

Article

## Selection-Mutation Dynamics of Signaling Games

Josef Hofbauer<sup>1</sup> and Simon M. Huttegger<sup>2,\*</sup>

<sup>1</sup> Department of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria; E-Mail: josef.hofbauer@univie.ac.at

<sup>2</sup> Department of Logic and Philosophy of Science, University of California at Irvine, 3151 Social Science Plaza A, Irvine, CA 92697, USA

\* Author to whom correspondence should be addressed; E-Mail: shuttegg@uci.edu; Tel.: +1-949-824-3220.

Academic Editor: Ulrich Berger

Received: 26 September 2014 / Accepted: 10 December 2014 / Published: 9 January 2015

---

**Abstract:** We study the structure of the rest points of signaling games and their dynamic behavior under selection-mutation dynamics by taking the case of three signals as our canonical example. Many rest points of the replicator dynamics of signaling games are not isolated and, therefore, not robust under perturbations. However, some of them attract open sets of initial conditions. We prove the existence of certain rest points of the selection-mutation dynamics close to Nash equilibria of the signaling game and show that all but the perturbed rest points close to strict Nash equilibria are dynamically unstable. This is an important result for the evolution of signaling behavior, since it shows that the second-order forces that are governed by mutation can increase the chances of successful signaling.

**Keywords:** selection-mutation dynamics; replicator dynamics; signaling games; structural stability

---

### 1. Introduction

In recent years, there has been considerable interest in the strategic aspects of signaling (Skyrms [1] provides an overview). The basic signaling model explored by game theorists is the sender-receiver game, which was introduced by the philosopher David Lewis [2], and has been re-discovered, in one

form or another, in economics and biology [3,4]. In a sender-receiver game, there is complete common interest between a sender, who is informed about the state of the world, and a receiver, who chooses an action that influences both the sender's and the receiver's payoff, which also depends on the state of the world. The receiver does not know this state, but can be informed about it by the sender's signal.

The crucial aspect of this game is that the signals do not have any pre-assigned meaning. Instead, senders and receivers may evolve toward a state of perfect communication [4,5]. It is well known that states of perfect communication are identical with the evolutionarily-stable strategies of the sender-receiver game, provided that they exist [6]. There are, however, many other kinds of Nash equilibria. These other equilibria are not isolated and, therefore, give rise to connected components of rest points for many classes of evolutionary dynamics. Since connected components of rest points are not robust, perturbations of the evolutionary dynamics lead to qualitative changes in the dynamic behavior and actually can destroy most of those rest points. This is a very general feature for games with a non-trivial extensive form where players move sequentially [7–10]. Along a component of rest points, selection is not operating, because there are no differences in payoffs. Evolutionary outcomes will thus often be determined by perturbing factors, such as mutation or drift.

In signaling games, many components of rest points are not asymptotically stable. However, subsets of (some of those) components attract open sets of initial conditions under the replicator dynamics [11,12]. The elements of such a component are not states of perfect communication; they only allow for partial communication. Hence, evolution can lead to states of partial communication, even in cases where the interests of sender and receiver completely coincide. Yet, because of the non-robustness of the components of rest points, the significance of this result depends on whether it persists under perturbations of the replicator dynamics.

Hofbauer and Huttegger [13] investigate the simple case of two signals and find that mutation rates can determine whether perturbed states of partial communication will be asymptotically stable. The purpose of this paper is to extend the analysis to the case of three signals. Such an extension is important, since signaling games with two signals exhibit certain features that distinguish them markedly from signaling games with more than two signals. In particular, the states of partial communication referred to above do not exist when there are only two signals. On the other hand, signaling games with three signals seem to exhibit the salient features of all signaling games with finitely many signals and can, thus, serve as a canonical example for this class of sender-receiver games.

We devote most of our attention to studying the stability of perturbed rest points. However, in the first step, it is necessary to investigate in detail the Nash equilibrium structure of the signaling game. We do this in Section 3 after briefly reviewing some general facts about sender-receiver games in Section 2. Our dynamical analysis consists of two parts. In Section 4, we study selection-mutation dynamics on the space of mixed strategies. Because of the high dimensionality of this space, we then consider selection-mutation dynamics on the space of behavior strategies in Section 5. This more tractable approach allows us to obtain specific information about the existence and stability of perturbed rest points.

## 2. The Model

Our description of the type of signaling game introduced by Lewis [2] follows closely the equivalent and mathematically more convenient setup of Trapa and Nowak [14]. A signaling game has two players, a sender and a receiver. A (pure) sender strategy can be represented as an  $n \times m$  matrix  $P$ . Each row of  $P$  has one entry being equal to 1; all other entries are equal to 0. There are  $n$  states of the world and  $m$  signals. Each row of  $P$  is associated with a state of the world and each column with a signal. If  $p_{ij} = 1$ , then the sender sends signal  $j$  when state  $i$  has occurred. Notice that  $n$  may be unequal to  $m$ , in which case the number of states is not the same as the number of signals.

Similarly, a (pure) receiver strategy will be represented by an  $m \times n$  matrix  $Q$ . A row of  $Q$  contains only zeros, except for one entry which is 1. Each row of  $Q$  is associated with a signal, while each column of  $Q$  corresponds to a state of the world (or, alternatively, one of  $n$  actions the receiver may perform). If  $q_{ji} = 1$ , then the receiver associates state  $i$  with signal  $j$  (or chooses action  $i$  upon receiving signal  $j$ ).

A mixed sender strategy (*i.e.*, a convex combination of pure strategies) results in a stochastic  $n \times m$  matrix  $P$ ; *i.e.*,  $p_{ij} \geq 0$  for all  $i = 1, \dots, n, j = 1, \dots, m$  and  $\sum_{j=1}^m p_{ij} = 1$  for all  $i = 1, \dots, n$ . Analogously, a mixed receiver strategy results in a stochastic  $m \times n$  matrix  $Q$ . A stochastic matrix determines a behavior strategy [9]. In the extensive form of a signaling game, each state of the world leads to a particular information set for the sender and each signal to one for the receiver. A stochastic matrix thus defines a behavioral strategy of a player, *i.e.*, a probability distribution over choices at an information set. This representation of behavior strategies will become important when we consider evolutionary dynamics on behavior strategies (*cf.* Section 5).

The payoff function in a Lewis signaling game is given by

$$\pi(P, Q) = \frac{1}{n} \sum_{i,j} p_{ij} q_{ji} = \frac{1}{n} \text{tr}(PQ) \tag{1}$$

where  $\pi(P, Q)$  is the payoff to each player if their joint strategy is  $(P, Q)$ . Hence, a Lewis signaling game is a partnership game, where both players get the same payoff for each outcome [15]. This symmetry yields quite special best response properties. For convenience, we state a result from Pawlowitsch [12] (BR denotes the best-response correspondence).

**Lemma 1.** *Let  $\bar{P}$  be a mixed  $n \times m$  sender strategy and  $\bar{Q}$  be a mixed  $m \times n$  receiver strategy. If  $Q \in \text{BR}(\bar{P})$ , then for all  $j$*

$$\sum_{i \in \text{argmax}_i(\bar{p}_{ij})} q_{ji} = 1 \quad \text{and} \quad q_{ji} = 0 \quad \text{if} \quad i \notin \text{argmax}_i(\bar{p}_{ij});$$

moreover:

$$\max_Q \text{tr}(\bar{P}Q) = \sum_j \max_i \bar{p}_{ij};$$

analogous relations hold for sender strategies  $P \in \text{BR}(\bar{Q})$ .

Lemma 1 implies that in order to find the maximum possible payoff given a certain strategy matrix, one just has to add its column maxima and normalize by  $\frac{1}{n}$ .

A signaling game with  $n$  states of the world and  $n$  signals has  $n!$  strict Nash equilibria, which have a very specific form; see e.g., [14].  $(P, Q)$  is a strict Nash equilibrium if and only if  $P$  is a permutation matrix and  $Q = P^\top$ . Strict Nash equilibria represent states of perfect communication and are often called signaling systems [2]. There are no strict Nash equilibria when the number of states is not equal to the number of signals [14]. There exist non-strict Nash equilibria besides signaling systems, though. A partial characterization of Nash equilibria is given in Trapa and Nowak [14]. If  $B$  is an  $n \times m$  matrix with no zero column, let us call  $B$  uniform if there exist real numbers  $b_1, \dots, b_m$ , such that for each  $j$ , the  $j$ -th column of  $B$  has 0 or  $b_j$  as entries.

**Lemma 2.** *Let  $P$  be an  $n \times m$  sender matrix and  $Q$  be an  $m \times n$  receiver matrix and assume that neither  $P$  nor  $Q$  contains a zero-column. Then,  $(P, Q)$  is a Nash equilibrium iff  $P$  and  $Q$  are uniform and  $p_{ij} = p_j$  iff  $q_{ji} = q_i$ .*

Hence,  $\text{supp}(P) = \text{supp}(Q^\top)$  (as subsets of  $\{1, \dots, n\} \times \{1, \dots, m\}$ ). Note that, at such a Nash equilibrium,

$$\pi(P, Q) = \frac{1}{n} \sum_{i,j} p_{ij} q_{ji} = \frac{1}{n} \sum_{(i,j) \in \text{supp}P} p_j q_i = \frac{1}{n} \sum_i q_i \sum_{j:(i,j) \in \text{supp}P} p_j = \frac{1}{n} \sum_i q_i$$

where the last equality follows from the fact that for each  $i$ ,  $\sum_{j:(i,j) \in \text{supp}P} p_j = 1$ . Analogously,

$$\pi(P, Q) = \frac{1}{n} \sum_j p_j,$$

and, thus,  $\sum_i p_i = \sum_j q_j$ .

Suppose that  $(P, Q)$  is a Nash equilibrium of the Trapa–Nowak class that is not strict. Then, it is clear that by continuous changes of the real numbers  $p_i, q_j$ ,  $i, j = 1, \dots, n$ , subject to the constraint that  $\sum_i p_i = \sum_j q_j$ , the resulting profile  $(P', Q')$  will still be a Nash equilibrium. Hence, non-strict Nash equilibria of the Trapa–Nowak class are part of the components of Nash equilibria (and, therefore, never isolated).

As one of the  $p_i$  or  $q_j$  approaches 0,  $P$  or  $Q$  will have a zero-column. In this case, the assumptions of Lemma 2 are not met, and its conclusions do not hold. Pawlowitsch [12] discusses these profiles at length. If signal  $j$  is never sent by the sender, then the receiver can choose arbitrary probability weights  $q_{ji}$  as long as the column maxima are preserved (this condition is necessary in order to preserve the best response correspondences of the players). Similarly, if the receiver never chooses state  $i$  in response to any signal, then the probability weights  $p_{ij}$  can take arbitrary values as long as column maxima are preserved.

As we already have noted above, signaling systems are the only strict Nash equilibria when  $n = m$ . In order to single out plausible ones among the huge number of other Nash equilibria, we mitigate the concept of strict Nash equilibrium in two ways. It is clear that strict Nash equilibria in partnership games are strict local maximizers of the payoff function  $\pi$ . Let us call a strategy pair  $(P, Q)$  stable iff the common payoff function is locally maximal at  $(P, Q)$ , i.e., for  $(P', Q')$  close to  $(P, Q)$ :  $\pi(P', Q') \leq \pi(P, Q)$ . Such a strategy will be Lyapunov stable under many evolutionary dynamics, including the replicator dynamics.

