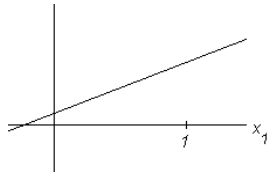
[Two additional cases, in which there is an interior stationary point but the orbits on both sides go in the same direction, are impossible here, because the function $y = (b - a)x_1 - b$ is linear and must change sign at a zero.]

We note also that except in the trivial case $a = b = 0$, all stationary points are either unstable or asymptotically stable. We will have to consider 3×3 games to find nontrivial situations with Lyapunov stable stationary points that are not asymptotically stable.

Excepting the trivial case $a = b = 0$, there are in fact eight distinct cases for the line $y = (b - a)x_1 - b$; each leading to one of the orbit diagrams above.

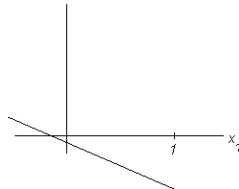
Case 0. $a = b = 0$. In this trivial case all $x \in \Delta_2$ are stationary points, and they are necessarily Lyapunov stable.

Case 1. $b - a > 0, \frac{b}{b-a} \leq 0$. [That is, $a < b \leq 0$.]



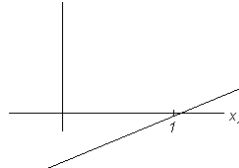
Here $\dot{x}_1 > 0$ for all $x_1 \in (0, 1)$, so the orbits are as in B above.

Case 2. $b - a < 0, \frac{b}{b-a} \leq 0$. [That is, $a > b \geq 0$.]



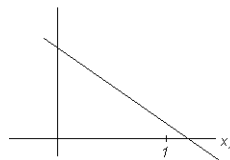
Here $\dot{x}_1 < 0$ for all $x_1 \in (0, 1)$, so the orbits are as in A above.

Case 3. $b - a > 0, \frac{b}{b-a} \geq 1$. [That is, $b > a \geq 0$.]



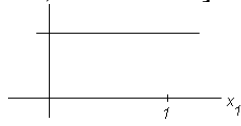
Here $\dot{x}_1 < 0$ for all $x_1 \in (0, 1)$, so the orbits are as in A above.

Case 4. $b - a < 0, \frac{b}{b-a} \geq 1$. [That is, $b < a \leq 0$.]



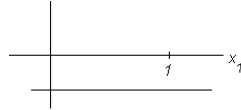
Here $\dot{x}_1 > 0$ for all $x_1 \in (0, 1)$, so the orbits are as in B above.

Case 5. $b - a = 0, b < 0$. [That is, $b = a < 0$.]



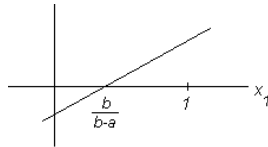
Here $\dot{x}_1 > 0$ for all $x_1 \in (0, 1)$, so the orbits are as in B above.

Case 6. $b - a = 0, b > 0$. [That is, $b = a > 0$.]



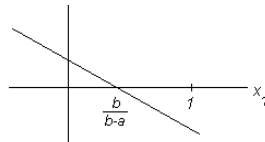
Here $\dot{x}_1 < 0$ for all $x_1 \in (0, 1)$, so the orbits are as in A above.

Case 7. $b - a > 0, 0 < \frac{b}{b-a} < 1$. [That is, $a < 0 < b$.]



Here $\dot{x}_1 < 0$ for $x_1 < \frac{b}{b-a}$ and $\dot{x}_1 > 0$ for $x_1 > \frac{b}{b-a}$, so the orbits are as in D above.

Case 8. $b - a < 0, 0 < \frac{b}{b-a} < 1$. [That is, $a > 0 > b$.]

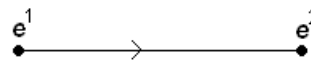
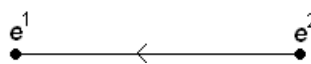
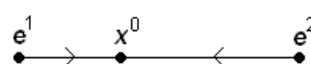
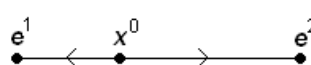


Here $\dot{x}_1 > 0$ for $x_1 < \frac{b}{b-a}$ and $\dot{x}_1 < 0$ for $x_1 > \frac{b}{b-a}$, so the orbits are as in C above.

