



# Feasibility of communication in binary signaling games

Josef Hofbauer<sup>a</sup>, Simon M. Huttegger<sup>b,\*</sup>

<sup>a</sup> Department of Mathematics, University of Vienna, Nordbergstraße 15, A-1090 Vienna, Austria

<sup>b</sup> Department of Logic and Philosophy of Science, 5100 Social Science Plaza, University of California, Irvine, CA 92697-5100, USA

## ARTICLE INFO

### Article history:

Received 28 February 2008

Received in revised form

1 July 2008

Accepted 1 July 2008

Available online 17 July 2008

### Keywords:

Signaling

Game theory

Selection–mutation dynamics

Bifurcation

## ABSTRACT

In signaling games the replicator dynamics does not almost always converge to states of perfect communication. A significant portion of the state space converges to components of Nash equilibria that characterize states of partial communication. Since these components consist of non-hyperbolic rest points, the significance of this result will depend on the dynamic behavior of specific perturbations of the replicator equations. In this paper we study selection–mutation dynamics of signaling games, which may be considered as one plausible perturbation of the replicator dynamics. We find that the long term behavior of the dynamics depends on the mutation rates of senders and receivers and on the relevance of communication.

© 2008 Elsevier Ltd. All rights reserved.

## 1. Introduction

Signaling enables coordinated activity on all levels of biological organization. Information transmission by signaling is found on the level of cells and microorganisms (Crespi, 2001; England et al., 1999), in animal communication (Maynard Smith and Harper, 2003) and, of course, in human language (Nowak and Krakauer, 1999). Mathematical models of signaling may help us understand the structure of signaling interactions and possible explanations for the emergence of successful information transmission.

Signaling poses two kinds of problems. One is concerned with possible conflicts of interest between the sender and the receiver, leading to the problem of honest signaling (Bergstrom and Lachmann, 1998; Grafen, 1990; Maynard Smith and Harper, 2003; Spence, 1973). If the interests of sender and receiver are aligned, other problems arise, which can be characterized in terms of signaling games (Lewis, 1969). In such games, the sender and the receiver may fail to communicate because of an uncoordinated use of signals. The strategies of a signaling game can be introduced by using the sets  $E$  and  $S$ , where  $|E| = |S| = n$ .  $E$  is the set of events (states) and  $S$  is the set of signals. A (pure) strategy of the sender can be represented as an  $n \times n$  matrix  $P$  with each row containing exactly one 1, the other entries being zero. Thus if  $p_{ij} = 1$  then the sender sends signal  $j$  given that state  $i$  has occurred. Similarly, a strategy of the receiver will be represented by an  $n \times n$  matrix  $Q$  with  $q_{ij} = 1$  for some  $j$  and  $q_{ik} = 0$  for all  $k \neq j$ . If  $q_{ij} = 1$

then the receiver associates event  $j$  with signal  $i$ . It is assumed that the sender and the receiver get the same payoff 1 if  $p_{ij} = 1 = q_{ji}$  given that state  $i$  has occurred. If  $(P, Q)$  is a profile of strategies, then the overall payoff for the sender and the receiver is given by

$$\frac{1}{n} \sum_{ij} p_{ij} q_{ji} = \frac{1}{n} \text{tr}(PQ) \quad (1)$$

(see Trapa and Nowak, 2000). Note the presumption that every state is weighed equally for computing the payoffs. More generally, the payoff resulting from strategy profile  $(P, Q)$  is given by

$$\sum_i w_i \sum_j p_{ij} q_{ji} = \text{tr}(WPQ), \quad (2)$$

with  $W = \text{diag}(w_1, \dots, w_n)$ , where  $w_i$  is the weight associated with event  $i$ ,  $w_i \geq 0$  and  $\sum_i w_i = 1$  for  $i = 1, \dots, n$ .  $w_i$  will in general depend on the probability of event  $i$  and the importance attached to communicating  $i$ .

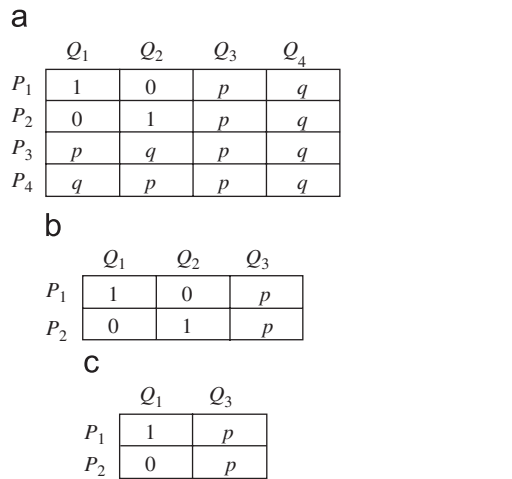
For instance, a binary signaling game ( $n = 2$ ) has four strategies for each player (see Fig. 1a for the payoff matrix):

$$P_1 = Q_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P_2 = Q_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ P_3 = Q_3 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad P_4 = Q_4 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \quad (3)$$

Signaling games have been studied extensively using the one-population replicator dynamics where each individual can be a sender or a receiver (Huttegger, 2007; Komarova and Niyogi, 2004; Nowak and Krakauer, 1999; Pawlowitsch, 2008; Skyrms,

\* Corresponding author.

E-mail addresses: [josef.hofbauer@univie.ac.at](mailto:josef.hofbauer@univie.ac.at) (J. Hofbauer), [simon.huttegger@kli.ac.at](mailto:simon.huttegger@kli.ac.at), [shuttegger@gmail.com](mailto:shuttegger@gmail.com) (S.M. Huttegger).