It is well known that there is a one-to-one correspondence between strict Nash equilibria of asymmetric games and evolutionarily-stable strategies of their symmetrization (Selten's theorem, see [9]). Similarly, it follows that stable strategy pairs correspond to neutrally stable strategies of the symmetrization (see [12] for definitions of neutral and evolutionary stability in signaling games). Therefore, we will refer to neutrally stable strategy pairs  $(P, Q)$  also in the context of asymmetric games with the understanding that we really mean stable strategy pairs.

Pawlowitsch [12] established the following important characterization of neutrally-stable strategies:

**Lemma 3.** *Let  $P$  be an  $n \times m$  sender matrix and  $Q$  be an  $m \times n$  receiver matrix. Then,  $(P, Q)$  is neutrally stable iff  $P$  or  $Q$  has no zero columns, and neither  $P$  nor  $Q$  has a column with multiple maximal elements that are elements of the open interval  $(0, 1)$ .*

Neutrally-stable strategies again give rise to components of Nash equilibria. Their significance derives from their role in the replicator dynamics of signaling games. For instance, neutrally-stable strategies are important for  $n \neq m$  when there are no strict equilibria. Then, the Pareto optimal component is asymptotically stable [6]. However, as we will see below, they are also of crucial importance in the case  $m = n$ .

Another weakening of strict Nash equilibria that will turn out to be crucial for analyzing signaling games are quasi-strict Nash equilibria. In a two-player game, a Nash equilibrium  $(x, y)$  is called quasi-strict if  $\text{BR}(y)$  is contained in the simplex spanned by  $\text{supp}(x)$  and  $\text{BR}(x)$  is contained in the simplex spanned by  $\text{supp}(y)$ . This means that there are no alternative best responses that are not contained in the faces spanned by  $x$  and  $y$ .

### 3. The Structure of Nash Components

In this section, we will try to get some more specific insights into the Nash equilibrium structure of signaling games. Knowledge of the geometry of Nash equilibria will prove to be particularly important once we start looking at the selection-mutation dynamics of signaling games.

We start with a useful general observation:

**Lemma 4.** *In any partnership game, there are only finitely many Nash equilibrium payoffs.*

**Proof.** Recall that if  $(\bar{p}, \bar{q})$  is a Nash equilibrium and  $p$  is a pure or mixed strategy for Player 1, whose support is contained in that of  $\bar{p}$ , then  $\pi(\bar{p}, \bar{q}) = \pi(p, \bar{q})$ . Suppose now that  $(p, q)$  and  $(p', q')$  are Nash equilibria with the same support. Then,  $\pi(p, q) = \pi(p', q) = \pi(p', q')$ . The second equality follows since Player 2 has the same payoff function. Since there are only finitely many possible supports, there are only finitely many equilibrium payoffs.  $\square$

Therefore, even though the set of Nash equilibria may be infinite, the set of Nash equilibrium payoffs of a (finite strategy) partnership game is always finite. The reason is that along any continuum of Nash equilibria, the payoff function is constant.

Let us now determine the payoff levels of a signaling game with  $n = m = 3$  at which Nash equilibria can be found. There are 6 strict Nash equilibria on the payoff level 1, given by a permutation matrix  $P$  and  $Q = P^T$  (these are the signaling systems). An example of this class is

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, Q = P^T. \tag{2}$$

The lowest payoff level at which Nash equilibria can be found is  $\frac{1}{3}$ . This payoff level includes Nash equilibria where the sender always uses the same signal or the receiver always chooses the same response. It also includes completely mixed equilibria, such as the barycenter. The  $\frac{1}{3}$  component is given by

$$P = \begin{pmatrix} \lambda & \mu & 1 - \lambda - \mu \\ \lambda & \mu & 1 - \lambda - \mu \\ \lambda & \mu & 1 - \lambda - \mu \end{pmatrix}, Q = \begin{pmatrix} \nu & \eta & 1 - \nu - \eta \\ \nu & \eta & 1 - \nu - \eta \\ \nu & \eta & 1 - \nu - \eta \end{pmatrix}, \tag{3}$$

where  $0 \leq \lambda, \mu, \nu, \eta$  and  $\lambda + \mu \leq 1, \nu + \eta \leq 1$ . It follows from Lemma 2 that these are Nash equilibria. The  $\frac{1}{3}$  component is convex and invariant under permutations of strategies.

The Trapa–Nowak characterization of Nash equilibria assumes that strategy matrices contain no zero columns. However, there are Nash equilibria that do not satisfy this assumption. In the case of three signals, these additional equilibria can be found on the  $\frac{1}{3}$  and on the  $\frac{2}{3}$  payoff level. Strategy matrices on the former payoff level are of the form

$$N = \begin{pmatrix} \lambda & 1 - \lambda & 0 \\ \lambda & 1 - \lambda & 0 \\ \lambda & 1 - \lambda & 0 \end{pmatrix},$$

with  $0 \leq \lambda \leq 1$ , where  $N$  can be a sender or a receiver matrix. These equilibria are all on the boundary of the component (3); *i.e.*, they are limits of the strategy matrices (3).

At the payoff level  $\frac{1}{2}$ , there are Nash equilibria that are mixtures of signaling systems. An example is given by

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, Q = P^T. \tag{4}$$

Notice that Equation (4) is isolated in the space of behavioral strategy profiles, but in the space of mixed strategies, it corresponds to a continuum of Nash equilibria. Each entry of  $P$  or  $Q$  can be realized by convex combinations of different pure strategies (in this case, eight pure strategies). Including permutations, the number of such mixed equilibria is 6.

The last, and most interesting, level is  $\frac{2}{3}$ . On the  $\frac{2}{3}$  level, there are equilibria, such as

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu & 1 - \mu \\ 0 & \mu & 1 - \mu \end{pmatrix}, \tag{5}$$

with  $0 \leq \lambda, \mu \leq 1$ . None of these Nash equilibria is neutrally stable (see Lemma 3). Note that the two-dimensional set of all these equilibria forms a Nash set [16,17]: every pair of such equilibria

$(P, Q), (P', Q')$  is interchangeable, *i.e.*,  $(P, Q')$  and  $(P', Q)$  are equilibria in this set, as well; *cf.* [18]. After renumbering events and signals, we obtain nine Nash sets like (5). The interior equilibria  $(P, Q)$  of these Nash sets (*i.e.*, for  $0 < \lambda, \mu < 1$ ) are quasi-strict.

A typical neutrally stable strategy is given by

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, Q = \begin{pmatrix} \mu & 1 - \mu & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6}$$

where  $0 < \lambda < 1$  or  $0 < \mu < 1$  (*cf.* Lemma 3 and Example 1 [12]). All of these equilibria are also quasi-strict. When both  $\lambda$  and  $\mu$  approach 0 or 1, these strategy pairs are still Nash equilibria, but no longer neutrally stable. The points on an edge (e.g.,  $0 < \lambda < 1, \mu = 0$ ) are neutrally stable, but not quasi-strict. After permutations, we again obtain 9 two-dimensional Nash sets of this form.

Two more classes of Nash equilibria can be found on the  $\frac{2}{3}$  payoff level. One class is given by

$$\begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 - \lambda & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \mu & 0 & 1 - \mu \end{pmatrix}. \tag{7}$$

with  $0 \leq \lambda, \mu \leq 1$ . Strategies of type (7) are not neutrally stable, because both  $P$  and  $Q$  have zero columns. Moreover, they are not quasi-strict (as shown in Proposition 8). There are 18 (two-dimensional) Nash sets of type (7).

There is another class of neutrally-stable strategies that has, to our knowledge, not been noted in the literature before. A typical example is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} \lambda & 1 - \lambda & 0 \\ 0 & 0 & 1 \\ \mu & \nu & 1 - \mu - \nu \end{pmatrix}. \tag{8}$$

where  $0 \leq \mu \leq \lambda$  and  $0 \leq \nu \leq 1 - \lambda$ . Notice that interior points in Nash sets such as (8) are neutrally stable by Lemma 3. There are 18 (plus 18 more after interchanging the two players) three-dimensional Nash sets of type (8).

Note that equilibria of the last two types are not limit points of Trapa–Nowak equilibria. Hence, not all Nash equilibria with zero columns are limit points of Nash equilibria whose strategy matrices contain no zero columns.

We provide an overview of the different types of Nash equilibria and their properties given in Figure 1.

As we shall see in the next two sections, neutrally-stable strategies are of major importance from the point of view of evolutionary dynamics. Signaling systems are of course also neutrally stable by virtue of being strict Nash equilibria. As we have just seen, there are two other types of neutrally stable strategies besides signaling systems, namely strategy profiles (6) and (8).

| Payoff Level  | Nash Sets | Number | Morse Index | Poincaré Index |
|---------------|-----------|--------|-------------|----------------|
| 1             | (2)       | 6      | 0           | +6             |
| $\frac{2}{3}$ | (5)       | 9      | 1           | −9             |
|               | (6)       | 9      | 2           | +9             |
|               | (7)       | 18     | −           | 0              |
| $\frac{1}{2}$ | (8)       | 36     | −           | 0              |
|               | (4)       | 6      | 3           | −6             |
| $\frac{1}{3}$ | (3)       | 1      | 4           | +1             |

**Figure 1.** Summary for the Nash equilibria of a signaling game with  $n = 3$ . The Morse index and the Poincaré index refer to the rest points of the selection-mutation dynamics in behavior strategies (14) for the case  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ . The Morse index is the number of positive eigenvalues of the Jacobian. At a maximizer of the potential function, the Morse index is 0. The Poincaré index is  $-1$  if the number of positive eigenvalues is odd and  $+1$  if it is even. The sum of the Poincaré indices is 1, which is equal to the Euler characteristic of the state space; cf. [15], Ch. 13.2.

There is an important difference between Nash sets, such as (5), and Nash sets as given by (6)–(8). The two-dimensional Nash set (5) is a subset of a larger polyhedron spanned by four sender strategies and four receiver strategies. The other Nash sets on the  $\frac{2}{3}$  payoff level are maximal in the sense that they contain no mixtures of pure strategies that are not included in the Nash set. Note that the  $\frac{2}{3}$  component is not convex (that it is connected will be shown in Proposition 7 below).

In the next proposition, we show that there are no Nash equilibria besides the ones listed above.

**Proposition 5.** *In a signaling game with  $n = m = 3$ ,  $(P, Q)$  is a Nash equilibrium iff  $(P, Q)$  belongs to one of the Nash sets of type (2)–(8).*

**Proof.** Let  $(P, Q)$  belong to the Trapa–Nowak class of Nash equilibria. Suppose first that  $0 < p_{ij}, q_{ij} < 1$  for all  $i, j$ . Then, Lemma 2 implies that  $(P, Q)$  belongs to class (3).

Suppose now that one entry in a particular row is zero, e.g.,  $p_{11} = 0$  (this can be achieved by renumbering the events and signals or interchanging the players) and that no entry of  $P$  equals 1. Set  $p_{12} = \lambda$  and  $p_{13} = 1 - \lambda$ ,  $0 < \lambda < 1$ . If  $p_{21} = 0$ , then  $p_{31} > 0$ . Since  $p_{31}$  is the unique column maximum, this implies that  $q_{13} = 1$  (Lemma 1), and from Lemma 2, it follows that  $p_{31} = 1$ , a contradiction. Hence,  $p_{21} > 0$ . If  $p_{22} = \lambda$ , then  $p_{21} = 1 - \lambda$ ; and if  $p_{23} = 1 - \lambda$ , then  $p_{21} = \lambda$ . Assume the latter without loss of generality, and suppose  $p_{33} > 0$ . Then,  $p_{33} = 1 - \lambda$  by Lemma 2. Thus, either  $p_{31} = 0$  or  $p_{32} = 0$ , and, again, by Lemma 1,  $q_{12} = 1$  or  $q_{21} = 1$ , which implies that



$\lambda = 1$ . Thus,  $p_{33} = 0$ , and we get  $\lambda = 1 - \lambda$ . Therefore, allowing one entry to be zero results in a Nash equilibrium belonging to class (4).

