Journal of Theoretical Biology 253 (2008) 131–141

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Evolutionary stability conditions for signaling games with costly signals

Gerhard Jäger *

University of Bielefeld, Faculty of Linguistics and Literature, PF 10 01 31, 33615 Bielefeld, Germany

ARTICLE INFO

Article history:
Received 30 July 2007
Received in revised form
18 February 2008
Accepted 19 February 2008
Available online 5 March 2008

Keywords:
Signaling games
Evolutionary game theory
Costly signaling
Evolutionary stability
Neutral stability

ABSTRACT

The paper investigates the class of signaling games with the following properties: (a) the interests of sender and receiver coincide, (b) different signals incur differential costs, and (c) different events (meanings/types) have different probabilities. Necessary and sufficient conditions are presented for a profile to be evolutionarily stable and neutrally stable, and for a set of profiles to be an evolutionarily stable set.

The main finding is that a profile belongs to some evolutionarily stable set if and only if a maximal number of events can be reliably communicated. Furthermore, it is shown that under the replicator dynamics, a set of states with a positive measure is attracted to "sub-optimal" equilibria that do not belong to any asymptotically stable set.

1. Introduction

In his book *Convention*, Lewis (1969) gave a game theoretic formalization of strategic communication. Lewis showed that a convention which guarantees successful communication can be self-reinforcing provided the interests of the communicators are sufficiently aligned. In game theoretic parlance, communication conventions are Nash equilibria. As the phenomenon of communication is of high relevance for many scientific disciplines, Lewis style *signaling games* and similar game theoretic models of communication received a great deal of attention since then (see for instance Spence, 1973; Crawford and Sobel, 1982 in economics, Grafen, 1990; Nowak and Krakauer, 1999; Hurd, 1995 in biology, Skyrms, 1996 in philosophy, Hurford, 1989; van Rooij, 2004 in linguistics and much subsequent work in all mentioned disciplines). The common theme of all these models can be summarized as follows:

- There are two players, the sender and the receiver.
- The sender has private information about an event that is unknown to the receiver. The event is chosen by nature according to a certain fixed probability distribution.
- The sender emits a signal which is revealed to the receiver.
- The receiver performs an action, and the choice of action may depend on the observed signal.
- The utilities of sender and receiver may depend on the event, the signal and the receiver's action.

Depending on the precise parameters, signaling games may have a multitude of equilibria. Therefore the question arises how a stable communication convention can be established. A promising route is to assume that such equilibria are the result of biological or cultural evolution. Under this perspective, communication conventions should be evolutionarily stable in the sense of evolutionary game theory.

Trapa and Nowak (2000) consider the class of signaling games where signaling is *costless* (i.e. the utility of sender and receiver does not depend on the emitted signal) and the interests of sender and receiver completely coincide. Furthermore, they assume that the actions of the receiver are isomorphic to the set of events. So the task of the receiver is essentially to guess the correct event. They also assume a uniform probability distribution over events. Under these conditions it turns out that the evolutionarily stable states (in the sense of Maynard Smith, 1982) are exactly those states where the sender strategy is a bijection from events to signals, and the receiver strategy is the inverse of the sender's strategy. This means that in an evolutionarily stable state, the receiver is always able to reliably infer the private information of the sender.[1]

Pawlowitsch (2007) investigates the same class of games, with the additional restriction that the number of events and signals must be identical. She shows that each such game has an infinite number of neutrally stable strategies (NSSs) (again in the sense of Maynard Smith, 1982) that are not evolutionarily stable. In these states, communication is not optimal because certain events

* Tel.: +49 521 106 3576; fax: +49 521 106 5844.
E-mail address: Gerhard.Jaeger@uni-bielefeld.de

[1] Similar results have also been obtained by Wärneryd (1993). Since he only considers pure strategies though, his results are perhaps less general.

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2008.02.039

cannot be reliably communicated. Perhaps surprisingly, these sub-optimal equilibria attract a set of states with a positive measure under the replicator dynamics. Natural selection alone thus does not necessarily lead to perfect communication.

In many naturally occurring signaling scenarios emitting or observing a signal may incur a cost to the players. Games with *costly signaling* have been studied extensively by economists (like Spence, 1973) and biologists (as Grafen, 1990) because costs may help to establish credibility in situations where the interests of sender and receiver are not completely aligned (an effect that is related to Zahavi's, 1975 famous *handicap principle*).

The idea that signals have differential costs is also a standard assumption in linguistics. It comes in many variants, including Grice's (1975) *Maxim of Manner*, Zipf's (1949) *Principle of Least Effort*, the concept of *markedness* in functional linguistics (see for instance Greenberg, 1966), etc. An evolutionary interpretation of the idea that high complexity of a signal may lead to low utility (i.e. low fitness under cultural evolution) can be traced back to the 19th century, as the following quotation from Charles Darwin shows:

> The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel. ... Max Müller has well remarked: 'A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their inherent virtue.' To these important causes of the survival of certain words, mere novelty and fashion may be added; for there is in the mind of man a strong love for slight changes in all things. The survival or preservation of certain favored words in the struggle for existence is natural selection. (Darwin, 1871:465f.)

van Rooij (2004) and Jäger (2007) formalize certain linguistic phenomena as signaling games with completely aligned interests of sender and receiver, costly signaling, and non-uniform probability distributions over events. The purpose of these works is to show that grammatical patterns that are typologically common or even universal among the languages of the world are evolutionarily stable.

The present paper studies the issue of evolutionary stability in this class of games (costly signaling and a non-uniform probability distribution over events) systematically. For simplicity's sake, the investigation is confined to games where any two events have pairwise different probabilities and any two signals incur pairwise different costs. This limitation seems legitimate as almost all parameter settings belong to this class.

In the subsequent sections, necessary and sufficient (static) conditions of neutral and evolutionary stability, as well as for a class of profiles to form an *evolutionarily stable set* (cf. Thomas, 1985), are developed. Briefly put, evolutionary stability always amounts to a bijective map between events and signals (such that the receiver's strategy is the inverse of the sender's strategy), possibly extended by non-communicable events or unused signals if the number of signals and events do not coincide. If the number of signals exceeds the number of events, the most expensive signals are never used by the sender. If, on the other hand, the number of events exceeds the number of signals, some events (not necessarily the least likely ones) are never inferred by the receiver.

In a neutrally stable state, the sender always plays a pure strategy that is the unique best response to the receiver's strategy. It is possible though that in such a state certain signals are never used by the sender, such that the best response of the receiver to the sender's strategy may be mixed. There are neutrally stable states that do not belong to an evolutionarily stable set. These are sub-optimal states where certain signals remain unused even though using them (and interpreting them appropriately) would increase the utility of both players. As in the class of games investigated by Pawlowitsch (2007), the set of sub-optimal neutrally stable states attracts a set of states with a positive measure under the replicator dynamics.

2. Setting the stage

A signaling game is an extensive form game between two persons, the sender and the receiver. The sender has some private information about an event e that is chosen by nature according to some fixed probability distribution from the set of events $\mathcal{E}$. The sender makes the first move by emitting a signal σ from some set $\mathcal{F}$ that can be observed by the receiver. The receiver in turn chooses some action a from a set $\mathcal{A}$. The utility of the players depend on e, σ and a.