**Fig. 1.** Payoffs in a binary signaling game: (a) the full game. If  $p = q = \frac{1}{2}$  then the signaling game has uniform weights. If  $p \neq q$  then the signaling game has non-uniform weights. (b) Truncated binary signaling game. (c) Truncated game where only one signaling system is present.

1996). We will instead focus on the two-population replicator dynamics

$$\dot{x}_i = x_i((A\mathbf{y})_i - \mathbf{x} \cdot A\mathbf{y}), \tag{4a}$$

$$\dot{y}_j = y_j((B\mathbf{x})_j - \mathbf{y} \cdot B\mathbf{x}), \tag{4b}$$

where  $A$  is the  $m \times m$  payoff matrix of the sender,  $m = n^n$ , and  $B = A^T$  is the payoff matrix of the receiver (T denotes the transpose of a matrix). Dynamics (4) lives on  $S_m \times S_m$ , where  $S_m$  is the unit-simplex in  $\mathbb{R}^m$ . Studying (4) rather than the one-population model on  $S_{m^2}$  has the advantage of considerably reducing the dimension of the state space. Moreover, the dynamics of the two-population model can be imbedded in the state space of the one-population model, inducing trajectories on an invariant manifold which is equivalent to the Wright manifold of two-locus genetic models (see Section 3.4.1 of Cressman, 2003).

The most relevant results for the asymptotic properties of the dynamics of signaling games have been established independently by Huttegger (2007) and by Pawlowitsch (2008). The results are formulated for the one-population model, but they carry over to the two-population setting in a straightforward way. The only asymptotically stable states are given by the strict Nash equilibria of the signaling game. Strict Nash equilibria are signaling systems (this term was introduced by Lewis, 1969).  $(P, Q)$  is a signaling system iff  $P$  is a permutation matrix and  $Q = P^T$ .  $(P_1, Q_1)$  and  $(P_2, Q_2)$  in (3) are examples of signaling systems. For  $n$  signals there are  $n!$  signaling systems.

Although signaling systems are asymptotically stable for (4), it is in general not true that the set of signaling systems attracts almost all solutions of dynamics (4) (in the sense that the set of solutions which do not converge to a signaling system have Lebesgue measure zero). Consider

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

where  $0 \leq \lambda, \mu \leq 1$ .  $(\bar{P}, \bar{Q})$  is a Nash equilibrium of the signaling game for  $0 \leq \lambda, \mu \leq 1$ . This defines a two-dimensional linear manifold  $M$  of rest points. Using techniques from center manifold theory (Carr, 1981) one can show that the relative interior of  $M$  attracts an open set of the state space while parts of the boundary are unstable (for details see Huttegger, 2007).

This result generalizes to signaling games with more than three signals. The average payoff on manifolds of rest points such

as  $M$  can get arbitrarily small as  $n$  grows. Consider  $(\bar{P}, \bar{Q})$  as given by

$$\begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_{n-1} & 0 \\ 0 & 0 & \dots & 0 & 1 \\ \vdots & & & \vdots & \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \mu_1 & \dots & \mu_{n-1} \end{pmatrix},$$

with  $0 \leq \lambda_i, \mu_j \leq 1$  for all  $i, j$ . This set of profiles shares the properties of (5) (Pawlowitsch, 2008). But  $\text{tr}(\bar{P}, \bar{Q}) = \sum_i \lambda_i + \sum_j \mu_j = 2$  irrespective of  $n$ .

What is the significance of this result? The existence of a linear manifold of rest points like  $M$  implies that the dynamics is degenerate and not structurally stable (Guckenheimer and Holmes, 1983). This means that if dynamics (4) is slightly perturbed, then the qualitative behavior of the dynamics near  $M$  will change. Some perturbations might lead to a rest point close to  $M$  that is asymptotically stable or linearly unstable. Or the rest points of  $M$  might disappear altogether. One way to assess the significance of the convergence of the replicator dynamics to manifolds like  $M$  consists in looking at plausible perturbations of this dynamics or at other evolutionary dynamics. The second approach was taken by Pawlowitsch (2007), who considers signaling games under a frequency-dependent Moran process in a finite population. Our approach stays closer to the replicator dynamics, since we shall study signaling games under a perturbation of the replicator equations by small mutation terms.

## 2. Selection–mutation dynamics

One plausible perturbation of the replicator equations is selection–mutation dynamics. This dynamics is studied in Bürger (2000), Hofbauer (1985), Hofbauer and Sigmund (1998). In the context of language games (in particular, the evolution of universal grammar) a general selection–mutation dynamics was introduced by Nowak (2000) and Nowak et al. (2002). Eqs. (6) below can be considered as a limiting case of the latter model (cf. Hofbauer and Sigmund, 1998, Section 20.1). Focusing on the limiting case (6) enhances the tractability of the problems associated with the dynamics of signaling games considerably.

The heuristics behind selection–mutation dynamics is that the change in the relative frequency of one type is not only given by selection but also by a term which describes the rate of mutation from one type to another; to be more specific, with a certain probability an offspring is a mutant, and in this case each type is equally likely to mutate into any other type. Then the two-population selection–mutation dynamics is given by

$$\dot{x}_i = x_i((A\mathbf{y})_i - \mathbf{x} \cdot A\mathbf{y}) + \varepsilon(1 - mx_i), \tag{6a}$$

$$\dot{y}_j = y_j((B\mathbf{x})_j - \mathbf{y} \cdot B\mathbf{x}) + \delta(1 - my_j), \tag{6b}$$

where  $m = n^n$ , for possibly different mutation rates  $\varepsilon, \delta \geq 0$ . For  $\varepsilon = \delta = 0$  dynamics (6) reduces to the replicator equation (4). In a biological context  $\varepsilon$  and  $\delta$  can be interpreted as uniform mutation rates. In a non-biological context  $\varepsilon$  and  $\delta$  can be understood as rates of experimentation or as the rates of mistakes. As we shall show below, the ratio  $\delta/\varepsilon$  plays a particularly important role in determining the evolutionary outcome of a signaling game.