Combining these, we get

14.6 Proposition. For the replicator dynamics as in (14.1), with $A = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix}$ and $x^0 = \frac{1}{b-a}[b \ -a]^T$, the orbits and dynamic stability properties of the stationary points are, except for the trivial case $a = b = 0$, as shown in Table 14.7. \square

Table 14.7

	orbits	e^1	e^2	x^0	Cases
$a \geq 0, b \geq 0$	A 	USP	ASSP	—	2,3,6
$a \leq 0, b \leq 0$	B 	ASSP	USP	—	1,4,5
$a > 0 > b$	C 	USP	USP	ASSP	8
$a < 0 < b$	D 	ASSP	ASSP	USP	7

15. The 2x2 case: static stability properties

Now we want to place the stationary points on the diagram of implications in Table 12.6; that is, decide whether they are Nash equilibria, neutrally stable strategies, or evolutionarily stable strategies. There are three kinds of stationary point:

For any unstable stationary point, we need to ascertain whether or not it is a Nash equilibrium.

For any Lyapunov stable stationary point that is not asymptotically stable, we need to ascertain whether or not it is neutrally stable. (Such points exist only in the trivial case $a = b = 0$.)

And for any asymptotically stable stationary point, we need to ascertain whether it is evolutionarily stable, neutrally stable, or only a Nash equilibrium.

Orbit diagram A: In this case $a \geq 0$ and $b \geq 0$, and at least one is strictly positive.

e^1 is an unstable stationary point. It is a Nash equilibrium if and only if $u(e^1; e^1) \geq u(e^2; e^1)$, by Proposition 3.3; that is, if and only if $0 \geq a$. So e^1 is a Nash equilibrium if and only if $a = 0$. This is Case 6 of the previous section.

e^2 is an asymptotically stable stationary point. By Theorem 6.1 and its counterpart for neutrally stable strategies (mentioned at the end of Section 7), e^2 is [evolutionarily, neutrally] stable if and only if $u(e^2; x) [> , \geq] u(x; x)$ for all x near e^2 ; that is, for $x = [\epsilon \quad 1 - \epsilon]$ for small positive ϵ . But for such x ,

$$\begin{aligned} u(e^2; x) - u(x; x) &= a\epsilon + b(1 - \epsilon) - (1 - \epsilon)(a\epsilon + b(1 - \epsilon)) \\ &= \epsilon(a\epsilon + b(1 - \epsilon)) \\ &= (a - b)\epsilon^2 + b\epsilon. \end{aligned}$$

This is positive for small positive ϵ if either $b > 0$ or $b = 0 < a$; but in the situation of orbit diagram A, one of these conditions holds. So e^2 is evolutionarily stable.

Orbit diagram B: Here $a \leq 0$ and $b \leq 0$ and at least one is strictly negative. The situation is symmetric to that of orbit diagram A, with a and b replaced by $-b$ and $-a$, respectively. So the unstable stationary point e^2 is a Nash equilibrium if $b = 0$, and the asymptotically stable stationary point e^1 is evolutionarily stable.

[The transformation from diagram A into diagram B is this: subtract the second row of A from both rows, and then exchange the roles of the two players by interchanging the rows and the columns of the matrix.]

Orbit diagram C: Here $a > 0 > b$.

The unstable stationary point e^1 is a Nash equilibrium if and only if $u(e^1; e^1) \geq u(e^2; e^1)$; that is, if and only if $0 \geq a$. So e^1 is not a Nash equilibrium. Similarly, the unstable stationary point e^2 is not a Nash equilibrium.

The asymptotically stable stationary point x^0 is [evolutionarily, neutrally] stable if and only if $u(x^0; x) [> , \geq] u(x; x)$ for all x near x^0 . Notice that in this case $x^0 = \frac{1}{a-b}[-b \quad a]$ and $a - b$ is positive. Points x near x^0 are of the form $\frac{1}{a-b}[-b + \epsilon \quad a - \epsilon]$ for small ϵ , which may be positive or negative. So x^0 is [evolutionarily, neutrally] stable if and only if $u(x^0 - x; x) [> , \geq] 0$. But $x^0 - x = \frac{1}{a-b}[-\epsilon \quad \epsilon]$, and so

$$\begin{aligned} u(x^0 - x; x) &= \frac{1}{(a-b)^2}[-\epsilon \quad \epsilon] \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \begin{bmatrix} -b + \epsilon \\ a - \epsilon \end{bmatrix} \\ &= \frac{1}{(a-b)^2}[-\epsilon \quad \epsilon] \begin{bmatrix} 0 \\ (a-b)\epsilon \end{bmatrix} = \frac{\epsilon^2}{a-b}, \end{aligned}$$

which is positive. So in this case x^0 is evolutionarily stable.