A similar argument applies when we assume that one entry in one row is 1. If  $p_{11} = 1$ , then  $p_{12} = p_{13} = 0$ , and by Lemma 2,  $p_{21} = p_{31} = 0$ . If  $p_{22} = 0$ , then  $p_{23} = 1$ , which contradicts the assumption that  $P$  has a single entry 1. Therefore,  $p_{22} = \lambda > 0$ , and hence,  $p_{23} = 1 - \lambda$  for  $0 < \lambda < 1$ . Lemma 2 implies that  $(P, Q)$  has the form (5).

The argument in the last paragraph shows that assuming that one entry of  $P$  equals 1 implies that the other two rows contain at least one zero. Nash equilibria of class (6) result from the assumption that two entries in two rows equal 1. Indeed, if there are two entries in two rows that equal 1, then they must be elements of the same column by Lemma 2; e.g.,  $p_{11} = p_{21} = 1$  (we are again assuming that  $P$  is not a permutation matrix). If  $p_{31} > 0$ , then  $p_{31} = 1$ , a contradiction. Thus,  $p_{31} = 0$ ,  $p_{32} = \lambda$  and  $p_{33} = 1 - \lambda$ ,  $0 < \lambda < 1$ .

If we allow for three 1 entries (one in each row), then  $P$  is a permutation matrix, and by Lemma 2,  $Q = P^T$ . Hence,  $(P, Q)$  is a signaling system. This yields the last class of Trapa–Nowak Nash equilibria.

Other Nash equilibria have either two zero columns or one zero column. If there are two zero columns in, say,  $P$ , then  $(P, Q)$  is a Nash equilibrium if, and only if,  $Q$  is as in (3). The resulting Nash equilibria are part of the boundary of the set defined by (3).

If, however,  $P$  has only one zero column, then it looks like

$$\begin{pmatrix} \lambda & 1 - \lambda & 0 \\ \mu & 1 - \mu & 0 \\ \nu & 1 - \nu & 0 \end{pmatrix}.$$

The case  $\lambda = \mu = \nu$  leads again to Nash equilibria belonging to the boundary of the component defined by (3). Suppose  $\lambda > \mu > \nu$ . Then, by Lemma 1, the maximum payoff against such a matrix is  $1 + \lambda - \nu$ . Hence, such a matrix is part of a Nash equilibrium only if  $\lambda = 1$  and  $\nu = 0$  ( $\mu$  is arbitrary).  $P$  is in Nash equilibrium with a strategy  $Q$ , if and only if  $(P, Q)$  belongs to the class of Nash equilibria given by (7).

If  $\lambda = \mu > \nu$ , then at a Nash equilibrium, we must have  $\lambda = \mu = 1$  and  $\nu = 0$ . The best response to this strategy is constrained to look like the right-hand part of (8). If  $\lambda > \mu = \nu$ , then a similar argument leads again to Nash equilibria of type (8).

This shows that if  $(P, Q)$  is a Nash equilibrium, then it must belong to one of the classes (2)–(8). Applying Lemma 2 together with simple calculations shows that a joint strategy belonging to one of these classes is a Nash equilibrium. □

As an immediate consequence, we get information about the payoffs obtained in equilibrium. In a signaling game with three signals, all Nash equilibria are found on a specific payoff level. At a signaling system (2), players earn a payoff of 1; at (5)–(8), players get  $\frac{2}{3}$ ; at (4), players get  $\frac{1}{2}$ ; and at (3), players get  $\frac{1}{3}$ .

The following proposition summarizes the number of each kind of Nash set, which can be calculated by using simple combinatorial arguments.

**Proposition 6.** *In the signaling game with  $n = m = 3$ , there are 6 signaling systems and 6 Nash equilibrium components of type (4), 9 Nash sets of type (5), 9 Nash sets of type (6), 18 Nash sets of type (7), 36 Nash sets of type (8) and, finally, one component of type (3).*

**Proof.** It is clear that there are  $3! = 6$  signaling systems and one Nash set of type (3). Nash sets of type (4) are determined by the location of the three 0 entries, which cannot be in the same column or the same row. Thus, there are again  $3! = 6$  such Nash sets. Nash sets of type (5) and (6) are determined by the location of a 1 entry. In both cases, the 1 entry determines which entries must be equal to 0, respectively 1. There clearly are nine possible Nash sets of each of these types. Nash sets of type (7) are given by the location of the zero column, of which there are 3 possibilities, and the location of the two 1 entries, which can vary in 6 possible ways. We thus have  $3 \times 6 = 18$  sets of this type. For type (8), we look again at the 1 entries. There are three 1 entries, two of which must be in the same column. For each choice of location for the third entry, there are two possible locations for the two in the same column. Hence, there are 18 different configurations. We have to double this number, because the same holds for the other player.  $\square$

As we have already remarked, the  $\frac{2}{3}$  component is of particular importance, since it contains all neutrally stable strategies. We can, in fact, show that it is a component. Call two Nash equilibria  $x$  and  $y$  path connected if there is a continuous path  $\phi$  of strategy profiles connecting  $x$  and  $y$ , such that every strategy profile on  $\phi$  is itself a Nash equilibrium. The following proposition establishes that each pair of Nash equilibria which yields a payoff of  $\frac{2}{3}$  is path connected (indeed, by a piecewise linear path).

**Proposition 7.** *The Nash equilibria on the  $\frac{2}{3}$  payoff level are path connected. Thus, the corresponding 72 Nash sets constitute a single component of Nash equilibria.*

**Proof.** First, notice that each Nash set of type (8) is path connected to a Nash set of type (6), since the signaler part of the former is a pure strategy. Likewise, it is clear that Nash sets of type (7) are connected to Nash sets of type (5). Thus, it will suffice to show that: (i) Nash sets of type (5) are connected; (ii) Nash sets of type (6) are connected; and (iii) one type of (5) and one of type (6) are connected.

Nash sets of type (5) can be identified by the column of the sender matrix  $j$ , which has  $p_{ij} = 1$  for some  $i$  and  $p_{kj} = 0$  for  $k \neq i$ . This follows from the special properties of Nash equilibria of the Trapa–Nowak class (see Lemma 2). For the same reasons, Nash sets of type (6) can be identified by locating the column of the sender matrix  $j$ , which has  $p_{ij} = 0$  for some  $i$  and  $p_{kj} = 1$  for  $k \neq i$ . For Nash sets of type (5), we will show that if  $p_{11} = 1, p_{21} = 0, p_{31} = 0$ ; then, the 1 entry can be moved horizontally to  $p_{12}, p_{13}$  and vertically to  $p_{21}$  and  $p_{31}$ , such that no changes lead to a strategy profile, which is not a Nash equilibrium. Horizontal moves are possible, because we can get from (5) to certain neutral Nash sets (6), which also shows that claim (iii) holds. Vertical moves are possible because of Nash sets of type (7). By symmetry, connectedness is established for all other sender strategies and the corresponding receiver strategies, as well. Similar arguments will establish claim (ii).

We start with the strategy matrices

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu & 1 - \mu \\ 0 & \mu & 1 - \mu \end{pmatrix}. \tag{9}$$

Suppose  $\lambda = 1, \mu = 1$  and call the corresponding strategy pair  $(P', Q')$ .  $P'$  is in Nash equilibrium with all receiver matrices of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \nu & 1 - \nu & 0 \end{pmatrix} \tag{10}$$

(cf. (7)). If  $Q''$  is the receiver strategy with  $\nu = 1$ , then  $(P', Q')$  and  $(P', Q'')$  are connected by a line of Nash equilibria. However,  $(P', Q'')$  is part of the neutrally-stable Nash set

$$\begin{pmatrix} \lambda & 0 & 1 - \lambda \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mu & 1 - \mu \\ 1 & 0 & 0 \end{pmatrix}, \tag{11}$$

where  $0 \leq \lambda, \mu \leq 1$ . This establishes claim (iii). Moreover, from this neutrally-stable Nash set, we can move to a Nash set of type (5), where  $p_{13} = 1$ . To this end, set  $\lambda = 0$  and  $\mu = 0$ , and apply arguments analogous to the ones above (using Nash sets of type (7)) to move along a path of Nash equilibria to the Nash set

$$\begin{pmatrix} 0 & 0 & 1 \\ \lambda & 1 - \lambda & 0 \\ \lambda & 1 - \lambda & 0 \end{pmatrix}, \begin{pmatrix} 0 & \mu & 1 - \mu \\ 0 & \mu & 1 - \mu \\ 1 & 0 & 0 \end{pmatrix}.$$

If we return to the Nash set given by (9), we can apply the same kinds of arguments to move from the vertex where  $\lambda = 0, \mu = 0$  to the neutrally stable Nash set given by

$$\begin{pmatrix} \lambda & 1 - \lambda & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \mu & 1 - \mu \end{pmatrix}$$

and from there to the Nash set

$$\begin{pmatrix} 0 & 1 & 0 \\ \lambda & 0 & 1 - \lambda \\ \lambda & 0 & 1 - \lambda \end{pmatrix}, \begin{pmatrix} 0 & \mu & 1 - \mu \\ 1 & 0 & 0 \\ 0 & \mu & 1 - \mu \end{pmatrix}$$

which has  $p_{21} = 1$ . This shows that the 1 entry can be moved horizontally.

From the corner  $(P', Q'')$ , we can move on a path of Nash profiles to another Nash set of type (5), which is given by

$$\begin{pmatrix} \lambda & 0 & 1 - \lambda \\ 0 & 1 & 0 \\ \lambda & 0 & 1 - \lambda \end{pmatrix}, \begin{pmatrix} \mu & 0 & 1 - \mu \\ 0 & 1 & 0 \\ \mu & 0 & 1 - \mu \end{pmatrix}.$$

To do this, we just have to use the sender matrix (10). This shows that we can get from a 1 entry in the first row to a 1 entry in the second row. Hence, it is clear that we can also get to  $p_{21} = 1$  by using horizontal movements. Starting at the corner

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

one can show by using a similar kind of reasoning that moving to a Nash set of type (5) with  $p_{31} = 1$  is possible without ever passing through a strategy profile that is not a Nash equilibrium.

Concerning (ii), we have already shown in the first part of the proof that one can move the column with two 1 entries from the second to the third column by passing through Nash equilibria of types (5) and (7). Similarly from the corner of the profile (11) with  $\lambda = 0, \mu = 0$ , one can move to the profile

$$\begin{pmatrix} 0 & \lambda & 1 - \lambda \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \mu & 1 - \mu \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

This shows that horizontal moves of the columns defining neutrally-stable sets are feasible. Vertical moves are feasible because of the profiles of type (7). For instance, one can get from (11) to

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, \begin{pmatrix} \mu & 1 - \mu & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

by moving through

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 - \lambda & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \mu & 0 & 1 - \mu \end{pmatrix}.$$

All other claims are established in a similar manner.  $\square$

In the next two sections, we will make use of the concept of quasi-strictness. The following two results relate quasi-strictness to the Trapa–Nowak class of Nash equilibria. Concerning our use of the term ‘mixed strategy representative’, note that a mixed strategy is a convex combination of pure strategies. In signaling games, pure strategies are zero-one matrices. A stochastic matrix  $P$  generally corresponds to more than one mixed strategy ( $P$  determines a behavior strategy). Thus, if  $(P, Q)$  is a pair of sender and receiver strategies, then a mixed strategy representative of  $(P, Q)$  is a pair of mixed strategies corresponding to  $P$  and  $Q$ , respectively.