System (6) shares some properties with (4). In particular, (6) is still a gradient system with respect to the Shahshahani metric. The potential function for the one-population analogue of (6) has been found by Hofbauer (1985). By an analogous argument it can be shown that

$$V(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot A\mathbf{y} + \varepsilon \sum_{i=1}^m \log x_i + \delta \sum_{j=1}^m \log y_j \tag{7}$$

is the potential function for (6). In particular,

$$\dot{V} = \sum_i \frac{\dot{x}_i^2}{x_i} + \sum_j \frac{\dot{y}_j^2}{y_j} \geq 0. \tag{8}$$

The first term of  $V$  is just the mean fitness in the sender population and the receiver population (which must be the same since  $\mathbf{x} \cdot \mathbf{A}\mathbf{y} = \mathbf{y} \cdot \mathbf{A}^T \mathbf{x} = \mathbf{y} \cdot \mathbf{B}\mathbf{x}$ ). From the logarithmic terms of  $V$  it is clear that as  $x_i \rightarrow 0$  or  $y_j \rightarrow 0$ ,  $V \rightarrow -\infty$ . This implies that there can be no rest points on the boundary of  $S_m \times S_m$ . The increase of the potential function along solutions (8) excludes cyclic behavior of the dynamics (6). In addition, it implies that all orbits converge to the set of rest points of (6) (Hofbauer and Sigmund, 1998; Bürger, 2000).

Before we turn to the special case of two signals, let us state two simple but important facts about the location of rest points of (6) for small mutation rates (compare the proof of Theorem 13.4.1 in Hofbauer and Sigmund, 1998). To keep things simpler, we will assume in the following that  $\delta = \rho\varepsilon$ , with  $\rho > 0$  fixed, while  $\varepsilon$  goes to 0.

First, if  $(\mathbf{p}(\varepsilon), \mathbf{q}(\varepsilon))$  is a rest point of (6) then

$$\begin{aligned} (A\mathbf{q}(\varepsilon))_i - \mathbf{p}(\varepsilon) \cdot A\mathbf{q}(\varepsilon) - m\varepsilon &= -\frac{\varepsilon}{p_i(\varepsilon)} < 0, \\ (B\mathbf{p}(\varepsilon))_j - \mathbf{q}(\varepsilon) \cdot B\mathbf{p}(\varepsilon) - m\rho\varepsilon &= -\frac{\rho\varepsilon}{q_j(\varepsilon)} < 0. \end{aligned}$$

This implies that as  $\varepsilon \rightarrow 0$  every accumulation point  $(\mathbf{p}, \mathbf{q})$  of  $(\mathbf{p}(\varepsilon), \mathbf{q}(\varepsilon))$  must satisfy

$$(A\mathbf{q})_i - \mathbf{p} \cdot A\mathbf{q} \leq 0 \quad \text{and} \quad (B\mathbf{p})_j - \mathbf{q} \cdot B\mathbf{p} \leq 0,$$

i.e.  $(\mathbf{p}, \mathbf{q})$  must be a Nash equilibrium of the signaling game. This allows us to conclude that *there are no rest points of (6) close to rest points of (4) which are not Nash equilibria of the underlying game.*

The second general fact concerns the *existence, uniqueness and asymptotic stability of perturbed signaling systems.* To see this set

$$\begin{aligned} F_i(\mathbf{x}, \mathbf{y}, \varepsilon) &= x_i((A\mathbf{y})_i - \mathbf{x} \cdot A\mathbf{y}) + \varepsilon(1 - mx_i), \\ G_j(\mathbf{x}, \mathbf{y}, \varepsilon) &= y_j((B\mathbf{x})_j - \mathbf{y} \cdot B\mathbf{x}) + \rho\varepsilon(1 - my_j). \end{aligned}$$

Let  $(\mathbf{p}, \mathbf{q})$  be a signaling system. Let

$$H = (F_1, \dots, F_m, G_1, \dots, G_m).$$

$H$  is defined on an open set containing  $(\mathbf{p}, \mathbf{q}, 0)$  and  $H(\mathbf{p}, \mathbf{q}, 0) = 0$ . The Jacobian matrix of  $H$  with respect to  $\mathbf{x}$  and  $\mathbf{y}$  evaluated at  $(\mathbf{p}, \mathbf{q}, 0)$  is invertible. Hence, by the implicit function theorem, there exists a unique smooth function  $\phi : (-\varepsilon_0, \varepsilon_0) \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ ,  $\phi(\varepsilon) = (\mathbf{p}(\varepsilon), \mathbf{q}(\varepsilon))$  such that

$$\phi(0) = (\mathbf{p}, \mathbf{q}) \quad \text{and} \quad H(\phi(\varepsilon), \varepsilon) = 0.$$

This establishes the existence of unique equilibria  $\phi(\varepsilon) \in S_m \times S_m$  for small  $\varepsilon > 0$ . Since  $\phi(\varepsilon)$  is close to a signaling system of the signaling game, we will call it a *perturbed signaling system.*

This argument works more generally for regular Nash equilibria of a game; compare the proof of Theorem 13.4.1 in Hofbauer and Sigmund (1998), or Section 2.2 in Bürger and Hofbauer (1994). However, since regular equilibria are isolated, the only regular equilibria in signaling games are signaling systems.

The asymptotic stability of perturbed signaling systems follows from the fact that the entries of the Jacobian matrix for (6) are continuous in  $\varepsilon$ . Therefore, each of the  $n!$  perturbed signaling systems is a local maximizer of the potential function (7).