Orbit diagram D: Here $a < 0 < b$.

By Theorem 9.3, a stationary point in the interior of Δ_2 is a Nash equilibrium, so the unstable stationary point x^0 is a Nash equilibrium in this case.

The asymptotically stable stationary point e^1 is [evolutionarily, neutrally] stable if and only if $u(e^1; x) - u(x; x) \begin{bmatrix} > \\ \geq \end{bmatrix} 0$ for all x near e^1 ; that is, for all $x = [1 - \epsilon \quad \epsilon]$ for small positive ϵ . Now $e^1 - x = [\epsilon \quad -\epsilon]$ for such x , so

$$\begin{aligned} u(e^1; x) - u(x; x) &= [\epsilon \quad -\epsilon] \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \begin{bmatrix} 1 - \epsilon \\ \epsilon \end{bmatrix} \\ &= [\epsilon \quad -\epsilon] \begin{bmatrix} 0 \\ a(1 - \epsilon) + b\epsilon \end{bmatrix} \\ &= -\epsilon((b - a)\epsilon + a) \\ &= -(b - a)\epsilon^2 - a\epsilon. \end{aligned}$$

This is positive for small positive ϵ because $a < 0$ in this case. So e^1 is evolutionarily stable.

Again we could make the transformation used for orbit diagram B to see that e^2 is evolutionarily stable, but we do the calculations as a check. e^2 is [evolutionarily, neutrally] stable if and only if $u(e^2; x) - u(x; x) \begin{bmatrix} > \\ \geq \end{bmatrix} 0$ for all x near e^2 ; that is, for $x = [\epsilon \quad 1 - \epsilon]$ for small positive ϵ . For such x , $e^2 - x = [-\epsilon \quad \epsilon]$, so

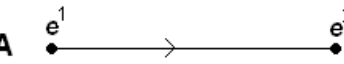
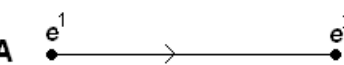

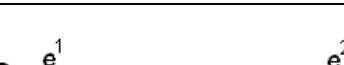
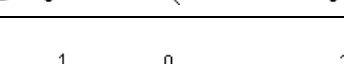
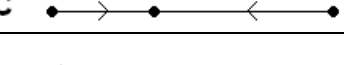
$$\begin{aligned} u(e^2; x) - u(x; x) &= [-\epsilon \quad \epsilon] \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \begin{bmatrix} \epsilon \\ 1 - \epsilon \end{bmatrix} \\ &= [-\epsilon \quad \epsilon] \begin{bmatrix} 0 \\ a\epsilon + b(1 - \epsilon) \end{bmatrix} \\ &= \epsilon(a\epsilon + b(1 - \epsilon)) \\ &= b\epsilon + (a - b)\epsilon^2. \end{aligned}$$

Since $b > 0$ in this case, this is positive, so e^2 is evolutionarily stable.

The trivial case. Here $u(x; y) = 0$ for all x and y and all $x \in \Delta_2$ are Lyapunov stable, but not asymptotically stable, stationary points. A point x is neutrally stable if and only if $u(x; y) \geq u(y; y)$ for all y near x . Since these are both zero, all x are neutrally stable.

15.1 Proposition. For the replicator dynamics as in (14.1), with $A = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix}$ and $x^0 = \frac{1}{b-a}[b \ -a]^T$, the orbits and the static and dynamic stability properties of the stationary points are as shown in Table 15.2. In each case, the indicated property is the strongest that can be asserted. □

Table 15.2

	orbits	e^1	e^2	x^0
$a = 0, b > 0$	A 	NE	ESS	—
$a > 0, b \geq 0$	A 	USP	ESS	—
$a < 0, b = 0$	B 	ESS	NE	—
$a \leq 0, b < 0$	B 	ESS	USP	
$a > 0 > b$	C 	USP	USP	ESS
$a < 0 < b$	D 	ESS	ESS	NE
$a = b = 0$	All points in Δ_2 are NSS.			