In this paper I will study the conditions for evolutionary stability of a subclass of normalized and symmetrized signaling games. In particular, I restrict attention to games where

- $\mathcal{E}, \mathcal{F}$, and $\mathcal{A}$ are finite,
- $\mathcal{E} = \mathcal{A}$ (the receiver's action is to guess an event),
- there are functions c_1 and c_2 that assign to each signal σ certain numbers $c_1(\sigma)$ and $c_2(\sigma)$ (intuitively, $c_1(\sigma)$ represents the sender's benefit/cost of using σ, and likewise c_2 is the receiver's cost function), and
- the extensive form utility functions are

  $$u_i(e, \sigma, a) = \delta_{e,a} + c_\rho(\sigma)$$

  (with $\rho \in \{1, 2\}$).

(The δ used in the last condition is the Kronecker function, i.e. $\delta_{e,a} = 1$ if $e = a$, and $\delta_{e,a} = 0$ otherwise.)

The last condition states that the interests of the players are identical except for the costs that the transmitted signal incurs. Also, the condition entails that the cost of sending/receiving a certain signal only depends on the signal itself, not on its intended or inferred interpretation.

The set of pure sender strategies is the set of functions $\mathcal{E} \mapsto \mathcal{F}$, and the set of pure receiver strategies is the set of functions $\mathcal{F} \mapsto \mathcal{E}$. Let us assume there are n different events and m different signals. Then each pure sender strategy can be represented by an $n \times m$-matrix that has one 1 per row and 0 otherwise. Likewise, a receiver strategy corresponds to an $m \times n$-matrix of this form.[2] Let $\mathcal{S}$ be the set of sender strategies, and $\mathcal{R}$ the set of receiver strategies (in matrix notation).

The matrix notation can easily be extended to mixed strategies. Let x be a mixed sender strategy, i.e. a probability distribution over $\mathcal{S}$. We then define

$$S^x = \sum_{S \in \mathcal{S}} x(S) S.$$

Mixed strategy matrices for receiver strategies are defined likewise. Note that S^x (R^y) is always a stochastic matrix (i.e. each row sums up to 1) if x (y) is a probability distribution.

Nature's probability distribution over the set of events $\mathcal{E}$ can be represented by a probability vector $\vec{e}$ of length n. The costs of the various signals can be represented by vectors $\vec{c}_1$ (for the sender) and $\vec{c}_2$ (for the receiver) of length m. Intuitively all $c_{\rho,i}$ should be negative or zero because costs are negative payoff, but this

[2] See for instance Trapa and Nowak (2000) on the matrix representation of signaling games.

restriction has no impact on the properties of the game and can thus be ignored.

The utility function over pure strategies can be represented in the following way (where $\rho \in \{1,2\}$):

$$u_\rho(S,R) = \sum_i e_i \times \sum_j s_{ij}(r_{ji} + c_{\rho,j}).$$

Note that this can be rewritten as

$$u_\rho(S,R) = \sum_i e_i \sum_j s_{ij} r_{ij} + \sum_i e_i \sum_j s_{ij} c_{\rho,j}.$$

Now suppose two signaling games G^* and G^{**}, with the utility functions u^*_ρ and u^{**}_ρ, respectively, are completely identical except for the receiver's cost vector, which is $\vec{c}^{\,*}_2$ for G^* and $\vec{c}^{\,**}_2$ for G^{**}. (This means that n and m, $\vec{e}$ and $\vec{c}_1$ are identical for the two games.) It then follows from the above equation that the sender's utility function is identical in the two games, and that for all S and R,

$$u^*_2(S,R) - u^{**}_2(S,R) = \sum_i e_i \sum_j s_{ij}(c^*_{2,j} - c^{**}_{2,j}).$$

This means that the receiver's utilities $u^*_2(S,R)$ and $u^{**}_2(S,R)$ in the two games always differ by an amount that only depends on the sender's strategy S. So the receiver's payoff matrix for G^{**} (where the receiver is assumed to be the column player) can be obtained from the corresponding matrix for G^* by adding a constant amount to each row. In other words, there is a function f from sender strategies to real numbers such that

$$u^*_2(S,R) = u^{**}_2(S,R) + f(S).$$

G^* and G^{**} are thus equivalent for all intents and purposes that are relevant in the context of this paper.[3] Likewise, it is easy to see that similar equivalences hold between the symmetrized versions of G^* and G^{**}. They have the same evolutionarily stable states, the same evolutionarily stable sets, the same neutrally stable states, and the same replicator dynamics. From the last point it follows that the symmetrized version of G^* and G^{**} have the same dynamic stability properties.

Since any two signaling games that only differ with respect to $\vec{c}_2$ are thus equivalent, it is convenient to consider only one representative of each equivalence class, namely one where $\vec{c}_1 = \vec{c}_2$, i.e. where sender and receiver happen to have the same cost function. This is a convenient choice because these games are partnership games. In the following, I will thus always assume, without restriction of generality, that $\vec{c}_1 = \vec{c}_2$, and drop the role subscript.

As a further simplification, we observe that adding the same constant to each element of $\vec{c}$ amounts to adding that constant to the utility of the normal form game. Since such a constant is immaterial to the character of the game, costs can always be normalized in such a way that the maximal cost is 0 and all other costs are negative:

$$\max_i c_i = 0.$$

Events with zero probability have no impact on the utility and can be ignored. Therefore I assume that $\forall i : e_i > 0$. Also, if the difference between the costs of two signals exceeds 1, it is never rationalizable to use the more expensive signal. We can therefore restrict attention to games were $\forall i : c_i \geqslant -1$.

Also, unless otherwise noted, I will only consider games that are *generic* in the sense that small changes of the vectors $\vec{e}$ and $\vec{c}$ do not change the qualitative character of the game. This amounts to the requirement that two signals are never equiprobable, and two signals are never equally costly: $e_i \neq e_j$ if $i \neq j$, and $c_i \neq c_j$ if $i \neq j$. Also, genericity entails that $\forall i : c_i > -1$.

It is convenient to assume that events are ordered according to probability and signals according to their costs. I therefore introduce the following convention without restriction of generality:

$$e_i > e_j \quad \text{iff} \quad i > j,$$
$$c_i > c_j \quad \text{iff} \quad i > j.$$

So the first row of the S-matrix corresponds to the most frequent event, the first column to the cheapest signal, etc.

To simplify notation, we introduce the following conventions:

- We construct a matrix P that is like S except that each row is multiplied with the probability of the corresponding event, and
- we construct a matrix Q that is like R except that to each cell, the costs of the signal corresponding to the row of that cell is added.

Definition 1.

$$p^S_{ij} \doteq s_{ij} \times e_i,$$
$$q^R_{ij} \doteq r_{ij} + c_i.$$

The S-matrix and R-matrix, as well as the parameters of the game, can easily be recovered from the P-matrix and Q-matrix.

$$e_i = \sum_j p_{ij},$$
$$c_j = \frac{\sum_i q_{ji} - 1}{m},$$
$$s_{ij} = \frac{p_{ij}}{e_i},$$
$$r_{ji} = q_{ji} - c_j.$$

It is therefore usually more convenient to work directly with P and Q, rather than with S and R.

With these conventions, the utility function for pure strategies reduces to

$$u(S,R) = \mathrm{tr}(P^S Q^R). \tag{1}$$

Let x and y be mixed strategies of the sender and the receiver, respectively. Since there is a one–one map between (S,R)-matrices and (P,Q)-matrices (for a given set of parameters), x and y uniquely correspond to probability distributions over P-matrices and Q-matrices, respectively. I will denote these derived probability distributions with x and y as well, because the context will always make clear which interpretation is intended.