**Proposition 8.** *If  $(P, Q)$  is a Nash equilibrium where both  $P$  and  $Q$  contain at least one zero-column, then no mixed strategy representative of  $(P, Q)$  is quasi-strict.*

**Proof.** Suppose without loss of generality that  $p_{i1} = 0$  and  $q_{j1} = 0$  for  $0 \leq i, j \leq 1$ . Suppose  $Q'$  is like  $Q$ , except that  $q_{11} = 1$ . Then, by Lemma 1,  $Q'$  is an alternative best response to  $P$ . However, since  $q_{11} = 1$ , there exists a pure strategy in  $\text{supp}(Q')$  that is not in  $\text{supp}(Q)$ . This implies that no mixed strategy representative of  $(P, Q)$  is quasi-strict.  $\square$

For  $n = m = 3$ , this implies that Nash equilibria of class (7) are not quasi-strict. Boundary equilibria of (3) with a zero column for the sender and the receiver matrix are also not quasi-strict. However, consider  $(P, Q)$  as in (3), and set  $\lambda = 0$ . Then,  $(P, Q)$  is not quasi-strict, since the sender strategy that sets  $\lambda = \mu = 0$  is an alternative best reply to  $Q$ , which is not in the support of any mixed-strategy representative of  $P$ . This shows that the converse of Proposition 8 does not hold in general.

**Proposition 9.** *If  $(P, Q)$  is a Nash equilibrium of the Trapa–Nowak class, then any mixed strategy representative of  $(P, Q)$  is quasi-strict.*

**Proof.** Since  $(P, Q)$  is a Trapa–Nowak Nash equilibrium, Lemma 2 implies that  $\text{supp}(P) = \text{supp}(Q^\top)$  and  $\text{supp}(P^\top) = \text{supp}(Q)$ . Thus, a pure strategy not in  $\text{supp}(P)$  or  $\text{supp}(Q)$  cannot be an alternative best reply to  $Q$  or to  $P$ .  $\square$

For the case  $n = m = 3$ , the only class of Nash equilibria where the two propositions above do not apply are given by (8) whenever  $Q$  has no zero column. To see when such Nash equilibria are quasi-strict, notice that  $P$  is the unique best reply to  $Q$  if  $Q$  has no zero column and  $0 < \mu < \lambda, 0 < \nu < 1 - \lambda$ . Moreover, by Lemma 1, a pure strategy best reply to  $P$  has a 1 entry only at the possibly positive entries of  $Q$ , given these constraints on  $Q$ . Note also that the equilibria in the Nash set (8) (except the pure ones) are neutrally stable according to Lemma 3. However, as noted above, if  $\mu = \nu = 0$ , then  $(P, Q)$  is neutrally stable, but not quasi-strict.

#### 4. Dynamics for Mixed Strategies

Several of the results obtained in the previous section are useful when studying the evolutionary dynamics of signaling games. For infinite population models, signaling games were studied most thoroughly in the context of the replicator dynamics [11,12]. The main problem for understanding the robust features of the replicator dynamics arises from the fact that in signaling games there are linear manifolds of rest points. The replicator dynamics of signaling games is thus not structurally stable, and perturbations of the dynamics will typically result in qualitatively different dynamical behavior.

For this reason, Hofbauer and Huttegger [13] investigated a natural perturbation of the replicator dynamics—the selection-mutation dynamics of Hofbauer [19]—for signaling games with two signals. The selection-mutation dynamics for two populations is given by

$$\dot{x}_i = x_i((Ay)_i - \mathbf{x} \cdot Ay) + \varepsilon(1 - nx_i) \tag{12a}$$

$$\dot{y}_j = y_j((B\mathbf{x})_j - \mathbf{y} \cdot B\mathbf{x}) + \delta(1 - my_j), \tag{12b}$$

where  $\mathbf{x} \in S_n$  and  $\mathbf{y} \in S_m$  describe the state of each population,  $(A, B)$  are the payoff matrices and  $\varepsilon, \delta$  are small, uniform mutation parameters. If  $\varepsilon = \delta = 0$ , the selection-mutation dynamics coincides with the two-population replicator dynamics. We assume that  $\varepsilon$  and  $\delta$  are of the same order as they go to zero; *i.e.*, there exists a fixed  $\rho > 0$ , such that  $\delta = \rho\varepsilon$  as  $\varepsilon \rightarrow 0$ .

In the case of signaling games, it is shown in Hofbauer and Huttegger [13] that the selection-mutation dynamics is a gradient system with respect to the Shahshahani metric. The corresponding potential function is given by

$$V(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot A\mathbf{y} + \varepsilon \sum_i \log x_i + \delta \sum_j \log y_j. \quad (13)$$

The first term on the right-hand side is the average payoff in the sender population, which equals the average payoff in the receiver population since  $A = B^\top$ . The logarithmic terms imply that rest points must be in the interior of the state space if  $\varepsilon, \delta > 0$ . Because of the potential function, there can be no limit cycle in the selection-mutation dynamics of signaling games.

There are some important relations between the selection-mutation dynamics and the replicator dynamics. The first one implies that the Nash equilibria of the underlying game (which are also rest points of the replicator dynamics) give some information about the location of rest points of the selection-mutation dynamics (for a proof, see 13).

**Proposition 10.** *If  $(\mathbf{x}, \mathbf{y})$  is a rest point of the two-population replicator dynamics, which is not a Nash equilibrium, then there exists no rest point for the selection-mutation dynamics (12) close to  $(\mathbf{x}, \mathbf{y})$  for sufficiently small  $\varepsilon, \delta$ .*

Another result asserts the existence of rest points close to the strict equilibria of signaling games. A Nash equilibrium  $(\mathbf{x}, \mathbf{y})$  is regular if the Jacobian matrix  $J$  of (12) with  $\varepsilon, \delta = 0$  has no zero eigenvalues, and it is hyperbolic if  $J$  has no eigenvalues with zero real part. Note that for partnership games, where  $B = A^\top$ , the replicator dynamics is a gradient, and hence, every regular equilibrium is hyperbolic. For such equilibria, a converse result to Proposition 10 holds.

**Proposition 11.** *Let  $(\mathbf{x}, \mathbf{y})$  be a regular Nash equilibrium. Then, there exist unique rest point of the selection-mutation dynamics (12) close to  $(\mathbf{x}, \mathbf{y})$  for sufficiently small  $\varepsilon, \delta$ . Moreover, if  $(\mathbf{x}, \mathbf{y})$  is hyperbolic, then the perturbed rest point has the same stability properties under the selection-mutation dynamics as  $(\mathbf{x}, \mathbf{y})$  has for the replicator dynamics.*

For proofs of Proposition 11, see Bürger and Hofbauer [20], Ritzberger [21], and, in the more specific context of signaling games, Hofbauer and Huttegger [13]. For signaling games, Proposition 11 implies that there exists a unique family of asymptotically-stable rest points of the selection-mutation dynamics close to each signaling system whenever  $\varepsilon, \delta$  are sufficiently small.

Regular Nash equilibria are isolated. As we have seen in the  $n = m = 3$  case, there are no isolated Nash equilibria other than signaling systems. This means that Proposition 11 only applies to signaling systems. In particular, Propositions 10 and 11 do not determine the dynamical properties of the selection-mutation dynamics (12) close to neutrally-stable Nash sets, such as (6). Neutrally-stable Nash sets are important in the replicator dynamics. As was shown in Pawlowitsch [12], neutrally-stable Nash sets attract an open set of initial conditions under the replicator dynamics without being asymptotically stable sets (neutrally-stable Nash sets cannot be asymptotically stable since the vertices of the set are unstable). No general results apply in this case. It is therefore unclear what will happen close to these sets under the selection-mutation dynamics.

The rest of the present paper is devoted to the analysis of the selection-mutation dynamics for signaling games. Our analysis will rely heavily on the results obtained in the previous section. Some of the questions we are interested in will be investigated in the context of the selection-mutation dynamics as given by (12), *i.e.*, in the context of mixed strategy spaces. This framework does not seem to be suitable for other problems, since the dimension of the state space is too large. For  $n = 3$  signals, the state space of the selection-mutation dynamics (12) is a 52-dimensional polyhedron (there are 27 strategies for each player). It is thus very difficult to investigate the behavior of the selection-mutation dynamics close to Nash equilibria other than signaling systems. Therefore, in the next section, we will consider behavioral strategies. In particular, we will study the dynamic features in the neighborhood of neutrally-stable Nash sets in the context of selection-mutation dynamics for behavioral strategies.

Consider, for example, the Nash set given in (5). For  $0 < \lambda, \mu < 1$ , these equilibria belong to the Trapa–Nowak class of Nash equilibria. Therefore, by Proposition 9, Nash equilibria given by (5) are quasi-strict as long as  $0 < \lambda, \mu < 1$ . If we view these equilibria in terms of mixed strategies, they are located on the boundary region  $M$  spanned by the product of the four sender strategies

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

with the four receiver strategies represented by the same matrices. Consider an equilibrium  $(\mathbf{p}, \mathbf{q})$  in the relative interior of  $M$ . Since  $(\mathbf{p}, \mathbf{q})$  is quasi-strict, all transversal eigenvalues in (12) with  $\varepsilon = \delta = 0$  are negative [9]. This implies that there exist orbits in the interior of  $S_{27} \times S_{27}$  that approach  $(\mathbf{p}, \mathbf{q})$  under the replicator dynamics. Hence, we may restrict our analysis of the stability properties of  $(\mathbf{p}, \mathbf{q})$  to the replicator dynamics on  $M$ . The set of Nash equilibria in  $M$  consists of two signaling systems and a linear manifold of Nash equilibria  $N$  corresponding to the mixed-strategy representatives of (5). It is straightforward to show that the linearization of the replicator dynamics at  $(\mathbf{p}, \mathbf{q})$  always has one eigenvalue with a positive real part. Hence, each  $(\mathbf{p}, \mathbf{q})$  is linearly unstable.

Let us now consider the selection-mutation dynamics (12) close to  $N$ . Suppose there exist rest points of (12) close to points in the relative interior of  $N$ . Since the entries of the Jacobian matrix of (12) are continuous in  $\varepsilon, \delta$ , each rest point close to a point in the relative interior of  $N$  will have an eigenvalue with positive real part for sufficiently small  $\varepsilon, \delta$ . Such a perturbed rest point will thus be linearly unstable.

The same kind of reasoning applies to Nash sets of type (4). In this case, there are eight pure sender strategies and eight pure receiver strategies in the support of  $(P, Q)$ . If  $(\mathbf{p}, \mathbf{q})$  is an interior equilibrium of  $(P, Q)$  in the boundary region spanned by these strategies, then  $(\mathbf{p}, \mathbf{q})$  is linearly unstable. Thus, any rest point of (12) close to  $(\mathbf{p}, \mathbf{q})$  will also be linearly unstable for sufficiently small values of the mutation parameters.

This kind of reasoning can be generalized to signaling games with  $n \geq 3$  signals. The key fact is that there are always at least two signaling systems in the support of the mixed strategy representatives of the corresponding Nash sets.

**Proposition 12.** *Let  $(P, Q)$  be a Nash equilibrium of the Trapa–Nowak class and  $(\mathbf{p}, \mathbf{q})$  a mixed strategy representative of  $(P, Q)$ , which has a signaling system in its support. Then,  $(\mathbf{p}, \mathbf{q})$  is linearly unstable*

if it is not a signaling system. Consequently, a rest point of (12) close to  $(\mathbf{p}, \mathbf{q})$  is linearly unstable for sufficiently small  $\varepsilon, \delta$ .