Counterexamples to implications not entailed in Table 12.6. At the end of Section 12 we listed the examples needed to show that only the implications entailed by Table 12.6 hold in general. We have now found all but one of these in the above analysis of 2×2 games:

A SP that is not a NE: Several cases as shown above.

A NE that is not a LSSP: x^0 in the case $a < 0 < b$.

A NSS that is not an ASSP: All points in the trivial case.

An ASSP that is not a NSS: There are no examples of this among 2×2 games.

[In case it is objected that one of the examples exists only in the trivial case, we mention that

for 3×3 games with payoff matrices of the form $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ b & c & d \end{bmatrix}$, the situation restricted to the

face containing e^1 and e^2 is trivial, although the game is not. For at least some nontrivial values of a, b, c , and d , all points on this face are neutrally stable stationary points but are not asymptotically stable. Such games, in which two of the pure strategies behave identically with respect to each other but differently with respect to a third, are studied extensively in the author's earlier survey, referred to in the Bibliography.]

16. Results and examples on evolutionary and neutral stability

Our main goal for the rest of this survey is to find an example of a game with a point x that is an ASSP but not a NSS. In this section we show a relatively simple way to determine whether an interior stationary point is a NSS or an ESS.

Recall that $x \in \Delta$ is an ESS [resp. NSS] if and only if $u(x; y) > [\text{resp. } \geq] u(y; y)$ for all y sufficiently near but not equal to x in Δ . (This is the characterization of evolutionary stability by local superiority given in Theorem 6.1 and its analog in Section 7.)

Recall that the support $S(x)$ of x is the set of indices i for which $x_i > 0$. It will be convenient to consider the set $R(x) = \{r \in \mathbb{R}^k : r \neq 0 \text{ and } x + r \in \Delta\}$. Notice that if $r = [r_1, \dots, r_k]^T \in R(x)$, then $\sum r_i = 0$, and $r_i \geq 0$ if $i \notin S(x)$.

16.1 Proposition. Suppose x is a NE of the game determined by $u(x; y) = x \cdot Ay$.

- (a) If $u(r; r) \leq 0$ for all $r \in R(x)$, then x is a NSS.
- (b) If x is an interior point of Δ , then x is a NSS [resp. ESS] if and only if $u(r; r) \leq 0$ [resp. < 0] for all $r \in R(x)$.

Proof. As noted in Section 7, x is an NSS if and only if it is weakly locally superior; that is, if and only if $u(y - x; y) \leq 0$ for y near x in Δ . This is the same as saying that $u(r; x + r) \leq 0$ for sufficiently small $r = y - x \in R(x)$. Since $u(r; x + r) = r \cdot Ax + r \cdot Ar$, we have that

$$x \text{ is a NSS} \iff r \cdot Ax + r \cdot Ar \leq 0 \text{ for } r \in R(x). \quad (16.2)$$

Similarly, from Theorem 6.1 on local superiority, we have that

$$x \text{ is an ESS} \iff r \cdot Ax + r \cdot Ar < 0 \text{ for } r \in R(x). \quad (16.3)$$

Without loss of generality let $S(x) = \{1, \dots, j\}$. Since x is a NE, it follows that

$$Ax = [c, \dots, c, c_{j+1}, \dots, c_k]^T,$$

where $c = u(x; x)$ and $c_i \leq c$ for $i > j$. Moreover, if $r \in R(x)$, then $r_i \geq 0$ for $i > j$. So

$$r \cdot Ax = c(r_1 + \dots + r_j) + c_{j+1}r_{j+1} + \dots + c_k r_k \leq c(r_1 + \dots + r_k) = 0.$$

Assertion (a) follows because of (16.2).

If x is an interior NE, then $Ax = [c, \dots, c]$ and $r \cdot Ax = 0$, and assertion (b) follows because of (16.2) and (16.3). □

16.4 Example. Consider $u(x; y) = x \cdot Ay$ where

$$A = \begin{bmatrix} 3 & 40 & 0 \\ 0 & 8 & 35 \\ 15 & 0 & 28 \end{bmatrix}.$$

Since all the row sums of A are equal, it follows that $x = \frac{1}{3}[1 \ 1 \ 1]^T$ satisfies $Ax = \frac{1}{3}[43 \ 43 \ 43]^T$. So $u(e^i; x) = \frac{43}{3}$ for $i = 1, 2, 3$, and thus x is an interior NE.

a. We show using Proposition 16.1 that x is not a NSS.