Also note that both matrix multiplication and the trace function are linear functions. Let $\mathcal{P}$ be the set of P-matrices that

[3] To see why, observe that for each pair of mixed strategies (x,y) of the asymmetric game,

$$u^*_2(x,y) = u^{**}_2(x,y) + \sum_S x(S)f(S).$$

So we have for all x, y, and z: $u^*_2(x,y) - u^*_2(x,z) = u^{**}_2(x,y) - u^{**}_2(x,z)$. This entails that G^* and G^{**} have the same Nash equilibria and strict Nash equilibria, and that they share their asymmetric replicator dynamics.

The same point can be made for the symmetrized version of the two games. It follows from the definition of the symmetrization of an asymmetric game (see below) that there is function g from mixed symmetric strategies to real numbers such that

$$u^*_{sym}(x,y) = u^{**}_{sym}(x,y) + g(y),$$

where u^*_{sym} and u^{**}_{sym} are the mixed strategy utility functions of the symmetrized versions of G^* and G^{**} respectively. So here we have for all x, y, and z: $u^*_{sym}(x,y) - u^*_{sym}(z,y) = u^{**}_{sym}(x,y) - u^{**}_{sym}(z,y)$. This entails that the symmetrized games also share their Nash equilibria, strict Nash equilibria and replicator dynamics, and they also share their ESSs and ESSets.

correspond to pure strategies in $\mathcal{S}$, and $\mathcal{Q}$ the set of Q-matrices corresponding to pure R-matrices in $\mathcal{R}$. The utility function for pure strategies given in (1) thus readily extends to mixed strategies:

$$\begin{aligned}
u(x,y) &= \sum_{S\in\mathcal{S}} x(S) \sum_{R\in\mathcal{R}} (y(R)\, u(S,R)) \\
&= \sum_{S\in\mathcal{S}} x(S) \sum_{R\in\mathcal{R}} (y(R)\mathrm{tr}(P^S Q^R)) \\
&= \sum_{P\in\mathcal{P}} x(P) \sum_{Q\in\mathcal{Q}} (y(Q)\mathrm{tr}(PQ)) \\
&= \mathrm{tr}(P^x Q^y). \qquad (2)
\end{aligned}$$

The utility function (2) defines an asymmetric partnership game. Such a game can be transformed into a doubly symmetric game (i.e. a symmetric partnership game) in the standard way. A strategy of the symmetrized game is a pair of strategies for the asymmetric game, one for each role. The symmetric utility function is derived from the asymmetric one in the following way:

$$u_{sym}((S_1,R_1),(S_2,R_2)) = u_{asym}(S_1,R_2) + u_{asym}(S_2,R_1).$$

Most authors multiply the symmetrized utility by $\frac{1}{2}$ to capture the intuition that each player finds himself in each role with equal probability. I drop this constant factor for convenience, as it has no bearing on the structure of the game. Also, I will denote the symmetric utility function simply as u from now on.

Let x be a mixed strategy of the symmetrized game. The matrices P^x and Q^x corresponding to this strategy are defined as

$$P^x \doteq \sum_{(P,Q)\in\mathcal{P}\times\mathcal{Q}} x((P,Q))P,$$

$$Q^x \doteq \sum_{(P,Q)\in\mathcal{P}\times\mathcal{Q}} x((P,Q))Q.$$

It is easy to see that the symmetric utility function for mixed strategies comes down to

$$u(x,y) = \mathrm{tr}(P^x Q^y) + \mathrm{tr}(P^y Q^x).$$

Nash equilibria and strict Nash equilibria can be characterized in terms of P-matrices and Q-matrices.

Lemma 1. *x is a Nash equilibrium if and only if the following conditions are fulfilled*:

1. *If $p^x_{ij} > 0$, then $q^x_{ji} = \max_{j'} q^x_{j'i}$.*
2. *If $q^x_{ji} > c_j$, then $p^x_{ij} = \max_{i'} p^x_{i'j}$.*

See the Appendix for a proof.

Lemma 2. *x is a strict Nash equilibrium if and only if*

1. *x is a Nash equilibrium, and*
2. *Each column in P^x and in Q^x has a unique maximum.*

The proof is given in the Appendix.

3. Some examples

Example 1. One might hypothesize that the number of events n provides an upper bound on the number of useful signals m. This is not necessarily so, as the following example demonstrates:

$$\vec{e} = \langle .6, .4\rangle,$$
$$\vec{c} = \langle 0, -.1, -.9\rangle,$$

$$S^x = \begin{pmatrix} .5 & .5 & 0 \\ .75 & 0 & .25 \end{pmatrix}, \quad R^x = \begin{pmatrix} .9 & .1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$P^x = \begin{pmatrix} .3 & .3 & 0 \\ .3 & 0 & .1 \end{pmatrix}, \quad Q^x = \begin{pmatrix} .9 & .1 \\ .9 & -.1 \\ -.9 & .1 \end{pmatrix}$$

x is a Nash equilibrium. Even though there are more signals than events, each signal is used in equilibrium.

Example 2. Likewise, there are Nash equilibria where the number of events exceeds the number of signals, like the following:

$$\vec{e} = \langle .5, .3, .2\rangle,$$
$$\vec{c} = \langle 0, -.1\rangle,$$

$$S^x = \begin{pmatrix} 1 & 0 \\ 1/3 & 2/3 \\ 0 & 1 \end{pmatrix}, \quad R^x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & .1 & .9 \end{pmatrix},$$

$$P^x = \begin{pmatrix} .5 & 0 \\ .1 & .2 \\ 0 & .2 \end{pmatrix}, \quad Q^x = \begin{pmatrix} 1 & 0 & 0 \\ -.1 & 0 & .8 \end{pmatrix}.$$

In this equilibrium, each event is a possible interpretation of some signal.

We now turn to a game that is perhaps the simplest conceivable non-trivial example for this class of games. Let $\vec{e} = \langle .75, .25\rangle$ and $\vec{c} = \langle 0, -.1\rangle$.

Example 3. In symmetrized asymmetric games, the ESSs are exactly the strict Nash equilibria. In the game at hand, there are two of them, namely:

$$S^{x_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R^{x_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$P^{x_1} = \begin{pmatrix} .75 & 0 \\ 0 & .25 \end{pmatrix}, \quad Q^{x_1} = \begin{pmatrix} 1 & 0 \\ -.1 & .9 \end{pmatrix}$$

and

$$S^{x_2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad R^{x_2} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$P^{x_2} = \begin{pmatrix} 0 & .75 \\ .25 & 0 \end{pmatrix}, \quad Q^{x_2} = \begin{pmatrix} 0 & 1 \\ .9 & -.1 \end{pmatrix}.$$

Example 4. Recall that a profile x in a symmetric game is neutrally stable if and only if it is a Nash equilibrium, and $u(x,y) \geqslant u(y,y)$ for all (alternative best replies) y with $u(y,x) = u(x,x)$. (The only difference to evolutionary stability is that in the latter notion, the last inequality has to be strict.) Next to the two ESSs just given, there is an infinity of NSSs that form a linear manifold, namely

$$S^x = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad R^x = \begin{pmatrix} 1 & 0 \\ \alpha & 1-\alpha \end{pmatrix},$$

$$P^x = \begin{pmatrix} .75 & 0 \\ .25 & 0 \end{pmatrix}, \quad Q^x = \begin{pmatrix} 1 & 0 \\ \alpha - .1 & .9 - \alpha \end{pmatrix}$$

for $\alpha \in (.9, 1]$.