There are no similar general statements about perturbations of equilibria other than signaling systems. But we can use degree theory and Morse index arguments to infer the existence of further rest points of (6). We will use these tools for the case of two signals. The case of three or more signals needs more space

and will be treated elsewhere. We would like to emphasize, however, that an analysis of the case of two signals is somewhat more important than an analysis of more general signaling games. Binary signaling games represent the most basic kind of signaling; we expect communication to evolve in binary signaling games given that it evolves at all.

Binary signaling games may be considered in their own right for other reasons as well. There are examples of binary signaling systems in living organisms (e.g. female fruit flies, *Drosophila subobscura*, send signals to approaching males when they have already mated, Maynard Smith and Harper, 2003). And third, as we shall see, the behavior of selection–mutation dynamics close to manifolds such as  $M$  can to some extent already be studied for  $n = 2$ .

### 3. Results

#### 3.1. Uniform weights

The case  $w_1 = w_2 = \frac{1}{2}$  has been numerically analyzed by Huttegger et al. (2008). Here, we present a full analytical treatment for illustrative purposes, since the emergence of communication in this kind of signaling game poses no problem to the replicator dynamics (see Huttegger, 2007, Theorem 9).

Consider the signaling game with strategies as given by (3) and payoff function (1). Let  $x_i$  be the relative frequency of  $P_i$  and  $y_j$  be the relative frequency of  $Q_j$ . The Nash equilibria are given by the two signaling systems  $(P_i, Q_i)$ ,  $i = 1, 2$  and a component of Nash equilibria given by

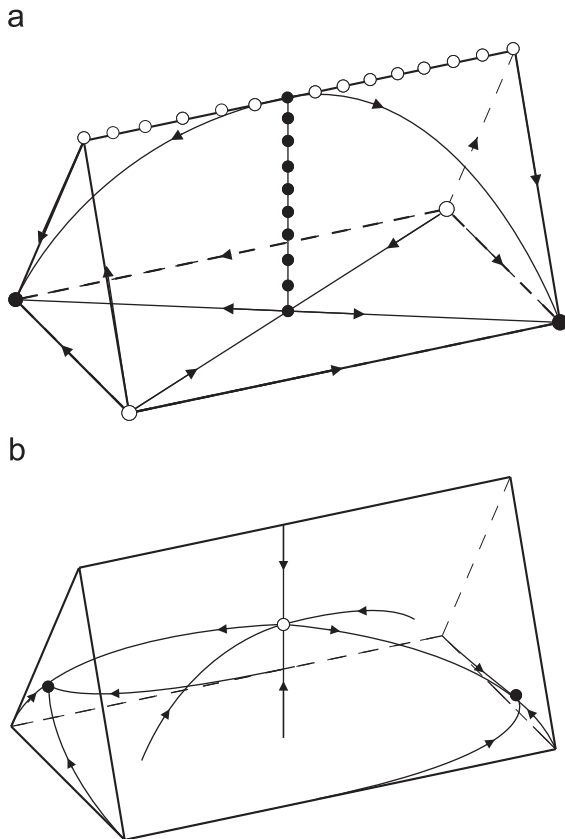
$$\begin{pmatrix} \lambda & 1 - \lambda \\ \lambda & 1 - \lambda \end{pmatrix}, \quad \begin{pmatrix} \mu & 1 - \mu \\ \mu & 1 - \mu \end{pmatrix}. \tag{9}$$

That these matrices describe Nash equilibria follows from Theorem 5.1 of Trapa and Nowak (2000). The matrix pair (9) defines a four-dimensional component of Nash equilibria given by the condition  $x_1 = x_2$  and  $y_1 = y_2$ . Let us denote this component by  $N$ .

To see that there are no other Nash equilibria, suppose first that  $P$  (or  $Q$ ) has a zero column. Then  $(P, Q)$  is a Nash equilibrium only if  $Q$  (or  $P$ ) is of form (9) for some  $0 \leq \mu \leq 1$  (or some  $0 \leq \lambda \leq 1$ ). If  $P$  (and  $Q$ ) has no zero column and if it is not one-to-one, then by Theorem 5.1 of Trapa and Nowak (2000)  $(P, Q)$  must be of form (9) in order to be a Nash equilibrium.

The second general fact for the dynamics of signaling games implies that, for small  $\varepsilon$ , there are two perturbed and asymptotically stable signaling systems for (6). The average payoff along  $N$  is  $\frac{1}{2}$ . The second and third terms of the potential function  $V$  attain their unique maximum at the barycenter  $\mathbf{b}$ . Hence if  $\varepsilon > 0$  the component  $N$  collapses into  $\mathbf{b}$ . Since there cannot be any other rest points of the perturbed dynamics,  $\mathbf{b}$  must be linearly unstable having one positive eigenvalue. This follows from the Morse inequalities (Milnor, 1963): a gradient system with two asymptotically stable rest points must have at least one rest point having one positive eigenvalue. An illustration of the dynamics is shown in Fig. 2. (This is the dynamics of the truncated game of Fig. 1b.)

Hence we see that from almost all initial conditions solutions of (6) converge to one of the perturbed signaling systems. For uniform weights this result is already true for the unperturbed dynamics (4). However, as we shall see in the next section, in the case of non-uniform weights dynamics (4) and (6) are indeed qualitatively different.



**Fig. 2.** State space and dynamics for the truncated game with  $p = \frac{1}{2}$ . (a) The selection–mutation dynamics for  $\varepsilon = \delta = 0$ . Black dots indicate Nash equilibria and white dots indicate rest points which are not Nash equilibria. There is one component of unstable Nash equilibria and two asymptotically stable signaling system. (b) Now  $\varepsilon, \delta > 0$ . The component of Nash equilibria collapses into the barycenter, which is linearly unstable. The two signaling systems become perturbed rest points which remain asymptotically stable.