We need only show that $u(r; r) = r \cdot Ar$ is positive for some r in $R(x)$. Such r are of the form $r = \frac{1}{18}[\epsilon \ \delta \ -\epsilon - \delta]^T$ for small ϵ and δ . For such an r we have

$$\begin{aligned} r \cdot Ar &= \frac{1}{18^2}[\epsilon \ \delta \ -\epsilon - \delta] \begin{bmatrix} 3 & 40 & 0 \\ 0 & 8 & 35 \\ 15 & 0 & 28 \end{bmatrix} \begin{bmatrix} \epsilon \\ \delta \\ -\epsilon - \delta \end{bmatrix} \\ &= \frac{1}{18^2}[16\epsilon^2 + \delta^2 + 46\epsilon\delta], \end{aligned}$$

which is certainly positive for some values of ϵ and δ .

b. We show incidentally that x is the only interior stationary point.

For any $w = [w_1 \ w_2 \ w_3]^T$ we have

$$Aw = \begin{bmatrix} 3w_1 + 40w_2 \\ 8w_2 + 35w_3 \\ 15w_1 + 28w_3 \end{bmatrix}.$$

If w is an interior stationary point, then all the entries of Aw equal $w \cdot Aw$, because $\dot{w}_i = w_i(e^i \cdot Aw - w \cdot Aw)$. So we have

$$3w_1 + 40w_2 = 8w_2 + 35w_3 = 8w_2 + 35(1 - w_1 - w_2),$$

which implies that

$$38w_1 + 47w_2 = 35, \tag{16.5}$$

and also

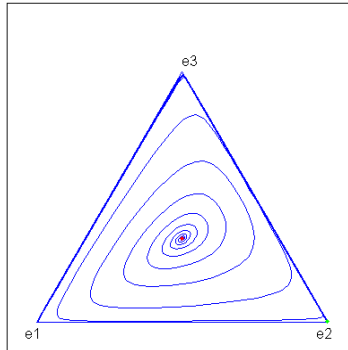
$$3w_1 + 40w_2 = 15w_1 + 28w_3 = 15w_1 + 28(1 - w_1 - w_2),$$

which implies that

$$16w_1 + 68w_2 = 28. \tag{16.6}$$

One can check that (16.5) and (16.6) imply that $w_1 = w_2 = \frac{1}{3}$, and the assertion follows. \square

A plot of one orbit, generated using MATLAB, suggests that x is (globally) asymptotically stable:

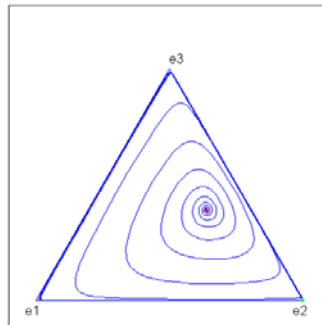


We will see in Section 17 that x is indeed an ASSP. Note that because x is not a NSS, and thus not an ESS, the Kullback-Leibler relative entropy (see Section 12) will not be a Lyapunov function.

16.7 Example. Weibull's Example 3.9 is the game $u(x; y) = x \cdot Ay$ where

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 0 & 1 & 5 \\ 5 & 0 & 4 \end{bmatrix}.$$

Here arguments similar to those in Example 16.4 show that the point $x = \frac{1}{18}[3 \ 8 \ 7]^T$ is the unique interior SP and is a NE but is not an NSS; and the plot of an orbit suggests that x is asymptotically stable.