The unique best response to Q^x is P^x. There are two pure best responses to P^x, namely

$$R_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},$$

$$Q_1 = \begin{pmatrix} 1 & 0 \\ .9 & -.1 \end{pmatrix}$$

and

$$R_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$Q_2 = \begin{pmatrix} 1 & 0 \\ -.1 & .9 \end{pmatrix}.$$

We have

$$\mathrm{tr}(P^x Q_1) = \mathrm{tr}(P^x Q_2) = \mathrm{tr}(P^x Q^x) = .75.$$

Hence x is in fact neutrally stable.

Example 5. If $\alpha = .9$, we get another equilibrium:

$$S^x = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad R^x = \begin{pmatrix} 1 & 0 \\ .9 & .1 \end{pmatrix},$$

$$P^x = \begin{pmatrix} .75 & 0 \\ .25 & 0 \end{pmatrix}, \quad Q^x = \begin{pmatrix} 1 & 0 \\ .8 & 0 \end{pmatrix}.$$

One possible best response to this is

$$S^y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R^y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$P^y = \begin{pmatrix} .75 & 0 \\ 0 & .25 \end{pmatrix}, \quad Q^y = \begin{pmatrix} 1 & 0 \\ -.1 & .9 \end{pmatrix}.$$

Now we have

$$\mathrm{tr}(P^x Q^x) = \mathrm{tr}(P^y Q^x) = \mathrm{tr}(P^x Q^y) = .75 < \mathrm{tr}(P^y Q^y) = .975.$$

Hence x is not neutrally stable.

Example 6. There is yet another Nash equilibrium of this game, which is also not neutrally stable:

$$S^x = \begin{pmatrix} 1/3 & 2/3 \\ 1 & 0 \end{pmatrix}, \quad R^x = \begin{pmatrix} .9 & .1 \\ 1 & 0 \end{pmatrix},$$

$$P^x = \begin{pmatrix} .25 & .5 \\ .25 & 0 \end{pmatrix}, \quad Q^x = \begin{pmatrix} .9 & .1 \\ .9 & -.1 \end{pmatrix}.$$

A possible best response to this is

$$S^y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad R^y = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$P^y = \begin{pmatrix} 0 & .75 \\ .25 & 0 \end{pmatrix}, \quad Q^y = \begin{pmatrix} 0 & 1 \\ .9 & -.1 \end{pmatrix}.$$

Now we have

$$\mathrm{tr}(P^x Q^x) = \mathrm{tr}(P^y Q^x) = \mathrm{tr}(P^x Q^y) = .7 < \mathrm{tr}(P^y Q^y) = .925.$$

Hence x is not neutrally stable either.

There are no more Nash equilibria of this game.

4. Evolutionary stability

As shown by Selten (1980), the ESSs in a symmetrized asymmetric game are exactly the strict Nash equilibria. Consequently, neither the P-matrix nor the Q-matrix in an ESS contains any multiple column maxima. In particular, the P-matrix cannot contain any zero-columns. Also, strict equilibria are always pure strategies. So in an ESS, each row of the P-matrix contains exactly one positive entry; all other entries are 0. As a result, a game can only have an ESS if $m \leqslant n$.

Suppose $n = m$. If P is pure and does not contain zero-columns, the corresponding S-matrix must be a permutation matrix. If each column of P thus contains exactly one entry >0, the unique best response to this is an R-matrix that is a transpose of S. Since $c_i > -1$ for all i, it follows that all non-zero entries of R correspond to column maxima in Q. Hence P is the unique best response to Q.

These considerations can be summarized in the following

Observation 1. *If $n = m$, x is an ESS if and only if S^x is a permutation matrix and R^x its transpose.*

Finally, consider the case where $m < n$. Any pure S-matrix now necessarily contains columns with multiple one-entries. However, due to genericity, the corresponding P-matrix nevertheless has a unique maximum in each column. So there is a unique best response R to such a P-matrix. R has at least one zero-column. Again, due to genericity, the corresponding columns in the Q-matrix still have a unique maximum each, namely the entry in the first row (which is 0, while all other entries are negative). Since (P, Q) is a Nash equilibrium, the rows of P corresponding to the zero-columns of R have their unique non-zero entry in the first column. It thus follows that only the first column of P contains more than one positive entry. So we have:

Observation 2. *If x is an ESS and $m < n$, then*:

- *the first column of P^x has $n - m + 1$ positive entries*,
- *each other column of P^x has exactly one positive entry, and*
- $q^x_{ji} > c_j$ *iff* $i = \min(\{i' : p^x_{i'j} > 0\})$.

These observations can be combined into:

Theorem 1. *x is an ESS if and only if*

1. $m \leqslant n$,
2. *the first column of P^x has $n - m + 1$ positive entries*,
3. *each other column of P^x has exactly one positive entry, and*
4. $q^x_{ji} = 1 + c_j$ *iff* $i = \min(\{i' : p^x_{i'j} > 0\})$, *otherwise* $q^x_{ji} = c_j$.

The proof of the only–if direction is given above. The if-direction is proved in the Appendix.

If $m < n$, there are necessarily some events that are never inferred by the receiver in an ESS. Note that these need not be the least likely events though. The following profile is an ESS, even though the second event is never correctly communicated, while the less likely third event is correctly communicated.

$$\vec{e} = \langle .5, .3, .2 \rangle,$$
$$\vec{c} = \langle 0, -.1, \rangle,$$

$$S^x = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R^x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$P^x = \begin{pmatrix} .5 & 0 \\ .3 & 0 \\ 0 & .2 \end{pmatrix}, \quad Q^x = \begin{pmatrix} 1 & 0 & 0 \\ -.1 & -.1 & .9 \end{pmatrix}.$$

There is no guarantee that the evolutionary dynamics will carry a population to some ESS. Some games do not even have ESSs, like those games considered here where $m > n$. It is nevertheless possible to make predictions about the long-term behavior of populations that do not converge to some ESS. The notion of an *evolutionarily stable set* is of central importance here. Briefly put, it can be shown that the replicator dynamics plus a small amount of random drift will guarantee that a population eventually converges to such a set. The details of this are spelled out in Section 6.

The notion of evolutionarily stable sets goes back to Thomas (1985); the definition used here is taken from Cressman (2003).

Definition 2. A set A of symmetric Nash equilibria is an *evolutionarily stable set* (ESSet) if, for all $x^* \in A, u(x^*, x) > u(x, x)$ whenever $u(x, x^*) = u(x^*, x^*)$ and $x \notin A$.

Every ESS forms a singleton ESSet.

A simple example for a non-singleton ESSet would be the following, with—for instance—$\vec{e} = \langle .8, .2 \rangle$ and $\vec{c} = \langle 0, -.1, -.2 \rangle$:

$$\left\{ x : S^x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad R^x = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \alpha & 1-\alpha \end{pmatrix} \text{ and } \alpha \in [0,1] \right\}.$$

In fact, all non-singleton ESSets are sets of a similar shape (or unions of such sets).

Theorem 2. *A set of strategies A is an ESSet iff for each $x \in A$, x is an ESS or*

1. *$m > n$,*
2. *the restriction of P^x to the first n columns and the restriction of Q^x to the first n rows form an ESS, and*
3. *for each y such that $P^x = P^y$, and Q^x and Q^y agree on the first n rows: $y \in A$.*

The proof, given in the Appendix, makes use of the notion of neutral stability, which is discussed in the next section.

The result is intuitively unsurprising. If $m > n$, we have more signals at our disposal than necessary to achieve optimal communication. Therefore only the n cheapest signals are ever used. Since the other signals are not used, their interpretation has no impact on fitness, and natural selection does not operate on this aspect of the strategies.