**Proof.** Let  $J$  be the Jacobian matrix of the replicator dynamics evaluated at the point  $(\mathbf{p}, \mathbf{q})$ . Since signaling games are partnership games,  $J$  is a self-adjoint linear operator relative to the Shahshahani inner product, *i.e.*,  $\langle (\xi, \eta), J(\zeta, \tau) \rangle_{(\mathbf{p}, \mathbf{q})} = \langle J(\xi, \eta), (\zeta, \tau) \rangle_{(\mathbf{p}, \mathbf{q})}$  for all  $(\xi, \eta), (\zeta, \tau)$  in the tangent space of the population state space; *cf.* [15], p. 259. Hence,  $J$  has at least one positive eigenvalue if  $J$  is not negative semi-definite. If we set  $\xi = p' - p$  and  $\eta = q' - q$  where

$$p' = \alpha \mathbf{p}^* + (1 - \alpha) \mathbf{p} \quad \text{and} \quad q' = \beta \mathbf{q}^* + (1 - \beta) \mathbf{q}$$

where  $(\mathbf{p}^*, \mathbf{q}^*)$  is a signaling system in  $\text{supp}(\mathbf{p}, \mathbf{q})$ , then

$$\langle (\xi, \eta), J(\xi, \eta) \rangle_{(\mathbf{p}, \mathbf{q})} = \mathbf{p}' \cdot A \mathbf{q}' - \mathbf{p} \cdot A \mathbf{q} > 0$$

(where  $A$  is the payoff matrix corresponding to the signaling game); the inequality follows from the fact that  $\mathbf{p}^* \cdot A \mathbf{q}^* > \mathbf{p} \cdot A \mathbf{q}$ . This shows that  $J$  is not negative semi-definite.  $\square$

Proposition 12 also applies to the Nash set of type (3). For this component, we can actually show more. As was already observed in Section 3, Equation (3) is compact and convex. This also holds for its counterparts in signaling games with  $n > 3$  [11]. The average payoff is constant along this component. The potential function  $V$  of (13) will thus take on a unique maximum on this component.

An analysis of the Nash sets (6)–(8) is more involved, however. One reason for this is the fact that (unlike Nash sets of type (3)–(5)) these Nash sets do not contain mixtures of pure strategies that are not included in the Nash set; in particular, no signaling system is in their support.

Most equilibria in Nash sets of type (6) are quasi-strict by Proposition 9. Thus, for the replicator dynamics, all transversal eigenvalues have a negative real part. The two remaining eigenvalues are zero. It is therefore *a priori* unclear what will happen under the selection-mutation dynamics (12). Numerical calculations for specific mutation parameters  $\varepsilon = \delta$  suggest that, in this case, there always exists a unique perturbed rest point close to each Nash set of type (6), which is linearly unstable, having two positive eigenvalues. Hence, its Morse index is two (see Figure 1 for a definition of the Morse index and the Poincaré index below). The same numerical procedures lead to the result that there are no perturbed rest points close to Nash sets of type (7) and (8); there are, however, unique perturbed rest points near each Nash set of type (3)–(5). The perturbed rest point close to the Nash set (3) has four positive eigenvalues (Morse index four); perturbed rest points close to (4) have three positive eigenvalues (Morse index three); and perturbed rest points close to (5) have one positive eigenvalue (Morse index one). These calculations were conducted with several specific values for  $\varepsilon$  close to zero.

These findings agree with index theory (Brouwer degree theory): the sum of the Poincaré indices of all (perturbed) rest points equals one, which is the Euler characteristic of  $S_{27} \times S_{27}$ . (The Poincaré index is equal to  $(-1)^{\text{Morse index}}$ .) For the perturbed dynamics (12), the sum of the Poincaré indices of the six signaling systems and the six rest points close to the Nash set (4) is zero; likewise, the sum of the Poincaré indices of the nine rest points close to Nash sets of type (5) and the nine rest points close to the Nash set (6) is zero, as well. The Poincaré index of the only remaining rest point is one. This information is summarized in Figure 1.



In addition to these numerical results, a more thorough analytical treatment of the selection-mutation dynamics would be desirable in order to find out more about the stability properties of neutrally-stable Nash sets. Because of the many dimensions of the mixed strategy space for the  $n = 3$  signaling game, such an analysis does not appear to be feasible for Nash sets of type (6)–(8). An analysis in terms of behavioral strategies is more promising.

## 5. Dynamics for Behavioral Strategies

The relationship between mixed strategies and behavioral strategies in extensive games of perfect recall is well understood [22]. In particular, if one is interested in the equilibrium structure of extensive form games, then no essential information is lost by focusing on behavioral strategies. We propose a similar move for the evolutionary dynamics of signaling games.

The selection-mutation dynamics on the space  $(S_n)^m \times (S_m)^n$  of behavioral strategies for signaling games is given by

$$\dot{p}_{ij} = p_{ij}(q_{ji} - a_i) + \varepsilon(1 - np_{ij}) \quad (14a)$$

$$\dot{q}_{ji} = q_{ji}(p_{ij} - b_j) + \delta(1 - mq_{ji}) \quad (14b)$$

where

$$a_i = \sum_l p_{il}q_{li} \quad \text{and} \quad b_j = \sum_k p_{kj}q_{jk}.$$

The selection-mutation dynamics in behavioral strategies can be thought of as a dynamics in the entries of the sender and receiver matrices, since these matrices are nothing but a representation of the game's behavioral strategies. Using (14) instead of (12) significantly reduces the dimensions of the state space. For  $n = m = 3$ , the mixed strategy space has effectively 52 dimensions, while the behavioral strategy space has only 12. For general  $m = n$  the mixed strategy space has  $2(n^n - 1)$  dimensions, whereas the space of behavioral strategies has  $2n(n - 1)$ . Moreover, it is easy to see that the Propositions 10 and 11 also hold for (14).

### 5.1. Replicator Dynamics

We begin by quickly reviewing some results for the replicator dynamics of signaling games, where the replicator dynamics is taken to be on behavior strategies; *i.e.*, we are looking at (14) with  $\varepsilon, \delta = 0$ . The results we state follow immediately from the corresponding results for mixed strategies [11,12].

First of all, it is clear that signaling systems are asymptotically stable. Next, the Nash sets (6) and (8) are not asymptotically stable, but they attract an open set of initial conditions (they are attractive). This follows from the fact that all Nash equilibria in such a Nash set, except the vertices are neutrally stable and, hence, Lyapunov stable under the replicator dynamics [12]. Straightforward calculations show that the transversal eigenvalues are negative, while the remaining eigenvalues are equal to zero. The center-manifold theorem then implies that the two types of Nash sets are attractive.

The other Nash sets are not attractive. They will be discussed in more detail in the following subsections. We start with the simple or less interesting cases, which include signaling systems and the Nash sets that are not even quasi-strict. We investigate the most interesting cases of neutrally stable Nash sets at the end of this section.

### 5.2. Signaling Systems

For  $n = m$ , the location of perturbed signaling systems can be easily computed. The state with  $p_{ii} = q_{ii} = 1$  and  $p_{ij} = q_{ij} = 0$  for  $i \neq j$  is the limit of a family of perturbed rest points of (14) as  $\varepsilon, \delta$  go to zero. We have  $a_i = b_j = 1$ , and hence, we obtain from (14) for  $i \neq j$ ,  $p_{ij} = \varepsilon$  and  $q_{ji} = \delta$  to first order; so,  $p_{ii} = 1 - (n - 1)\varepsilon$  and  $q_{jj} = 1 - (n - 1)\delta$ . The formulas for the other perturbed signaling systems are similar.

Since signaling systems are strict Nash equilibria, it is clear that perturbed signaling systems must be asymptotically stable for sufficiently small  $\varepsilon, \delta$ ; this follows from the continuity of the eigenvalues in  $\varepsilon$  and  $\delta$ .

### 5.3. The 1/3 Component

As was noted above, the  $\frac{1}{3}$  component is convex. As in the case of mixed strategy evolutionary dynamics, it follows that there is a unique maximum of the potential function (13) on this set. This determines a unique rest point of (14) in this set, which is given by

$$P = Q = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

By continuity, this rest point must be linearly unstable. Calculating the eigenvalues of the Jacobian matrix of (14) at  $(P, Q)$  for  $\varepsilon = \delta = 0$  shows that there are four zero eigenvalues, four positive eigenvalues and four negative eigenvalues. It follows that for small positive  $\varepsilon, \delta$ , the four zero eigenvalues become negative. Hence, the Morse index of the perturbed rest point is four.

### 5.4. Type (4) Rest Points

In terms of behavior strategies, this type of rest point is regular. It follows from the implicit function theorem that there exists a unique family of perturbed rest points of (14) close to (4) for sufficiently small  $\varepsilon$  and  $\delta$ . By continuity, the perturbed rest point must be linearly unstable. In the dynamics (14) with  $\varepsilon = \delta = 0$ , the rest points of type (4) have three positive rest points. Hence, the Morse index of the perturbed rest point is three.

### 5.5. Type (7) Nash Sets

Analyzing the dynamic behavior of (14) close to the other Nash sets requires more work. We start by considering Nash sets of type (7). We are going to show that Nash sets of type (7) do not persist under the selection-mutation dynamics (14). The intuitive reason for this is that

these Nash sets are not quasi-strict; there always is a signaling system that is not used in the sender population. Introducing senders who utilize the unused signals and receivers who respond to it appropriately will destabilize any equilibrium in this set. Since such mutations will be introduced in the dynamics (14), no rest point of type (7) can persist under selection-mutation dynamics.

**Proposition 13.** *The behavior strategy given by*

$$\begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 - \lambda & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ \mu & 0 & 1 - \mu \end{pmatrix}, \tag{15}$$

with  $0 < \lambda, \mu < 1$ , is not the limit of a sequence of perturbed rest points. Analogous statements are true for the other Nash sets of type (7).

**Proof.** Assuming the existence of perturbed rest point close to (15), we can calculate the first-order expressions of the following variables,

$$p_{12} = \varepsilon, p_{13} = \frac{\varepsilon}{1 - \mu}, p_{31} = \varepsilon, p_{33} = \frac{\varepsilon}{\mu}$$

and

$$q_{12} = \frac{\delta}{1 - \lambda}, q_{13} = \delta, q_{21} = \delta, q_{22} = \frac{\delta}{\lambda}.$$

Set  $x = p_{23}$  and  $y = q_{32}$ . Note that  $x, y \rightarrow 0$ , as  $\varepsilon, \delta \rightarrow 0$ . Assuming the existence of a perturbed rest point, we must have

$$\begin{aligned} \dot{x} &= x(y - a_2) + \varepsilon(1 - 3x) = 0 \\ \dot{y} &= y(x - b_3) + \delta(1 - 3y) = 0, \end{aligned}$$

where  $a_2 = xy + L\delta$  and  $b_3 = xy + M\varepsilon$  and

$$L = \frac{\lambda}{1 - \lambda} + \frac{1 - \lambda}{\lambda}, \quad M = \frac{\mu}{1 - \mu} + \frac{1 - \mu}{\mu}.$$

Hence

$$x(y - xy - L\delta) + \varepsilon(1 - 3x) = 0 \tag{16a}$$

$$y(x - xy - M\varepsilon) + \delta(1 - 3y) = 0. \tag{16b}$$

This system of equations has no solutions for sufficiently small  $\varepsilon, \delta$ . To see this, notice that (16b) implies

$$x(1 - y) < M\varepsilon$$

and, hence, from (16a)

$$\varepsilon = \frac{x}{1 - 3x} (L\delta - y(1 - x)) < \frac{x}{1 - 3x} L\delta < \frac{M\varepsilon L\delta}{(1 - 3x)(1 - y)}.$$

Therefore

$$(1 - 3x)(1 - y) < ML\delta,$$

which fails to hold as  $\delta \rightarrow 0$ .  $\square$

5.6. Type (5) Nash Sets

We now investigate the dynamics of (14) close to (5). By a symmetric perturbed rest point, we shall mean a rest point that exhibits certain symmetry properties (the meaning of which will become clear below). It should be emphasized that there might be other rest points of (14) close to the corresponding Nash sets that are not symmetric.