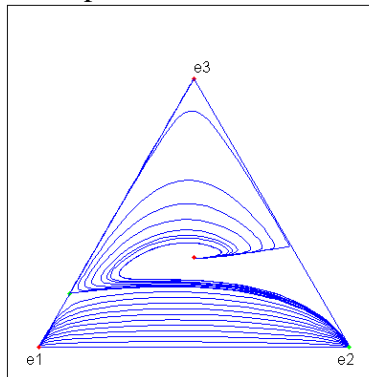
We got Example 16.4 from this example by using the “centering” transformation suggested by Hofbauer and Sigmund (Exercise 7.1.3, p. 68). They observe that if $x = [x_1 \cdots x_k]^T$ is a stationary point of the replicator dynamics for $u(x; y) = x \cdot Ay$, then $y = K[c_1x_1 \cdots c_kx_k]^T$ (with $K = 1/\sum_1^k c_jx_j$) is a stationary point of the replicator dynamics for $u(x; y) = x \cdot By$, where $b_{ij} = \frac{a_{ij}}{c_j}$.

It turns out that checking for an ASSP is much simplified if the interior NE in question is the center $\frac{1}{k}[1 \cdots 1]^T$. This happens when, as in Example 16.4, the row sums of A are all equal. The transformation that moves x to $y = \frac{1}{k}[1 \cdots 1]^T$ makes the row sums of A equal.

16.8 Example. On page 71 of Hofbauer and Sigmund's book is the matrix

$$A = \begin{bmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{bmatrix}.$$

This matrix has constant row sums, and as a result $x = \frac{1}{3}[1 \ 1 \ 1]^T$ is the unique interior NE. As in Example 16.4 we can check that it is not a NSS: specifically, for $r = [\epsilon \ \delta \ -\epsilon - \delta]^T$, we have that $r \cdot Ar = 5\epsilon^2 - 8\delta^2$, which is both positive and negative in any neighborhood of 0. Here the plot of some orbits is as follows.



This is strikingly different from the pictures for the previous two examples; nevertheless, as we will see, the center point x is asymptotically stable.

Thus, once we have proved asymptotic stability in the three examples of this section, we will have completed the program, set out at the end of Section 15, of showing that none of the implications not entailed in Table 12.6 hold in general.

17. Proving asymptotic stability using linearization

As we noted at the end of Example 16.4, when x is not a NSS we cannot prove that x is an ASSP using the Kullback-Leibler relative entropy as a Lyapunov function. (If we could, then x would be an ESS and thus a NSS.) We could try to find another Lyapunov function in such a case, but instead we linearize the replicator dynamics and use Theorem 11.3.

This turns out to be simplified greatly if the row sums of the matrix A are all equal, in which case $x = \frac{1}{k}[1 \ 1 \ \cdots \ 1]^T$ is the unique interior SP, and it is a NE.

As always, the game is determined by $u(x; y) = x \cdot Ay$ where $A = [a_{ij}]$ is an arbitrary $k \times k$ matrix. The replicator dynamics are $\dot{x} = f(x) = [f_1(x) \ \cdots \ f_k(x)]$; that is,

$$\dot{x}_i = f_i(x) = x_i(e^i \cdot Ax - x \cdot Ax). \quad (17.1)$$

The Jacobian of f at x is the $k \times k$ matrix $J = Jf(x)$ whose ij entry is $\frac{\partial f_i}{\partial x_j}(x)$.

17.2 Proposition. $\frac{\partial}{\partial x_j}(e^i \cdot Ax) = a_{ij}$.

Proof. $e^i \cdot Ax = a_{i1}x_1 + \cdots + a_{ik}x_k$. □

17.3 Proposition. $\frac{\partial}{\partial x_j}(x \cdot Ax) = e^j \cdot (A + A^T)x$.

Proof. $x \cdot Ax = a_{jj}x_j^2 + \sum_{r \neq j} a_{jr}x_jx_r + \sum_{s \neq j} a_{sj}x_sx_j + (\text{terms not involving } x_j)$.

Therefore

$$\begin{aligned} \frac{\partial}{\partial x_j}(x \cdot Ax) &= 2a_{jj}x_j + \sum_{r \neq j} a_{jr}x_r + \sum_{s \neq j} a_{sj}x_s \\ &= \sum_{r=1}^k a_{jr}x_r + \sum_{s=1}^k a_{sj}x_s \\ &= \sum_{r=1}^k (a_{jr} + a_{rj})x_r \end{aligned}$$

This is precisely the j^{th} row of $A + A^T$ times x , or $e^j \cdot (A + A^T)x$. □