**3.2. Non-uniform weights: truncated game**

We now consider signaling games with two signals and a non-uniform weight matrix  $W$ . Suppose without loss of generality that  $w_1 > w_2 = 1 - w_1$ . For a more concise notation we set  $w_1 = p$  below.

The structure of the equilibrium components is different in the case of a non-uniform weight matrix. To simplify matters we will first look at a truncated version of the game (shown in Fig. 1b) with pure strategies  $P_1, P_2, Q_1, Q_2$  and  $Q_3$ . Here, in the spirit of classical game theory, we omit the strictly dominated strategy  $Q_4$ , and the strategies  $P_3, P_4$  which are equivalent to mixtures of  $P_1$  and  $P_2$ . We shall see in the last section that the results obtained for the truncated game are qualitatively the same as the results for the full game.

$(P_1, Q_1)$  and  $(P_2, Q_2)$  are signaling systems. There is another component of rest points given by

$$\begin{pmatrix} \alpha & 1 - \alpha \\ 1 - \alpha & \alpha \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad 0 \leq \alpha \leq 1. \tag{10}$$

Not all points along this line are Nash equilibria, however. But (10) contains a component of Nash equilibria  $K$ , given by  $q \leq \alpha \leq p$ . The size of  $K$  relative to (10) depends on  $p$ . The average payoff along  $K$  is  $p$ .

For the unperturbed replicator dynamics (4) the two signaling systems are asymptotically stable. The component  $K$  is not asymptotically stable, but the points in the relative interior of  $K$  are Lyapunov stable. Moreover, like the points in the relative

interior of the component  $M$  (see 5),  $K$  attracts an open set of initial conditions under (4) (a precise argument proving this conclusion is given in Huttegger, 2007). Note that a component like  $K$  does not exist for the uniform case according to the remarks at the end of the previous section.

What happens under the perturbed dynamics (6) for the truncated game? We know that, like in the uniform case, there exist two perturbed rest points of (6) close to the signaling systems for sufficiently small  $\varepsilon$ . Both rest points are asymptotically stable. Index or degree theory (see e.g. Section 13.4 in Hofbauer and Sigmund, 1998) tells us that there is at least one further rest point. From the analysis above it is clear that this rest point must be near  $K$ . By using a perturbation expansion we will study its stability properties.

The perturbed dynamics for the truncated game is given by

$$\begin{aligned} \dot{x}_1 &= x_1(y_1 + py_3 - \bar{a}) + \varepsilon(1 - 2x_1), \\ \dot{x}_2 &= x_2(y_2 + py_3 - \bar{a}) + \varepsilon(1 - 2x_2) \end{aligned} \tag{11a}$$

and

$$\begin{aligned} \dot{y}_1 &= y_1(x_1 - \bar{a}) + \delta(1 - 3y_1), \\ \dot{y}_2 &= y_2(x_2 - \bar{a}) + \delta(1 - 3y_2), \\ \dot{y}_3 &= y_3(p - \bar{a}) + \delta(1 - 3y_3), \end{aligned} \tag{11b}$$

where  $\bar{a} = x_1y_1 + x_2y_2 + py_3$ . We look for a symmetric rest point satisfying  $x_1 = x_2 = \frac{1}{2}$ . Setting (11a) to zero shows that  $y_1 = y_2$ , and thus  $\bar{a} = y_1 + py_3$ . By (11b) a rest point of the system must therefore satisfy

$$y_1(3\delta + p - \frac{1}{2} + y_1(1 - 2p)) = \delta.$$

A Taylor expansion in terms of  $\delta$  leads to the solution

$$\hat{y}_1 = \hat{y}_2 = \frac{2\delta}{2p - 1} + O(\delta^2), \quad \hat{y}_3 = 1 - 2\hat{y}_1.$$

In order to study the stability properties of this rest point we introduce two new variables  $\xi, \eta$ :

$$x_1 = \frac{1}{2} + \xi \quad \text{and} \quad \eta = y_1 - y_2.$$

Then (11) can be written as

$$\begin{aligned} \dot{\xi} &= (\frac{1}{4} - \xi^2)\eta - 2\varepsilon\xi, \\ \dot{\eta} &= \xi(1 - y_3) - \eta(3\delta + (p - \frac{1}{2})y_3) - \xi\eta^2, \\ \dot{y}_3 &= (p - \frac{1}{2})y_3(1 - y_3) - y_3\xi\eta + \delta(1 - 3y_3). \end{aligned} \tag{12}$$

Linearizing (12) at the rest point  $\xi = \eta = 0, y_3 = \hat{y}_3$  yields

$$\begin{aligned} \dot{\xi} &= \frac{1}{4}\eta - 2\varepsilon\xi, \\ \dot{\eta} &= \xi(1 - \hat{y}_3) - \eta(3\delta + (p - \frac{1}{2})\hat{y}_3), \end{aligned} \tag{13}$$

where we can ignore the third equation, since  $\dot{y}_3$  contains no linear terms in  $\xi, \eta$ . The Jacobian matrix of (13) at the rest point is given by

$$J = \begin{pmatrix} -2\varepsilon & 1/4 \\ 4\delta/(2p - 1) & -\delta - (p - 1/2) \end{pmatrix},$$

up to higher order terms in  $\varepsilon$  and  $\delta$ . Then

$$\det(J) = \varepsilon(2p - 1) - \frac{\delta}{2p - 1}.$$

This is less than zero if and only if

$$\frac{\delta}{\varepsilon} > (2p - 1)^2. \tag{14}$$

In this case (which includes  $\delta = \varepsilon$ ), the rest point close to  $K$  is a saddle. If, on the other hand, the inequality sign in (14) is reversed, then the rest point with  $x_1 = x_2 = \frac{1}{2}$  will be a sink since the trace of  $J$  is negative. Index theory implies that there must be at least two further rest points, each one having one positive eigenvalue.