5. Neutral stability

As illustrated in Example 4, there may be equilibria that are not evolutionarily but neutrally stable. To characterize the properties of neutral stability, the following lemma proves useful.

Lemma 3. *If x is a Nash equilibrium, and both $0 < p^x_{i^*j^*} < e_{i^*}$ and $c_{j^*} < q^x_{j^*i^*} < 1 + c_{j^*}$ for some i^*, j^*, then x is not neutrally stable.*

A proof is given in the Appendix.

We can make the following observation:

Lemma 4. *If x is an NSS, P^x is a pure strategy.*

There are thus no neutrally stable states in the interior of the state space.

Q^x may be mixed in an NSS, as Example 4 illustrates. Regarding to Q^x, we can make the following observation:

Lemma 5. *If x is an NSS, Q^x does not contain multiple column maxima.*

The proofs of these lemmas are given in the Appendix.

The opposite direction actually holds as well, which leads to a concise characterization of neutral stability:

Theorem 3. *x is an NSS if and only if it is a Nash equilibrium and Q^x does not contain multiple column maxima.*

Proof. The forward direction corresponds to Lemma 5. As for the backward direction, suppose x is Nash, and Q^x does not contain multiple column maxima. It follows immediately that for all $y \in BR(x) : P^y = P^x$. So there cannot be a y with $2\,\mathrm{tr}(P^x Q^x) = \mathrm{tr}(P^x Q^y) + \mathrm{tr}(P^y Q^x) < 2\,\mathrm{tr}(P^y Q^y)$. Hence x is neutrally stable. □

In an NSS, non-determinism thus can only occur in the receiver strategy. It is quite restricted insofar as it can only occur as response to some zero-column in P:

Observation 3. *If x is an NSS and there are some i, i', j with $c_j < q^x_{ji}, q^x_{ji'} < 1 + c_j$, then $\forall i' : p^x_{i'j} = 0$.*

This follows directly from the facts that non-determinism can only occur in response to multiple column maxima, which, due to genericity, can only occur in a zero-column of P if P is pure.

Note that neutral stability without evolutionary stability is quite a pervasive phenomenon.

Observation 4. *If $m, n \geqslant 2$, there is always at least one NSS that is not element of an ESSet.*

For instance, putting all probability mass into the first column both on the sender side and the receiver side leads to an NSS that is obviously not contained in any ESSet.

6. Dynamic stability and basins of attraction

The games considered in this paper are symmetrized asymmetric (or bimatrix) games. As developed in detail in Cressman (2003), there is a tight connection between static stability and dynamic stability under the replicator dynamics for this class of games. Most notably, a set of strategies is asymptotically stable under the replicator dynamics if and only if it is an ESSet. As a corollary, it follows that the asymptotically stable states are exactly the ESSs.

Let us have a look at the dynamic properties of the set of neutrally stable equilibria. It is rather obvious that all ESSs are isolated points in the sense that each ESS has a neighborhood that does not contain any other Nash equilibria. This follows from the facts that (a) all Nash equilibria are fixed points under the replicator dynamics and (b) each ESS is asymptotically stable under the replicator dynamics.

The set of NSSs that are not ES has a richer topological structure. (I use the usual Euclidean norm here, i.e. $\|x - y\| = \sqrt{\sum_{(P,Q) \in \mathcal{P} \times \mathcal{Q}} (x(P,Q) - y(P,Q))^2}$.)

Lemma 6. *Let x^* be an NSS that is not an ESS. There is some $\varepsilon > 0$ such that for each Nash equilibrium y with $\|x - y\| < \varepsilon$,*

1. *y is itself neutrally stable, and*
2. *for each $\alpha \in [0,1]$, $\alpha x^* + (1-\alpha) y$ is neutrally stable.*

See the Appendix for a proof.

We say that two NSSs x and y belong to the same continuum of NSSs if for each $\alpha \in [0,1]$, $\alpha x^* + (1-\alpha) y$ is neutrally stable.

Theorem 4. *Each NSS x has some non-null neighborhood A such that each interior point in A converges to some neutrally stable equilibrium y under the replicator dynamics that belongs to the same continuum of NSSs as x.*

Proof. The proof basically follows the strategy of the corresponding proof in Pawlowitsch (2007). In Thomas (1985) it is shown that each NSS x is Lyapunov stable under the replicator dynamics. This means that for each ε-environment of x there is a neighborhood A of x such that each point in A remains within the ε-environment of x in positive time. We also know (part of the so-called "folk theorem of evolutionary game theory", cf. Hofbauer and Sigmund, 1998) that each convergent trajectory in the interior of the state space converges to a Nash equilibrium. Finally, in Akin and Hofbauer (1982) it is proven that in doubly symmetric games, each trajectory converges. According to Lemma 6, there is some ε-environment of x such that all Nash equilibria within it belong to the same continuum of NSSs as x. So there is some neighborhood A of x such that all points in A remain within the ε-environment of x. Hence all interior points within A converge to some NSS that belongs to the same continuum of NSSs as x. □

We get the immediate corollary:

Corollary 1. *Each NSS belongs to some continuum of NSSs that attracts a set of states with a positive measure.*

Proof. As A is a neighborhood of x, it has a positive measure. The boundary of the state space has measure zero. Hence the set of interior points within A has a positive measure. □

It is obvious that each ESSet has a basin of attraction with a positive measure—this follows directly from the fact that each ESSet is asymptotically stable. The corollary shows though that the basins of attraction of the ESSets do not exhaust the state space. As pointed out in Observation 4, there are NSSs that do not belong to any ESSet. As ESSets are asymptotically stable, each ESSet has a neighborhood that does not contain any NSSs. Hence if an NSS x does not belong to any ESSet, the entire continuum of NSSs that x belongs to is disjoint from the ESSets. We thus get the additional corollary:

Corollary 2. *The set of Nash equilibria that do not belong to any ESSet attracts a set of states with a positive measure.*

One might conjecture that Nash equilibria that are not stable do not have a basin of attraction with a positive measure—or, to put it the other way round, that almost all initial states converge to an NSS. While this seems plausible, this issue is, to my knowledge, still unsolved.

So in the absence of neutral drift, there is no guarantee that a population will converge to some ESSet under the replicator dynamics. However, it can be shown quite easily that a combination of natural selection and neutral drift leads into some ESSet from any arbitrary initial state. (A similar result is given by Trapa and Nowak (2000) for the class of games considered there.)

Theorem 5. *Given any strategy profile x_1, there is a finite sequence of profiles $(x_i)_{i\leqslant n}$ for some $n\in\mathbb{N}$ such that*

1. *there is an ESSet E such that $x_n\in E$, and*
2. *$u(x_{i+1},x_i)\geqslant u(x_i,x_i)$ $\forall i<n$.*

Proof. see Appendix.

If $u(x_{i+1},x_i)\geqslant u(x_i,x_i)$, this means that an x_i-population can successfully be invaded and replaced by x_{i+1}-mutants, either due to natural selection (if the inequality is strict) or by neutral drift (in case of equality). So the theorem entails that from each point there is at least one trajectory that leads to some ESSet if the effects of neutral drift are taken into account. ESSets, however, are protected against the effects of drift because they are asymptotically stable.