17.4 Proposition. With f as in (17.1),

$$\frac{\partial f_i}{\partial x_j}(x) = \begin{cases} e^i \cdot Ax - x \cdot Ax + x_i(a_{ii} - e^i \cdot (A + A^T)x) & \text{if } j = i, \\ x_i(a_{ij} - e^j \cdot (A + A^T)x) & \text{if } j \neq i. \end{cases}$$

Consequently, if x is an interior stationary point, so that $e^i \cdot Ax - x \cdot Ax = 0$, then

$$\frac{\partial f_i}{\partial x_j}(x) = x_i(a_{ij} - e^j \cdot (A + A^T)x). \quad (17.5)$$

Proof. This is immediate from Propositions 17.2 and 17.3. □

17.6 Corollary. Suppose the row sums of A are all equal. Then $x = \frac{1}{k}[1 \cdots 1]$ is an interior stationary point and a Nash equilibrium, and

$$\frac{\partial f_i}{\partial x_j}(x) = \frac{1}{k^2}(ka_{ij} - j^{\text{th}} \text{ column sum of } A + A^T). \quad (17.6)$$

Proof. If the row sums all equal s , then when $x = \frac{1}{k}[1 \cdots 1]$, we have $e^i \cdot Ax = \frac{s}{k}$ for all i , and so by Proposition 3.6, x is a NE. Also, in (17.5) $x_i = \frac{1}{k}$ and $e^j \cdot (A + A^T)x$ is the j^{th} row sum of $A + A^T$, which equals the j^{th} column sum because $A + A^T$ is symmetric. \square

17.7 Example. Returning to Example 16.4, we have

$$A = \begin{bmatrix} 3 & 40 & 0 \\ 0 & 8 & 35 \\ 15 & 0 & 28 \end{bmatrix}.$$

The row sums all equal 43 and $x = \frac{1}{3}[1 \ 1 \ 1]$ is the unique interior SP and a NE. We saw in 16.4 that it is not a NSS. It is asymptotically stable if the eigenvalues of $Jf(x)$ all have negative real parts. We have

$$A + A^T = \begin{bmatrix} 6 & 40 & 15 \\ 40 & 16 & 35 \\ 15 & 35 & 56 \end{bmatrix} \text{ and the column sums are } [61 \ 91 \ 106].$$

It follows from 17.6 that

$$Jf(x) = \frac{1}{9} \begin{bmatrix} -52 & 29 & -106 \\ -61 & -67 & -1 \\ -16 & -91 & -22 \end{bmatrix}.$$

One can check that the eigenvalues of this matrix are $-\frac{43}{3}$ and $-\frac{2}{3} \pm \frac{10\sqrt{5}}{3}i$. They all have negative real parts, and so by Theorem 11.3, x is an ASSP.

Thus, at last, we have completed the set of examples showing that no implications other than those entailed by Table 12.6 are true in general.

17.8 Example. The same holds for Example 16.8, where we have

$$A = \begin{bmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{bmatrix}.$$

The row sums are equal, $x = \frac{1}{3}[1 \ 1 \ 1]$ is the unique interior SP and a NE, and we showed that x is not a NSS. But

$$A + A^T = \begin{bmatrix} 0 & 3 & -5 \\ 3 & 0 & 8 \\ -5 & 8 & 0 \end{bmatrix}; \text{ the column sums are } [-2 \ 11 \ 3].$$

It follows from 17.6 that

$$Jf(x) = \frac{1}{9} \begin{bmatrix} 2 & 7 & -15 \\ -7 & -11 & 12 \\ -1 & -2 & -3 \end{bmatrix},$$

and the eigenvalues of this are $-\frac{2}{3}$ and $-\frac{1}{3} \pm \frac{\sqrt{2}}{3}i$; all have negative real parts. So again we have an ASSP that is not a NSS.

18. Bibliography

Gintis, Herbert: *Game Theory Evolving*. Princeton University Press, 2000.

Hofbauer, Josef and Karl Sigmund: *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

Kelly, Douglas G.: *Evolutionary Game Theory and the Prisoner's Dilemma: a survey*. Technical Report UNC-STOR2008-08, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill.

Osborne, Martin J.: *An Introduction to Game Theory*. Oxford University Press, 2004.

Osborne, Martin J. and Ariel Rubinstein: *A Course in Game Theory*. MIT Press, 1994.

Weibull, Jörgen: *Evolutionary Game Theory*. MIT Press, 1997.