**Theorem 14.** *There exists a unique family of (symmetric) perturbed rest points that converges to the equilibrium*

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

from the Nash set (5), as  $\varepsilon, \delta \rightarrow 0$ . To a first approximation, these perturbed rest points are given by

$$\tilde{P} = \begin{pmatrix} 1 - 2\varepsilon & \varepsilon & \varepsilon \\ 2\varepsilon & \frac{1}{2} - \varepsilon & \frac{1}{2} - \varepsilon \\ 2\varepsilon & \frac{1}{2} - \varepsilon & \frac{1}{2} - \varepsilon \end{pmatrix}, \tilde{Q} = \begin{pmatrix} 1 - 2\delta & \delta & \delta \\ 2\delta & \frac{1}{2} - \delta & \frac{1}{2} - \delta \\ 2\delta & \frac{1}{2} - \delta & \frac{1}{2} - \delta \end{pmatrix}.$$

Analogous statements are true for the other Nash sets of type (5).

**Proof.** If it exists, a symmetric perturbed rest point has the form

$$\begin{pmatrix} 1 - 2p & p & p \\ 2x & \frac{1-2x}{2} & \frac{1-2x}{2} \\ 2x & \frac{1-2x}{2} & \frac{1-2x}{2} \end{pmatrix}, \begin{pmatrix} 1 - 2q & q & q \\ 2y & \frac{1-2y}{2} & \frac{1-2y}{2} \\ 2y & \frac{1-2y}{2} & \frac{1-2y}{2} \end{pmatrix} \tag{17}$$

For this to exist, the following system must have a solution for  $\varepsilon, \delta > 0$ :

$$\begin{aligned} p[2y - 4yp - (1 - 2p)(1 - 2q)] + \varepsilon(1 - 3p) &= 0 \\ q[2x - 4xq - (1 - 2p)(1 - 2q)] + \delta(1 - 3q) &= 0 \\ \left(\frac{1}{2} - x\right) \left[\frac{1}{2} - y - 2xq - 2\left(\frac{1}{2} - x\right)\left(\frac{1}{2} - y\right)\right] + \varepsilon\left(3x - \frac{1}{2}\right) &= 0 \\ \left(\frac{1}{2} - y\right) \left[\frac{1}{2} - x - 2yp - 2\left(\frac{1}{2} - x\right)\left(\frac{1}{2} - y\right)\right] + \delta\left(3y - \frac{1}{2}\right) &= 0 \end{aligned} \tag{18}$$

The left-hand sides of these equations can be understood as a function  $F : \mathbb{R}^2 \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ , where the components of  $F(\varepsilon, \delta, p, q, x, y)$  are given by the left-hand side expressions of (18). We see that  $F(0, 0, 0, 0, 0, 0) = 0$ . Let  $M$  be the four-by-four matrix of partial derivatives of  $F$  with respect to  $x, y, p, q$  evaluated at the point  $(0, 0, 0, 0, 0, 0)$ . Then,  $M$  is given by

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Thus,  $M$  is regular, and it follows from the implicit function theorem that there exist unique solutions for (18) in terms of  $\varepsilon$  and  $\delta$  provided that  $\varepsilon, \delta$  are sufficiently close to zero. Indeed, linearizing the system (18) at the point  $p, q, x, y = 0$  gives

$$\begin{aligned} -p + \varepsilon &= 0 \\ -q + \delta &= 0 \\ \frac{1}{2}(x - \varepsilon) &= 0 \\ \frac{1}{2}(y - \delta) &= 0 \end{aligned} \tag{19}$$

Hence,  $p, x = \varepsilon + O(2)$  and  $q, y = \delta + O(2)$ , where  $O(2)$  stands for terms of order  $\varepsilon^2, \varepsilon\delta, \delta^2$  or higher.  $\square$

Points in the relative interior of Nash sets like (5) have one positive eigenvalue under the replicator dynamics. Hence, the Morse index is one. It follows that each rest point in the family of rest points given in Theorem 14 is linearly unstable for small  $\varepsilon, \delta$ . By considering the characteristic polynomial of the Jacobian of (14) evaluated at one of these perturbed rest points, we can obtain sharper results.

**Theorem 15.** *Let  $(\tilde{P}, \tilde{Q})$  be the perturbed rest point given by Theorem 14. Then, the following two statements are true for sufficiently small mutation rates:*

1. *If  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ , then the Jacobian matrix of the selection-mutation dynamics (14) is hyperbolic and has one positive eigenvalue.*
2. *If  $\varepsilon > 2\delta$  or  $\varepsilon < \frac{\delta}{2}$ , the Jacobian matrix of the selection-mutation dynamics (14) is hyperbolic and has two positive eigenvalues.*

**Proof.** Calculating and factorizing the characteristic polynomial  $\chi(x)$  of the Jacobian matrix evaluated at the rest point  $(\tilde{P}, \tilde{Q})$  of Theorem 14 yields four factors; up to higher order terms in  $\varepsilon, \delta$ , a Taylor expansion around  $(\varepsilon, \delta) = (0, 0)$  yields the following approximations of the factors (up to multiplication by a positive constant):

$$2\varepsilon + 2\delta - 1 + 8(\delta + \varepsilon)x + 4x^2 \tag{20}$$

$$2\delta - \varepsilon + (1 + 3\delta - 3\varepsilon)x + (3 + 2\delta - 6\varepsilon)x^2 + 2x^3 \tag{21}$$

$$2\varepsilon - \delta + (1 - 3\delta + 3\varepsilon)x + (3 - 6\delta + 2\varepsilon)x^2 + 2x^3 \tag{22}$$

$$1 - 7\delta - 7\varepsilon + (6 - 33\delta - 33\varepsilon)x + (13 - 50\delta - 50\varepsilon)x^2 + (12 - 24\delta - 24\varepsilon)x^3 + 4x^4 \tag{23}$$

Setting the first factor to zero yields two solutions for  $x$ , which are given by (again up to higher order terms)

$$-\delta - \varepsilon \pm \frac{1}{2}\sqrt{1 - 2\delta - 2\varepsilon} \approx \frac{1}{2} - \frac{3}{2}(\delta + \varepsilon), -\frac{1}{2} - \frac{3}{2}(\delta + \varepsilon)$$

For all sufficiently small  $\varepsilon, \delta$ , one will be positive and the other one negative.

The eigenvalues given by the second and the third factor depend on  $\varepsilon$  and  $\delta$ . By examining the discriminant of (21), it can be shown that the corresponding cubic equation has three solutions, which we will denote by  $x_1, x_2, x_3$ . Then

$$2\delta - \varepsilon + (1 + 3\delta - 3\varepsilon)x + (3 + 2\delta - 6\varepsilon)x^2 + 2x^3 = 2(x - x_1)(x - x_2)(x - x_3).$$

It follows that

$$2\delta - \varepsilon = -2x_1x_2x_3. \tag{24}$$

Now, setting  $\varepsilon, \delta$  to zero in the second factor of (21), and solving the cubic equation leads to three solutions for  $x$ , two of which are negative and one of which is zero (this last one is one of the zero eigenvalues of the replicator dynamics at this rest point). If  $\varepsilon, \delta$  are small, it follows that two solutions of  $x_1, x_2, x_3$  must also be negative. Returning to (24), this implies that if  $2\delta - \varepsilon > 0$ , then  $x_1, x_2, x_3 < 0$ . Conversely, if  $2\delta - \varepsilon < 0$ , then exactly one solution must be positive.

A similar argument leads to the same conclusions for the third factor of (22). In this case, if  $2\varepsilon - \delta > 0$ , then all eigenvalues coming from this factor are negative. If, on the other hand,  $2\varepsilon - \delta < 0$ , then one of the eigenvalues given by the third factor will be positive. Thus, from the second and the third factor, we get only negative eigenvalues if  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ ; and we get exactly one positive eigenvalue if  $\varepsilon > 2\delta$  or  $\varepsilon < \frac{\delta}{2}$ .

The fourth factor can only yield negative eigenvalues. This can be seen by first setting  $\varepsilon, \delta = 0$  and then solving the corresponding quartic equation for  $x$ . This yields four negative solutions, which will remain negative for (23) as long as  $\varepsilon, \delta$  are small.  $\square$

**Example 16.** The case  $\varepsilon = \delta$  is especially important and was considered in the previous section within the context of the selection-mutation dynamics in mixed strategies. There, we observed that numerical calculations of eigenvalues of perturbed rest points close to (5) yielded one positive eigenvalue. Theorem 15 is an analytical confirmation of this result, since if  $\varepsilon = \delta$ , the clause of its first statement is fulfilled.

### 5.7. Neutrally-Stable Nash Sets of Type (6)

By far the most important alternatives to signaling systems are neutrally-stable strategies. In this section, as in the previous one, we first prove the existence of certain perturbed rest points close to Nash sets of type (6) and then show that they will be linearly unstable under the selection-mutation dynamics.

Let us first note a simple observation concerning (6). The next lemma asserts that if a point in the relative interior of a neutrally-stable Nash set is a limit of perturbed rest points, it must be the barycenter of the Nash set. Notice, however, that this result does not imply that there is at most one perturbed rest point close to a neutrally-stable Nash set. Several branches of curves of rest points might meet at the barycenter of the Nash set.

**Lemma 17.** *Suppose that*

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \lambda & 1 - \lambda \end{pmatrix}, \begin{pmatrix} \mu & 1 - \mu & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{25}$$

*is a limit of perturbed rest points of the selection-mutation dynamics. Then,  $\lambda = 1/2$  and  $\mu = 1/2$ .*

**Proof.** If (25) is a limit of perturbed rest points, then approximate solutions up to higher order terms in  $\varepsilon, \delta$  are given by

$$\begin{pmatrix} 1 - \frac{2\varepsilon}{\mu} & \frac{\varepsilon}{\mu} & \frac{\varepsilon}{\mu} \\ 1 - \frac{2\varepsilon}{1-\mu} & \frac{\varepsilon}{1-\mu} & \frac{\varepsilon}{1-\mu} \\ \varepsilon & p_{32} & p_{33} \end{pmatrix}, \begin{pmatrix} q_{11} & q_{12} & \delta \\ \frac{\delta}{\lambda} & \frac{\delta}{\lambda} & 1 - \frac{2\delta}{\lambda} \\ \frac{\delta}{1-\lambda} & \frac{\delta}{1-\lambda} & 1 - \frac{2\delta}{1-\lambda} \end{pmatrix}$$

where

$$\begin{aligned} \dot{p}_{32} &= p_{32}(q_{23} - a_3) + \varepsilon(1 - 3p_{32}) = 0 \\ \dot{p}_{33} &= p_{33}(q_{33} - a_3) + \varepsilon(1 - 3p_{33}) = 0 \\ \dot{q}_{11} &= q_{11}(p_{11} - b_1) + \delta(1 - 3q_{11}) = 0 \\ \dot{q}_{12} &= q_{12}(p_{21} - b_1) + \delta(1 - 3q_{12}) = 0 \end{aligned}$$

and

$$\begin{aligned} a_3 &= p_{31}q_{13} + p_{32}q_{23} + p_{33}q_{33} \\ b_1 &= p_{11}q_{11} + p_{21}q_{12} + p_{31}q_{13}. \end{aligned}$$

Since  $\dot{p}_{32} = 0$ ,

$$p_{32}((1 - p_{32})(q_{23} - q_{33}) - p_{31}(q_{13} - q_{33})) + \varepsilon(1 - 3p_{32}) = 0$$

and thus

$$p_{32} \left( 2\delta(1 - p_{32}) \frac{2\lambda - 1}{\lambda(1 - \lambda)} + \varepsilon + O(\varepsilon^2) \right) + \varepsilon(1 - 3p_{32}) = 0.$$

The first term of the Taylor expansion in  $p_{32} = \lambda$  of the left-hand side yields

$$(2\delta - \varepsilon)(2\lambda - 1) = 0.$$

Thus, if  $\varepsilon \neq 2\delta$ , then  $\lambda = 1/2$ . A similar argument for  $q_{11}$  shows that if  $\delta \neq 2\varepsilon$ , then  $\mu = 1/2$ .