A perturbation expansion shows that one solution is approximately given by  $q < \bar{x}_1 < \bar{x}_2 < p$  and

$$\bar{y}_1 = \frac{\delta}{p - \bar{x}_1} + O(\delta^2), \quad \bar{y}_2 = \frac{\delta}{p - \bar{x}_2} + O(\delta^2),$$

where  $(\bar{x}_1, \bar{x}_2) \times (0, 0, 1)$  is the limit of perturbed equilibria as  $\varepsilon \rightarrow 0$ . The second solution is obtained by letting  $q < \bar{x}_2 < \bar{x}_1 < p$ .

Eq. (12) implies the following conditions for rest points close to  $K$ :

$$\eta = \frac{8\varepsilon\xi}{1 - 4\xi^2},$$

$$\xi = \frac{\eta(3\delta + (p - 1/2)y_3)}{1 - y_3 - \eta^2}.$$

Inserting  $\bar{y}_3 = 1 - \bar{y}_1 - \bar{y}_2$  and the expression for  $\eta$  into the second equation leads to the result that there exists no rest point other than the one for  $\xi = 0$  close to  $K$  whenever  $\delta/\varepsilon \geq (2p - 1)^2$ . Therefore, if  $\delta/\varepsilon \geq (2p - 1)^2$ , then dynamics (6) converges to a signaling system from almost all initial conditions.

It is instructive to look at the boundary face which is spanned by  $P_1, P_2, Q_1$  and  $Q_3$  (cf. the payoff matrix in Fig. 1c). The selection–mutation dynamics of this  $2 \times 2$  game can be written as

$$\dot{x}_1 = x_1(y_1 + py_3 - \bar{a}) + \varepsilon(1 - 2x_1),$$

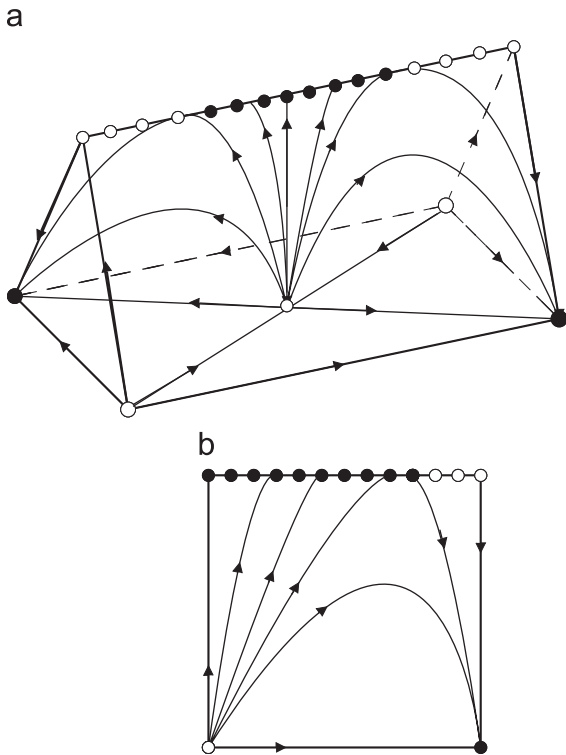
$$\dot{x}_2 = x_2(py_3 - \bar{a}) + \varepsilon(1 - 2x_2) \tag{15a}$$

and

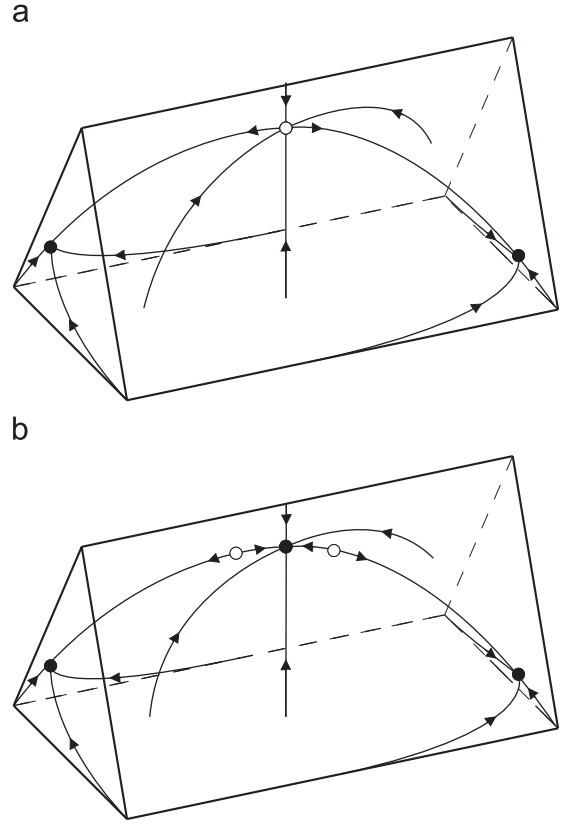
$$\dot{y}_1 = y_1(x_1 - \bar{a}) + \delta(1 - 2y_1),$$

$$\dot{y}_3 = y_3(p - \bar{a}) + \delta(1 - 2y_3), \tag{15b}$$

where  $\bar{a} = x_1y_1 + py_3$ . The unperturbed dynamics of (15) with  $\varepsilon, \delta = 0$  is illustrated in Fig. 3b. This truncated game has a



**Fig. 3.** State space and replicator dynamics (4) of the truncated game with  $p > \frac{1}{2}$ . Black dots indicate Nash equilibria and white dots indicate rest points which are not Nash equilibria. (a) A component of rest points attracts a set of positive measure. (b) The replicator dynamics restricted to the game where only  $P_1, P_2, Q_1$  and  $Q_3$  are present.