7. Relation to previous work

Results that are similar to those presented above have been obtained by Pawlowitsch (2007) for the class of signaling games where $m=n$, all events have the same probability, and signaling is costless (or, equivalently, all events incur the same costs). In these games too, if $n\geqslant 3$ there are sets of neutrally stable states not belonging to any ESSet that attract a set of initial states with a positive measure.[4] Despite this global similarity, the structure of these neutrally stable sets are quite different in the two classes of games. The only neutrally stable states of a game with generic probabilities and payoffs (and $n=m$) that remain neutrally stable when $\vec{e}$ and $\vec{c}$ are replaced by constant vectors (i.e. when a game from the class considered here is transformed into an isomorphic game from the class considered in Pawlowitsch's paper) are the ESSs.

To see why, consider some neutrally, but not evolutionarily, stable state x of a game from the class considered here, with $m=n$. It cannot be a strict equilibrium, because otherwise it would be an ESS. However, according to the Lemmas 4 and 5, P^x is a pure strategy that is the unique best response to Q^x. So there must be multiple best responses to P^x. As P^x is pure and probabilities are generic, P^x must contain a zero-column. Let us say that the j_1th column of P^x is a zero-column. Since $n=m$, P^x must also contain a column with multiple positive entries. Let us say that this column is j_2, and that i_1 and i_2 are the two smallest row indices such that $p^x_{i_1j_2},p^x_{i_2j_2}>0$. Since x is a Nash equilibrium, $r^x_{j_2i_1}=1$, and thus $r^x_{j_2i_2}=0$.

An example of this configuration was given in Example 4 and is repeated here:

$$S^x=\begin{pmatrix}1&0\\1&0\end{pmatrix},\quad R^x=\begin{pmatrix}1&0\\\alpha&1-\alpha\end{pmatrix},$$

$$P^x=\begin{pmatrix}.75&0\\.25&0\end{pmatrix},\quad Q^x=\begin{pmatrix}1&0\\\alpha-.1&.9-\alpha\end{pmatrix}$$

for some $\alpha\in(.9,1]$. Here $j_1=2$, $j_2=1$, $i_1=1$ and $i_2=2$. Now suppose x is also a Nash equilibrium for constant $\vec{e}$ and $\vec{c}$. Then each 1-entry in S^x must correspond to a column maximum in R^x. Since $r^x_{j_2i_2}=0$, the i_2th column of r^x must be a zero-column. In the example, this entails that $\alpha=1$.

$$S^x=\begin{pmatrix}1&0\\1&0\end{pmatrix},\quad R^x=\begin{pmatrix}1&0\\1&0\end{pmatrix}.$$

So both S^x and R^x have at least one zero-column. As proven in Pawlowitsch (2007), if $\vec{e}$ and $\vec{c}$ are constant vectors, such a strategy is not neutrally stable.

So if the vectors $\vec{e}$ and $\vec{c}$ are transformed from generic to constant (or vice versa), all properly neutrally stable states cease to be stable, while new continua of neutrally stable states emerge. Future research will have to show whether it is possible to develop a characterization of neutral stability that does not impose restrictions on $\vec{e}$ and $\vec{c}$, such that Pawlowitsch's and the present results come out as special cases. For the time being, I have to leave this issue open.

A similar but more general result is proven by Huttegger (2007). He investigates the class of signaling games where $n=m$ and signals are costless, but he drops the assumption that events are equiprobable. Neither does he assume that event probabilities are generic; no restrictions whatsoever are placed on $\vec{e}$. He shows that if $n\geqslant 3$, there is always a set of neutrally stable states that attracts a set of initial states with a positive measure. If event

[4] Pawlowitsch does not explicitly discuss ESSets beyond ESSs, but it is easy to see that there are no ESSets in the class of games investigated by her that do not consist solely of ESSs. Suppose otherwise. As shown in Cressman (2003), ESSets of symmetrized asymmetric games are sets of Nash equilibria that are closed under best response. So if E is an ESSet and $x\in E$ is not a strict equilibrium, there must be a best response y to x such that y is pure and $y\in E$. If S^y is a permutation matrix, then R^y must be its transpose (and vice versa), i.e. y would be strict. But since x must be a best response to y, x would also be strict, contra assumption. So both S^y and R^y must contain at least one zero-column. As shown by Pawlowitsch, such a state cannot be neutrally stable, but all elements of some ESSet must be neutrally stable, so we have a contradiction.

probabilities are generic, this result even extends to the case where $n = m = 2$.

Evolutionary stability conditions for games with costly signaling are also investigated in Blume et al. (1993). They consider a wider class of utility functions, which contains the one considered here as a special case. On the other hand, they restrict attention to games where $m > 2^{m+n} + n$, none of which has ESSs. The authors focus on the set-valued stability concept of an *equilibrium evolutionarily stable set* (EES set; cf. Swinkels, 1992), which is related, but not identical to the notion of an ESSet. Hence the issues addressed in their paper are complementary to the ones discussed here.

8. Conclusion

The paper investigated the conditions for static and dynamic stability for a certain class of signaling games. Specifically, I focused on games where the interests of sender and receiver are identical except possibly regarding signaling costs, signals incur differential costs and events have differential probabilities. The investigation was restricted to generic parameter configurations. It turned out that essentially those languages are evolutionarily stable where sender and receiver use the same bijective map between events and signals. If the number of events exceeds the number of signals, some events are never communicated though. If, on the other hand, the number of signals exceeds the number of events, the most expensive signals are never used.

One might argue that actually only the case with $m = n$ is relevant. It is always possible to introduce new signals via some kind of mutation if $m < n$, and signals that are never used (in case of $m > n$) are virtually non-existent. Whether or not this is really the case perhaps depends on the precise physical and cognitive boundary conditions. While for instance natural languages can use recursivity to create an unbounded number of signals, this complexity comes at a cost that in some cases may be so high that the signal in question turns out to be useless.

The second finding of the paper concerns neutral stability. Each of the investigated games has an infinity of neutrally stable states that are not included in any ESSets. Furthermore, these sub-optimal states jointly attract a set of states with a positive measure. So if evolution consisted only of natural selection, we would predict that such states are in fact empirically observable. They are not protected against neutral drift though, so depending on the force of drift, neutral stability may or may not be relevant empirically. If the population is sufficiently large (as it has to be if the replicator dynamics is applicable) and the mutation rate is low, it may take an arbitrarily long time until a population leaves some continuum of sub-optimal states.

These weakly stable states are characterized by the fact that certain potentially useful signals are in fact never used. One might argue that such situations are in fact attested in historical linguistics. For instance, the Proto-Indo-European long vowels /a:/ and /o:/ collapsed into /o:/ in Proto-Germanic, leaving the latter language with the phonetic slot of the long /a:/ unfilled. This process must have taken place within the first millennium BCE. Due to a seemingly unrelated process of linguistic change, Germanic short vowels were lengthened in certain words during the transition from Proto-Germanic to Old High German.[5] In this way, Old High German re-acquired the long vowel /a:/. This (probably) happened several centuries later. One might hypothesize that the Proto-Germanic vowel system was neutrally but not evolutionarily stable. If correct, this would be an indication that neutrally stable states can persist a significant amount of time even if they are not protected against drift.

Be this as it may, the issue whether or not—or under what conditions—neutrally but not evolutionarily stable states are expected to be observable empirically is an intriguing one that deserves further investigation.

A final point concerns the issue of symmetric versus asymmetric dynamics. The present paper only investigated the stability properties of symmetrized signaling games. This is justified in scenarios where each individual can act both as a sender and as a receiver. This is valid in many domains of application—especially in linguistics, but also for certain signaling systems in biology and economics. However, various scenarios for the evolution of signaling games involve a genuinely asymmetric situation, like signals being sent from males to females, interspecific communication, signaling from seller to buyer, etc. Fortunately all results for the symmetrized games straightforwardly carry over to the bi-population replicator dynamics of the corresponding asymmetric game. In particular, it follows from results given in Cressman (2003) that a set of profiles is asymptotically stable according to the symmetrized replicator dynamics if and only if it is asymptotically stable under the asymmetric replicator dynamics.