Suppose  $\delta = 2\varepsilon$ . Since  $\dot{q}_{12} = 0$ , we get:

$$q_{12} \left( 2\varepsilon(1 - q_{12}) \frac{1 - 2\mu}{\mu(1 - \mu)} + \delta + O(\varepsilon^2) \right) + \delta(1 - 3q_{12}) = 0$$

The first term of the corresponding Taylor expansion is given by:

$$2\varepsilon(1 - 2\mu) - 4\varepsilon\mu + 2\varepsilon = 0$$

Thus,  $\mu = 1/2$ . By a similar argument,  $\dot{p}_{32} = 0$  implies  $\lambda = 1/2$ . Along the same lines, it can be shown that  $\lambda = \mu = 1/2$  if  $\delta = 2\varepsilon$ . One just has to find the first terms of the corresponding Taylor expansions for  $\dot{p}_{33} = 0$  and  $\dot{q}_{11} = 0$ .  $\square$

We continue by proving the existence of certain symmetric perturbed rest points close to the two-dimensional Nash set (6). Recall that we had some results concerning Nash sets of type (5) for the selection-mutation dynamics in mixed strategies (12), whereas similar results for Nash sets of type (6) were out of reach for the system (12).

**Theorem 18.** *There exists a unique family of (symmetric) perturbed rest points that converges to the equilibrium*

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

from the Nash set (6), as  $\varepsilon, \delta \rightarrow 0$ . To a first approximation, these perturbed rest points are given by

$$\bar{P} = \begin{pmatrix} 1 - 4\varepsilon & 2\varepsilon & 2\varepsilon \\ 1 - 4\varepsilon & 2\varepsilon & 2\varepsilon \\ \varepsilon & \frac{1-\varepsilon}{2} & \frac{1-\varepsilon}{2} \end{pmatrix}, \bar{Q} = \begin{pmatrix} \frac{1-\delta}{2} & \frac{1-\delta}{2} & \delta \\ 2\delta & 2\delta & 1 - 4\delta \\ 2\delta & 2\delta & 1 - 4\delta \end{pmatrix}.$$

Analogous statements are true for the other Nash sets of type (6).

**Proof.** A symmetric perturbed rest point has the form

$$\begin{pmatrix} 1 - 2p & p & p \\ 1 - 2p & p & p \\ 2x & \frac{1-2x}{2} & \frac{1-2x}{2} \end{pmatrix}, \begin{pmatrix} \frac{1-2y}{2} & \frac{1-2y}{2} & 2y \\ q & q & 1 - 2q \\ q & q & 1 - 2q \end{pmatrix}, \tag{26}$$

provided that it exists. It exists if the following system has a solution for  $\varepsilon, \delta > 0$ :

$$\begin{aligned} p \left[ q - 2pq - \left( \frac{1}{2} - y \right) (1 - 2p) \right] + \varepsilon(1 - 3p) &= 0 \\ q \left[ p - 2pq - \left( \frac{1}{2} - x \right) (1 - 2q) \right] + \delta(1 - 3q) &= 0 \\ \left( \frac{1}{2} - x \right) [1 - 2q - 4xy - (1 - 2x)(1 - 2q)] + \varepsilon \left( 3x - \frac{1}{2} \right) &= 0 \\ \left( \frac{1}{2} - y \right) [1 - 2p - 4xy - (1 - 2y)(1 - 2p)] + \delta \left( 3y - \frac{1}{2} \right) &= 0 \end{aligned} \tag{27}$$

As in the case of the proof of Theorem 14, the left-hand sides of these equations can be understood as a function  $G : \mathbb{R}^2 \times \mathbb{R}^4 \rightarrow \mathbb{R}^4$ , where the components of  $G(\varepsilon, \delta, p, q, x, y)$  are given by the left-hand side expressions of (27). We see that  $G(0, 0, 0, 0, 0, 0) = 0$ . Let  $N$  be the four-by-four matrix of partial derivatives of  $G$  with respect to  $x, y, p, q$  evaluated at the point  $(0, 0, 0, 0, 0, 0)$ .  $N$  is given by

$$\begin{pmatrix} -\frac{1}{2} & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$



Thus,  $N$  is regular and the implicit function theorem implies that there exist unique solutions for (27) in terms of  $\varepsilon$  and  $\delta$  provided that  $\varepsilon, \delta$  are sufficiently close to zero. Indeed, linearizing the system (27) at the point  $p, q, x, y = 0$  yields

$$\begin{aligned} -\frac{p}{2} + \varepsilon &= 0 \\ -\frac{q}{2} + \delta &= 0 \\ x - \frac{\varepsilon}{2} &= 0 \\ y - \frac{\delta}{2} &= 0 \end{aligned} \tag{28}$$

Hence,  $p = 2\varepsilon + O(2)$ ,  $q = 2\delta + O(2)$ ,  $x = \frac{\varepsilon}{2} + O(2)$  and  $y = \frac{\delta}{2} + O(2)$ , where  $O(2)$  denotes terms of order  $\varepsilon^2, \varepsilon\delta, \delta^2$  or higher.  $\square$

Next, we determine the stability properties and the Morse index of the rest point whose existence was proven immediately above.

**Theorem 19.** *Let  $(\bar{P}, \bar{Q})$  be the perturbed rest point given by Theorem 18. Then, the following two statements are true for sufficiently small mutation rates:*

1. *If  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ , then the Jacobian matrix of the selection-mutation dynamics (14) is hyperbolic and has two positive eigenvalues.*
2. *If  $\varepsilon > 2\delta$  or  $\varepsilon < \frac{\delta}{2}$ , the Jacobian matrix of the selection-mutation dynamics (14) is hyperbolic and has one positive eigenvalue.*

**Proof.** By calculating and factorizing the characteristic polynomial  $\chi(x)$  of the Jacobian matrix evaluated at the rest point of Theorem 18, we get four factors (we omit higher terms in  $\varepsilon, \delta$ ):

$$1 - 3\varepsilon - 3\delta + (4 - 6\varepsilon - 6\delta)x + 4x^2 \tag{29}$$

$$2\delta - 4\varepsilon + (1 + 5\delta - 15\varepsilon)x + (4 + 2\delta - 14\varepsilon)x^2 + 4x^3 \tag{30}$$

$$2\varepsilon - 4\delta + (1 + 5\varepsilon - 15\delta)x + (4 + 2\varepsilon - 14\delta)x^2 + 4x^3 \tag{31}$$

$$1 - 11\varepsilon - 11\delta + (6 - 48\varepsilon - 48\delta)x + (13 - 67\varepsilon - 67\delta)x^2 + (12 - 30\varepsilon - 30\delta)x^3 + 4x^4 \tag{32}$$

The first and the last factors only yield negative eigenvalues. This can be seen by first setting  $\varepsilon, \delta = 0$  and then solving the resulting quadratic and the cubic equations. The solutions are negative and, thus, continue to be negative for small positive  $\varepsilon, \delta$ .

The eigenvalues given by the second and the third factor depend on  $\varepsilon$  and  $\delta$ . Since the discriminant of (30) is positive, the corresponding cubic equation has three solutions, which we will denote by  $x_1, x_2, x_3$ . Then, in this case, we have

$$2\delta - 4\varepsilon + (1 + 5\delta - 15\varepsilon)x + (4 + 2\delta - 14\varepsilon)x^2 + 4x^3 = 4(x - x_1)(x - x_2)(x - x_3)$$

and thus

$$2\delta - 4\varepsilon = -4x_1x_2x_3. \tag{33}$$

Now, setting  $\varepsilon, \delta$  to zero in the second factor of (30), and solving the resulting cubic equation leads to two negative solutions for  $x$  and another one where  $x = 0$ . If  $\varepsilon, \delta$  are small, it follows that two values among  $x_1, x_2, x_3$  must also be negative. By (33), this implies that if  $2\delta - 4\varepsilon > 0$ , then  $x_1, x_2, x_3 < 0$ ; and if  $2\delta - 4\varepsilon < 0$ , then exactly one solution must be positive. For the third factor of (31), it can be shown that if  $2\varepsilon - 4\delta > 0$ , then  $x_1, x_2, x_3 < 0$ ; and if  $2\varepsilon - 4\delta < 0$ , then exactly one solution must be positive. In sum, the rest point  $(\bar{P}, \bar{Q})$  has two positive eigenvalues, as long as  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ , and one if one positive eigenvalue if  $\varepsilon > 2\delta$  or  $\varepsilon < \frac{\delta}{2}$ .  $\square$

**Example 20.** Consider again the example  $\varepsilon = \delta$ . For the selection-mutation dynamics in mixed strategies, we noted in the last section that numerical calculations suggest that there exist unique rest points close to the barycenter of the Nash set (6) that have two positive eigenvalues. Theorem 19 is an analytical confirmation of the latter point in the framework of behavioral strategies. If  $\varepsilon = \delta$ , then the proviso of the first statement of the theorem clearly holds, and thus, the symmetric perturbed rest point has two positive eigenvalues.

Theorems 14 and 18 show that a family of (symmetric) perturbed rest points always exists close to Nash sets, such as (5) and (6). The existence of perturbed rest points does not depend on the ratio  $\delta/\varepsilon$ . Their stability properties on the other hand do: while the symmetric perturbed rest points are linearly unstable for sufficiently small  $\varepsilon, \delta$  regardless of the ratio  $\delta/\varepsilon$ , the number of positive eigenvalues depends on it.

### 5.8. Neutrally-Stable Nash Sets of Type (8)

The ratio  $\delta/\varepsilon$  is also relevant for the existence of perturbed rest points close to Nash sets of type (8), to which we now turn our attention.

**Theorem 21.** *Let*

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} \lambda & 1 - \lambda & 0 \\ 0 & 0 & 1 \\ \mu & \nu & 1 - \mu - \nu \end{pmatrix} \tag{34}$$

with  $0 < \mu < \lambda$  and  $0 < \nu < 1 - \lambda$ . Denote by  $(\bar{P}, \bar{Q})$  those Nash equilibria in (34) with  $\lambda = \frac{1}{2}$  and  $\mu = \nu$ .

1. If  $2 > \frac{\delta}{\varepsilon} > \frac{1}{2}$ , then there are no perturbed rest points converging to  $(\bar{P}, \bar{Q})$  as  $\varepsilon, \delta \rightarrow 0$ .
2. If  $\frac{\delta}{\varepsilon} < \frac{1}{2}$ , then there exists a family of perturbed rest points converging to some  $(\bar{P}, \bar{Q})$  as  $\varepsilon, \delta \rightarrow 0$ .
3. If  $\frac{\delta}{\varepsilon} > 2$ , then there exists a family of perturbed rest points converging to some  $(\bar{P}, \bar{Q})$  as  $\varepsilon, \delta \rightarrow 0$ .