**Fig. 4.** State space and perturbed dynamics with  $\varepsilon, \delta > 0$  for the truncated game with  $p > \frac{1}{2}$ . Black dots indicate sinks and white dots indicate saddles. (a) If  $\varepsilon, \delta > 0$  and  $\delta/\varepsilon > (2p - 1)^2$ , then there exists one rest point close to  $K$  which is linearly unstable. (b) If  $\delta/\varepsilon < (2p - 1)^2$ , then there exist three rest points close to  $K$ . One of them is asymptotically stable while the other two are linearly unstable.

component of Nash equilibria given by  $0 \leq x_1 \leq p$  and  $y_1 = 0$  which attracts an open set of initial conditions, but is not stable. A straightforward calculation shows that for small  $\varepsilon, \delta > 0$ , a pair of rest points (a sink and a saddle) exists near this component if

$$\frac{\delta}{\varepsilon} < 1 - 2\sqrt{p(1 - p)}. \tag{16}$$

If the inequality sign in (16) is reversed then there are no perturbed equilibria near this component. Then all orbits converge to the perturbed signaling system.

This truncated  $2 \times 2$  game is strategically equivalent to the Chain–Store game and the ultimatum minigame studied by Gale et al. (1995) and Binmore and Samuelson (1999). Their results on perturbed game dynamics also emphasize the role of the ratio  $\delta/\varepsilon$  in determining what happens close to the component of Nash equilibria (see Fig. 4).

### 3.3. Non-uniform weights: full game

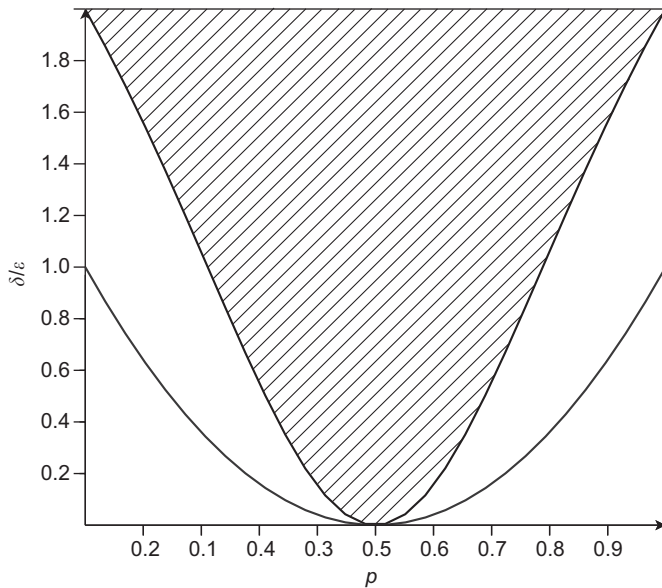
We now consider the full game of Fig. 1a. The selection–mutation dynamics of this game is given by

$$\dot{x}_1 = x_1(y_1 + py_3 + qy_4 - \bar{a}) + \varepsilon(1 - 4x_1),$$

$$\dot{x}_2 = x_2(y_2 + py_3 + qy_4 - \bar{a}) + \varepsilon(1 - 4x_2),$$

$$\dot{x}_3 = x_3(p(y_1 + y_3) + q(y_2 + y_4) - \bar{a}) + \varepsilon(1 - 4x_3),$$

$$\dot{x}_4 = x_4(p(y_2 + y_3) + q(y_1 + y_4) - \bar{a}) + \varepsilon(1 - 4x_4) \tag{17a}$$



**Fig. 5.** The graphs defining the regions of stability of the symmetric perturbed equilibrium corresponding to 14 (below) and 18 (above). The region of stability (i.e. the region above the graph) changes as we move from the truncated game to the full game, but there is no qualitative difference.

and

$$\begin{aligned}\dot{y}_1 &= y_1(x_1 + px_3 + qx_4 - \bar{a}) + \delta(1 - 4y_1), \\ \dot{y}_2 &= y_2(x_2 + qx_3 + px_4 - \bar{a}) + \delta(1 - 4y_2), \\ \dot{y}_3 &= y_3(p - \bar{a}) + \delta(1 - 4y_3), \\ \dot{y}_4 &= y_4(q - \bar{a}) + \delta(1 - 4y_4),\end{aligned}\quad (17b)$$

where  $p > q$  and

$$\bar{a} = x_1y_1 + x_2y_2 + x_3(py_1 + qy_2) + x_4(qy_1 + py_2) + py_3 + qy_4.$$

Using a perturbation expansion we find a symmetric perturbed equilibrium  $\mathbf{p}$  with  $x_i = \frac{1}{4}$  for  $i = 1, \dots, 4$ ,

$$\hat{y}_1 = \hat{y}_2 = \frac{2\delta}{2p-1} + O(\delta^2), \quad \hat{y}_3 = \frac{\delta}{p-q} + O(\delta^2)$$

and  $\hat{y}_4 = 1 - 2\hat{y}_1 - \hat{y}_3$ . We now use a computer algebra system to compute the Jacobian matrix of (17) at  $\mathbf{p}$  and find that  $\mathbf{p}$  is a sink for small  $\varepsilon, \delta > 0$  if

$$\frac{\delta}{\varepsilon} < 2 \frac{(1-2p)^2}{1-2p+2p^2}. \quad (18)$$

A calculation of the principal minors of the Jacobian matrix shows that in this case all eigenvalues are negative. Index theory again implies the existence of two more equilibria which are saddles. If the inequality in (18) is reversed then  $\mathbf{p}$  is a saddle (with one positive eigenvalue). Numerical simulations suggest that no further equilibria exist in this case and hence almost all orbits converge to the perturbed signaling systems. Hence, qualitatively, we obtain the same results as for the truncated game. However, the region of stability is different. For example, for  $\delta = \varepsilon$ , (18) holds for some  $p$  whereas in the truncated game, (14) holds. Compare the graphs corresponding to (14) and (18) in Fig. 5.