Likewise, the proofs of the statements from Section 6 regarding neutral stability carry over the asymmetric setting without problems. So in particular the two corollaries (with "ESSet" replaced by "SESet" in the second corollary) also hold for the bi-population dynamics.

Appendix

Proof of Lemma 1. The proof relies on the assumption that each stochastic matrix M can be represented as the expected value of a probability distribution over pure stochastic matrices, i.e. matrices with exactly one 1 per row and 0 everywhere else. This can be shown by induction over the number of cells with values in $(0, 1)$. If there are no such cells, M is already pure. So let us assume that M contains n values in $(0, 1)$, and let $m_{ij} = \alpha \in (0, 1)$ be the smallest of these values.

Then let M' be a pure stochastic matrix which has all its 1-entries at positions where M has positive entries. Furthermore, $m'_{ij} = 1$. $M'' = (1-\alpha)^{-1}(M - \alpha M')$ is then a stochastic matrix containing $n-1$ positive entries, and $M = (1-\alpha)M'' + \alpha M'$. By induction hypothesis, M'' is the weighted average of a set of pure matrices. Hence M is the weighted average of a set of pure matrices as well, which concludes the induction step.

Now suppose x is a Nash equilibrium and $p^x_{ij} > 0$. Also, suppose there is some j' such that $q^x_{j'i} > q^x_{ji}$. Let y be so that P^y is exactly like P^x except that $p^y_{ij} = 0$ and $p^y_{ij'} = p^x_{ij'} + p^x_{ij}$. (It follows from the construction given in the previous paragraph that such a y always exists.) Then $\mathrm{tr}(P^y Q^x) - \mathrm{tr}(P^x Q^x) = p^x_{ij}(q^x_{j'i} - q^x_{ji}) > 0$, so P^x is not a best response to Q^x. This contradicts the assumption that x is Nash, 1 must in fact hold. The proof of 2 is entirely analogous.

As for the other direction, suppose P^x is not a best response to Q^x. Hence there is some best response P^y to Q^x with $\mathrm{tr}(P^y Q^x) > \mathrm{tr}(P^x Q^x)$. So there must be at least one i with

$$\sum_j p^y_{ij} q^x_{ji} > \sum_j p^x_{ij} q^x_{ji}.$$

Let

$$i^* = \arg\max_i \left(\sum_j p^y_{ij} q^x_{ji} - \sum_j p^x_{ij} q^x_{ji} \right).$$

[5] More precisely, a Proto-Germanic sequence of short vowel + nasal consonant + voiceless uvular fricative was replaced by a sequence of long vowel + fricative in Old High German.

So necessarily

$$\sum_j p^y_{i^*j}q^x_{ji^*} > \sum_j p^x_{i^*j}q^x_{ji^*}.$$

Let j^* be such that $q^x_{j^*i^*}$ is maximal within its column. Then for some matrix P^z which is exactly like P^y except that $p^z_{i^*j^*} = 1$ we have

$$\sum_j p^z_{i^*j}q^x_{ji^*} = q^x_{j^*i^*}.$$

Hence, because P^y is a best response to Q^x,

$$\sum_j p^y_{i^*j}q^x_{ji^*} = q^x_{j^*i^*}$$

as well, and thus

$$\sum_j p^x_{i^*j}q^x_{ji^*} < q^x_{j^*i^*}.$$

So there must be some i' with $p^x_{i'j^*} > 0$ and $q^x_{j^*i'} < q^x_{j^*i^*}$.

By contraposition of this argument, we infer that if the assumption 1 is met, P^x is a best response to Q^x. By a similar argument it can be shown that assumption 2 entails that Q^x is a best response to P^x. These two implications jointly entail the lemma. □

Proof of Lemma 2. Suppose x is a strict Nash equilibrium, and suppose $p^x_{ij} = p^x_{i'j}$ are both column maxima with $i \neq i'$. Let y and z be so that (P^y, Q^y) and (P^z, Q^z) are exactly like (P^x, Q^x) except that $q^y_{ji} = 1 + c_j$ and $q^z_{ji'} = 1 + c_j$. Both y and z are then best responses to x, so x cannot be a strict Nash equilibrium.

Now suppose x is Nash, and each column in P^x and Q^x has a unique maximum. It follows from Lemma 1 that P^x has non-zero entries only at positions that correspond to column maxima in Q^x. Since column maxima are unique, there cannot be a y with $P^y \neq P^x$ such that y would be Nash as well. Likewise, Q^x has entries $q^x_{ji} > c_j$ only at positions that correspond to column maxima in P^x. By a similar argument, Q^x must be a unique best response to P^x. So x is in fact strict. □

Proof of Theorem 1. The proof for the only–if direction is given in the text. So suppose all four assumptions on the right-hand side hold. It follows that each column of P^x has exactly one maximum. According to Lemma 2, it remains to be shown that x is a Nash equilibrium.

It follows from the assumptions 2 and 3 that there are n positive entries in P^x—one per row. So if $p^x_{ij} > 0$, then $p^x_{ij} = e_i$. Due to genericity, therefore each column in P^x has unique maximum. For the first column, this must be the one with the lowest index. From assumption 4 it follows that the second condition of Lemma 1 is fulfilled.

Likewise, each column of Q^x has a unique maximum. If $q^x_{ji} = 1 + c_j$, q^x_{ji} is the unique maximum of the ith column of Q^x. If $q^x_{ji} = c_j$ for all j, the maximum of the ith column must be q^x_{1i}.

Suppose $j^* > 1$, and $p^x_{i^*j^*} > 0$. Because of assumption 3 $i^* = \min(\{i' : p^x_{i'j^*} > 0\})$, and hence $q^x_{j^*i^*} = 1 + c_{j^*}$ due to assumption 4. So $q_{j^*i^*}$ is a column maximum.

Now assume $p^x_{i^*1} > 0$. If $p^x_{i^*1}$ is a column maximum, $r^x_{1i^*} = 1$, and hence $q^x_{1j^*}$ is a column maximum. If $p^x_{i^*1}$ is not a column maximum, $r^x_{1i^*} = 0$. We have to show now that $q^x_{1i^*}$ is a column maximum. Suppose it is not. Then there must be a $j^* > 1$ with $q^x_{j^*i^*} = 1 + c_{j^*}$. By assumption 4, $p^x_{i^*j^*} > 0$. This is not possible though because the i^*th row of P^x contains exactly on positive entry (due to assumptions 2 and 3), which is $p^x_{i^*1}$ by assumption. Hence $q^x_{1i^*}$ is a column maximum. So according to Lemma 1, x is a Nash equilibrium, and according to Lemma 2, it is a strict equilibrium. Hence it is an ESS. □

Proof of Theorem 2. Suppose A is an ESSet, $x \in A$ and x is not an ESS. Cressman (2003, 46,47) proves that if $y \in A$ and $u(y,x) = u(x,x)$, then $u(x,y) = u(y,y)$. Furthermore, according to the definition, if $y \notin A$ and $u(y,x) = u(x,x)$, then $u(y,y) < u(x,x)$. So in either case, if y is a best response to x, then $u(y,y) \leqslant u(x,x)$, so x is neutrally stable.