**Proof.** The Nash equilibrium (34) is quasi-strict, with  $a_1 = \lambda, a_2 = 1 - \lambda, a_3 = 0, b_1 = b_2 = 1, b_3 = 0$ . Hence, to first order, a perturbed equilibrium nearby satisfies

$$p_{12} = \frac{\varepsilon}{\lambda}, p_{13} = \frac{\varepsilon}{\lambda - \mu}, p_{22} = \frac{\varepsilon}{1 - \lambda}, p_{23} = \frac{\varepsilon}{1 - \lambda}, p_{31} = \varepsilon, p_{33} = \frac{\varepsilon}{\mu + \nu}$$

and

$$q_{13} = \delta, \quad q_{21} = \delta, \quad q_{22} = \delta.$$

For the remaining variables, we make an ansatz

$$q_{11} = \lambda + x, \quad q_{31} = \mu + y, \quad q_{32} = \nu + z$$

with  $x, y, z$  small if  $\varepsilon, \delta$  are small. Hence,  $q_{12} = 1 - \lambda - x - \delta, q_{23} = 1 - 2\delta, q_{33} = 1 - \mu - \nu - y - z.$

Inserting all this into the three equations

$$\begin{aligned} \dot{q}_{11} &= q_{11}(p_{11} - b_1) + \delta(1 - 3q_{11}) = 0 \\ \dot{q}_{31} &= q_{31}(p_{13} - b_3) + \delta(1 - 3q_{31}) = 0 \\ \dot{q}_{32} &= q_{32}(p_{23} - b_3) + \delta(1 - 3q_{32}) = 0 \end{aligned} \tag{35}$$

and letting  $\delta, \varepsilon \rightarrow 0$  with  $\frac{\delta}{\varepsilon} = \text{const} > 0$  and  $x, y, z \rightarrow 0$  gives three equations for the ratio  $\delta/\varepsilon$

$$\begin{aligned} (1 - 2\lambda)\frac{\delta}{\varepsilon} &= 1 - 2\lambda + \lambda(1 - \lambda) \left( \frac{1}{\lambda - \mu} - \frac{1}{1 - \lambda - \nu} \right) \\ (1 - 3\mu)\frac{\delta}{\varepsilon} &= \mu \left( -\frac{1 - \mu}{\lambda - \mu} + \frac{\nu}{1 - \lambda - \nu} + \frac{1 - \mu - \nu}{\mu + \nu} \right) \\ (1 - 3\nu)\frac{\delta}{\varepsilon} &= \nu \left( -\frac{1 - \nu}{1 - \lambda - \nu} + \frac{\mu}{\lambda - \mu} + \frac{1 - \mu - \nu}{\mu + \nu} \right) \end{aligned} \tag{36}$$

and, hence, two equations between  $\lambda, \mu, \nu.$

Under the symmetry assumptions  $\lambda = \frac{1}{2}, \mu = \nu,$  these two equations hold, and (36) reduces to

$$\frac{\delta}{\varepsilon} = \frac{1 - 6\mu}{2 - 6\mu} \tag{37}$$

and thus

$$\mu = \frac{1}{6} \frac{\varepsilon - 2\delta}{\varepsilon - \delta}.$$

Hence, we obtain two disjoint intervals of solutions: for  $\frac{\delta}{\varepsilon} < \frac{1}{2},$  we have  $0 < \mu < \frac{1}{6},$  and for  $\frac{\delta}{\varepsilon} > 2,$  we have  $\frac{1}{3} < \mu < \frac{1}{2}.$  □

There are 18 Nash sets of the same structure as  $(P, Q)$  ( $P$  being a zero-one matrix). If we interchange  $P$  and  $Q,$  we obtain eighteen further Nash sets. In these cases,  $\varepsilon$  and  $\delta$  switch roles in the proof of Proposition 21. Therefore, we have 36 perturbed rest points near Nash sets of type (8) in cases 2 and 3 of the theorem. One can show that the perturbed rest points from case 2 have Morse index two, whereas those in case 3 have Morse index one. Hence, the total Poincare index of all 36 perturbed rest points is zero.

## 6. Conclusions

As we noted above, the main reason for studying the selection-mutation dynamics of signaling games is that the replicator dynamics of those games are not structurally unstable. We have seen in Section 3 that all Nash equilibria other than signaling systems are elements of components of equilibria. This implies that perturbations of the replicator equations will in general destroy the rest points corresponding to these equilibria. Thus, the results of the replicator dynamics need to be supplemented by results obtained for perturbations of the replicator dynamics. The selection-mutation dynamics arguably is an arguably plausible perturbation. It explicitly introduces an element of evolutionary processes (mutation) that is absent from the replicator equations; and, it does so in a way that leads back to the replicator dynamics in the no-mutation limit.

In the following discussion, we focus on the important case where  $\varepsilon$  is approximately the same as  $\delta$  (i.e.,  $2\varepsilon > \delta > \frac{\varepsilon}{2}$ ). For the selection-mutation dynamics, we have obtained a partial analysis of the stability properties of various rest points, most notably those which are close to Nash equilibria that are not signaling systems of the signaling game. In particular, we proved the existence of certain perturbed rest points and showed that these rest points are linearly unstable for the selection mutation dynamics. Only perturbed rest points close to signaling systems are asymptotically stable. Hence, a central result of the replicator dynamics, that there are open sets of points with trajectories converging to a neutrally-stable Nash set, does not appear to carry over to the selection-mutation dynamics. The zero eigenvalues of neutrally-stable rest point turn into positive eigenvalues under the selection-mutation dynamics.

These results are partial in two respects. In the first place, we did not analyze signaling games with more than three states, acts and signals. However, as mentioned above, the basic features of the replicator dynamics of signaling games with three signals are the same as those of signaling games with more than three signals. Thus, the same may be true for the selection-mutation dynamics.

In the second place, we do not show that there are no other families of perturbed rest points. Although we cannot prove it, we conjecture that this is not the case. There are two arguments supporting this conjecture. First, we conducted extensive numerical calculations for specific parameter values of  $\varepsilon, \delta$  by using Newton's method and were not able to find any other perturbed rest points. Second, as summarized in Figure 1, our conjecture is consistent with index theory.

The selection-mutation dynamics of signaling games with two signals, states and acts was explored in [13]. There, we distinguished between the cases of equal and unequal state probabilities. In the case of equal state probabilities, we found that almost all trajectories converge to one of the signaling systems, while in the case of unequal state probabilities, this is not the case for all values of the two mutation parameters. By adopting the Trapa–Nowak framework, we assume equal state probabilities here, as well. We believe that one can show for the case of three or more signals with unequal state probabilities that there is an open set of points with trajectories that do not converge to a signaling system. The reason is that, as in the case of two signals, if state probabilities are unequal, there is an act that is a strictly better guess than other acts in the absence of informative signals. This can make it advantageous for a receiver to ignore signals and, thus, stabilize non-communicative behavior under the replicator dynamics and under the selection mutation dynamics.

The kinds of problems analyzed in this article are closely connected to the problem of equilibrium selection in extensive form games. For many extensive form games, non-isolated Nash equilibria are typical and persist under perturbations of terminal-node payoffs [9,10]. The effect of mutation on the dynamics close to such Nash equilibria was also studied by Binmore and Samuelson [7,8] in the context of games other than signaling games. Their results are close to ours in that they also emphasize the dependence of dynamical outcomes on mutation rates. However, the signaling game studied here has a considerably larger strategy space than the games considered by Binmore and Samuelson.

There are other signaling games for which it would be important to know the stability of rest points under the selection-mutation dynamics. One example is signaling games with more signals than states. The equilibrium structure of these games was analyzed in Donaldson *et al.* [23]. It is shown there that the Nash equilibria where senders communicate as much as possible and receivers react optimally are elements of one component of Nash equilibria. Without a more detailed analysis, it does not seem to be clear how the selection-mutation dynamics behave close to this component. For this game, as in the signaling game studied in this paper, the selection-mutation dynamics is not the only process that should be considered when one is interested in the robustness of the results obtained for the replicator dynamics, but it is a rather simple and at least partially tractable dynamics for which sharp results can be obtained.

### Acknowledgments

Hofbauer's work was supported by the WWTF grant (Project Number MA09-017) of the city of Vienna, Austria. Huttegger's work was supported by the National Science Foundation under Grant No. EF-1038456. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

### Author Contributions

Both authors contributed equally to this article.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Skyrms, B. *Signals: Evolution, Learning, and Information*; Oxford University Press: Oxford, UK, 2010.
2. Lewis, D. *Convention. A Philosophical Study*; Harvard University Press: Harvard, MA, USA, 1969.
3. Crawford, V.; Sobel, J. Strategic information transmission. *Econometrica* **1982**, *50*, 1431–1451.
4. Nowak, M.A.; Krakauer, D.C. The evolution of language. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8028–8033.
5. Skyrms, B. *Evolution of the Social Contract*; Cambridge University Press: Cambridge, UK, 1996.

6. Wärneryd, K. Cheap talk, coordination and evolutionary stability. *Games Econ. Behav.* **1993**, *5*, 532–546.
7. Binmore, K.; Samuelson, L. Drift. *Eur. Econ. Rev.* **1994**, *38*, 859–867.
8. Binmore, K.; Samuelson, L. Evolutionary drift and equilibrium selection. *Rev. Econ. Stud.* **1999**, *66*, 363–394.
9. Cressman, R. *Evolutionary Dynamics and Extensive Form Games*; MIT Press: Cambridge, MA, USA, 2003.
10. Huttegger, S.M. Generic properties of evolutionary games and adaptationism. *J. Philos.* **2010**, *107*, 80–102.
11. Huttegger, S.M. Evolution and the explanation of meaning. *Philos. Sci.* **2007**, *74*, 1–27.
12. Pawlowitsch, C. Why evolution does not always lead to an optimal signaling system. *Games Econ. Behav.* **2008**, *63*, 203–226.
13. Hofbauer, J.; Huttegger, S.M. Feasibility of communication in binary signaling games. *J. Theor. Biol.* **2008**, *254*, 843–849.
14. Trapa, P.E.; Nowak, M.A. Nash equilibria for an evolutionary language Game. *J. Math. Biol.* **2000**, *41*, 172–188.
15. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
16. Jansen, M.J.M. On the definition of the strategic stability of equilibria. *Nav. Res. Logist. Q.* **1981**, *28*, 286–295.
17. Jansen, M.J.M.; Jurg, A.P.; Borm, P.E.M. On strictly perfect sets. *Games Econ. Behav.* **1994**, *6*, 400–415.
18. Nash, J. Noncooperative games. *Ann. Math.* **1951**, *54*, 289–295.
19. Hofbauer, J. The selection mutation equation. *J. Math. Biol.* **1985**, *23*, 41–53.
20. Bürger, R.; Hofbauer, J. Mutation load and mutation-selection balance in quantitative genetic traits. *J. Math. Biol.* **1994**, *32*, 193–218.
21. Ritzberger, K. *Foundations of Non-Cooperative Game Theory*; Oxford University Press: Oxford, UK, 2002.
22. Kuhn, H.W. *Lectures on the Theory of Games*; Princeton University Press: Princeton, NJ, USA, 2003.
23. Donaldson, M.C.; Lachmann, M.; Bergstrom, C. The evolution of functionally referential meaning in a structured world. *J. Theor. Biol.* **2007**, *246*, 225–233.