#### 4. Conclusion

We have shown that whether communication emerges in binary signaling games depends on the ratio  $\delta/\varepsilon$  and the payoff parameter  $p$ . After some thought, these results appear to be quite intuitive. If the receiver experiments much less frequently than

the sender (small  $\delta/\varepsilon$ ), the players might end up not communicating. As an example, one may think of two species where the receiver species has a much longer generation time than the sender species. This model might also be interpreted as a two-locus genetics model with possibly differing mutation rates. In all these cases, condition (14) or (18) relates  $p$  to the ratio  $\delta/\varepsilon$ . For given  $p$ , if  $\delta/\varepsilon$  is too small, then the receiver population is not responsive enough to the experimentation of the sender population.

The payoff parameter  $p$  measures the relevance of communication (see Nowak et al., 2002, Box 2). If  $p$  is close to  $\frac{1}{2}$ , communicating the event that has occurred is most important. As  $p$  goes to 1 (or to 0), communication becomes less important. In the limit  $p = 1$ , the ratio  $\delta/\varepsilon$  has to be large (at least 2) in order to get communication in the full binary signaling game. Thus, the range of  $\delta/\varepsilon$  that guarantees convergence to a perturbed signaling system reflects the relevance of communication.

As to the broader aspects of the theory of signaling games, our analysis shows that there is no clear-cut answer to the question whether selection–mutation dynamics is a mechanism for the evolution of optimality in sender–receiver games. The answer to this question depends on the parameters  $\delta, \varepsilon$  and  $p$ . The relations of these parameters as expressed by (14) and (18) determines the outcome of the evolutionary dynamics. Moreover, these relations have a clear interpretation in terms of responsiveness of receivers relative to the rate of experimentation of senders ( $\delta/\varepsilon$ ) and relevance of communication ( $p$ ). Thus, perfect communication is likely to evolve if communication is relevant and if receiver experimentation is responsive with respect to sender experimentation.

#### Acknowledgment

S. Huttegger's research was supported by the Konrad Lorenz Institute for Evolution and Cognition Research.

#### References

- Bergstrom, C., Lachmann, M., 1998. Signalling among relatives III. Talk is cheap. *Proc. Natl. Acad. Sci. USA* 95, 5100–5105.
- Binmore, K., Samuelson, L., 1999. Evolutionary drift and equilibrium selection. *Rev. Econ. Stud.* 66, 363–393.
- Bürger, R., 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, Chichester.
- Bürger, R., Hofbauer, J., 1994. Mutation load and mutation–selection–balance in quantitative genetic traits. *J. Math. Biol.* 32, 193–218.
- Carr, J., 1981. *Applications of Centre Manifold Theory*. Springer, NY.
- Crespi, 2001. The evolution of social behavior in microorganisms. *Trends Ecol. Evol.* 16, 178–183.
- Cressman, R., 2003. *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge, MA.
- England, R.R., Hobbs, G., Bainton, N.J., Roberts, D., 1999. *Microbial Signaling and Communication*. Cambridge University Press, Cambridge, UK.
- Gale, J., Binmore, K.G., Samuelson, L., 1995. Learning to be imperfect: the ultimatum game. *Games Econ. Behav.* 8, 56–90.
- Grafen, A., 1990. Biological signals as handicaps. *J. Theor. Biol.* 144, 517–546.
- Guckenheimer, J., Holmes, P., 1983. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer, NY.
- Hofbauer, J., 1985. The selection mutation equation. *J. Math. Biol.* 23, 1985.
- Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK.
- Huttegger, S.M., 2007. Evolution and the explanation of meaning. *Philos. Sci.* 74, 1–27.
- Huttegger, S.M., Skyrms, B., Smead, R., Zollman, K.J.S., 2008. Evolutionary dynamics of Lewis signaling games. *Synthese*, forthcoming.
- Komarova, N.L., Niyogi, P., 2004. Optimizing the mutual intelligibility of linguistic agents in a shared world. *Artif. Intell.* 154, 1–42.
- Lewis, D., 1969. *Convention. A Philosophical Study*. Harvard University Press, Harvard.
- Maynard Smith, J., Harper, D., 2003. *Animal Signals*. Oxford University Press, Oxford.
- Milnor, J., 1963. *Morse Theory*. Princeton University Press, Princeton, NJ.

- Nowak, M.A., 2000. Evolutionary biology of language. *Philos. Trans. R. Soc. B* 355, 1615–1622.
- Nowak, M.A., Krakauer, D.C., 1999. The evolution of language. *Proc. Natl. Acad. Sci. USA* 96, 8028–8033.
- Nowak, M.A., Komarova, N.L., Niyogi, P., 2002. Computational and evolutionary aspects of language. *Nature* 417, 611–617.
- Pawlowitsch, C., 2007. Finite populations choose an optimal language. *J. Theor. Biol.* 249, 606–616.
- Pawlowitsch, C., 2008. Why evolution does not always lead to a signaling system. *Games Econ. Behav.* 63, 203–226.
- Skyrms, B., 1996. *Evolution of the Social Contract*. Cambridge University Press, Cambridge, UK.
- Spence, A.M., 1973. Job market signaling. *Q. J. Econ.* 87, 355–374.
- Trapa, P.E., Nowak, M.A., 2000. Nash equilibria for an evolutionary language game. *J. Math. Biol.* 41, 172–188.