According to Lemma 4, P^x is a pure strategy, and according to Lemma 5, Q^x does not contain multiple column maxima. If x is not an ESS, there must be an alternative best response y with $u(y,x) = u(x,x)$ and $u(y,y) = u(x,x)$. Since P^x is the unique best response to Q^x, $P^y = P^x$. Since P^x is pure and $x \neq y$, either Q^x is mixed and $Q^y = Q^x$ or $Q^y \neq Q^x$. In either case, it follows that there are multiple best responses to P^x. So P^x must have at least one column with multiple maxima. Because P^x is pure, this must be a zero-column. Let the j^*th column be the zero-column of P^x with the lowest index. Since P^x is a best response to Q^x, no entry in the j^*th row of Q^x can be a column maximum. Now suppose there is some $j > j^*$ and some i such that q_{ji} is a column maximum. Let z be exactly like x except that $q^z_{j^*i} = 1 + c_{j^*}$ (and all other entries in this row are c_{j^*}). Since Q^z is still a best response to x, it follows from the assumptions that $z \in A$. On the other hand, z is not a Nash equilibrium because the only positive entry in the ith row of P^z is in the jth column, but the column maximum of the ith column of Q^z is in the j^*th row. So $z \in A$ and z is not Nash, which is a contradiction. Therefore, if the j^*th column of P^x is a zero-column, all column maxima in Q^x must have a lower row index than j^*.

All columns in P^x with an index $j < j^*$ are not zero-columns. Therefore each of these columns has a unique maximum. As x is a Nash equilibrium, the corresponding rows in R^x are pure (do not contain entries strictly between 0 and 1). Hence there cannot be several column maxima within one row in Q^x. So $j^* = n + 1$, and hence $n < m$.

We thus established that the first n rows of Q^x contain all the column maxima of Q^x. Since P^x is a pure best response to Q^x, all positive entries in P^x must be located within the first n columns. But this entails that the restriction of S^x to the first n columns is a permutation matrix. Since Q^x is a best response to P^x, the restriction of R^x to the first n rows must be a transpose of S^x, so the restriction of P^x to the first n columns and of Q^x to the first n rows do in fact form an ESS.

Now suppose $m > n$, $P^x = P^y$, and Q^x and Q^y agree on the first n columns. By the argumentation given above, all column maxima of Q^x are located in the first n rows. Hence Q^x and Q^y agree on the location of the column maxima. They only disagree on the rows with an index $> n$, and these rows correspond to zero-columns in P^x/P^y. Hence $u(x,x) = u(y,x) = u(y,y)$, and therefore $y \in A$.

Now let us turn to the other direction. Suppose the conditions 1–3 are fulfilled and $x \in A$. To prove that A is an ESSet, we have to show that for each y with $u(y,x) = u(x,x)$ and $u(x,y) \leqslant u(y,y)$, it holds that $y \in A$. (I use the transpose of the formulation given in Definition 2.) So let us assume that $u(y,x) = u(x,x)$ and $u(x,y) \leqslant u(y,y)$.

According to assumptions 1 and 2 the restriction of R^x to the first n rows is a permutation matrix. Hence the first n rows of R^x are pure. This entails that Q^x cannot contain multiple column maxima. Rather, all column maxima of Q^x are located within the first n rows where R^x contains 1-entries. So P^x is a best response to Q^x. Likewise, according to assumption 2, Q^x is a best response to P^x. Hence x is a Nash equilibrium. Since $u(y,x) = u(x,x)$, y is a best response to x. Since Q^x does not contain multiple column maxima, there is a unique best response to it, so $P^x = P^y$. Likewise, the first n columns of P^x have unique column maxima, so Q^x and Q^y must agree on the first n rows. According to assumption 3, $y \in A$. □

Proof of Lemma 3. Suppose x is a Nash equilibrium with $0 < p^x_{i^*j^*} < e_{i^*}$ and $c_{j^*} < q^x_{j^*i^*} < 1 + c_{j^*}$. If $p^x_{i^*j^*} > 0$, $q^x_{j^*i^*}$ must be a column maximum according to Lemma 1. Also, there must be some $j \neq j^*$ with $p^x_{i^*j} > 0$, and hence $q^x_{ji^*}$ is a column maximum as well, hence $q^x_{ji^*} = q^x_{j^*i^*}$. Likewise, if $q^x_{j^*i^*} > c_{j^*}$, there must be some $i \neq i^*$ with $q^x_{j^*i} > c_{j^*}$, and hence $p^x_{ij^*} = p^x_{i^*j^*}$ are both column maxima.

We can construct a strategy y such that (P^y, Q^y) is exactly like (P^x, Q^x), except that:

$$p^y_{i^*j^*} = p^x_{i^*j^*} + p^x_{i^*j},$$
$$p^y_{i^*j} = 0,$$
$$q^y_{j^*i^*} = q^x_{j^*i^*} + q^x_{j^*i} - c_{j^*},$$
$$q^y_{j*i} = c_{j^*},$$

$\mathrm{tr}(P^xQ^y) = \mathrm{tr}(P^yQ^x) = \mathrm{tr}(P^xQ^x)$, so $u(y,x) = u(x,x)$. However, $\mathrm{tr}(P^yQ^y) - \mathrm{tr}(P^xQ^x) = p^x_{i^*j}(q_{j^*i} - c_{j^*}) > 0$, hence $u(y,y) > u(x,x)$. So x cannot be neutrally stable. □

Proof of Lemma 4. Suppose otherwise. Then there must be at least one row i in P^x with multiple positive entries that are smaller than e_i. This in turn implies that the corresponding column in Q^x has multiple maxima. So at least one of these positive entries in P^x must correspond to some q^x_{ji} with $c_j < q^x_{ji} < 1 + c_j$. According to Lemma 3, x cannot be neutrally stable then. □

Proof of Lemma 5. Suppose x is an NSS, and Q^x contains multiple column maxima. Let us say that the ith column of Q^x has multiple maxima. Let $j_1 = \min(\{j | q^x_{ji} = \max_{j'} q^x_{j'i}\})$, and let $j_2 \in \{j | q^x_{ji} = \max_{j'} q^x_{j'i}\} - \{j_1\}$. Obviously $j_1 < j_2$.

Since $q^x_{j_2i} \leqslant 1 + c_{j_2}$ and $c_{j_1} > c_{j_2}$, $q^x_{j_1i} < 1 + c_{j_1}$. So there must be some i_1 with $q^x_{j_1i_1} > c_{j_1}$. Hence both $p^x_{ij_1}$ and $p^x_{i_ij_1}$ are column maxima because of Lemma 1. As x is NSS, P^x is pure according to Lemma 4. Hence the j_1th column of P^x is a zero-column, so $p^x_{ij_1} = 0$. Hence there must be some $j^* \neq j_1$ such that $p^x_{ij^*} > 0$. Therefore $q^x_{j^*i}$ is a column maximum, so $q^x_{j^*i} = q^x_{j_1i}$.

We construct a y such that (P^y, Q^y) is exactly like (P^x, Q^x), except that

$$p^y_{ij_1} = p^x_{ij^*},$$
$$p^y_{ij^*} = 0,$$
$$q^y_{j_1i} = q^x_{j_1i} + q^x_{j_1i_1} - c_{j_1},$$
$$q^y_{j_1i_1} = c_{j_1}.$$

It follows that $\mathrm{tr}(P^yQ^x) = \mathrm{tr}(P^xQ^y) = \mathrm{tr}(P^xQ^x)$, but $\mathrm{tr}(P^yQ^y) - \mathrm{tr}(P^xQ^x) = p^x_{ij^*}(q^x_{j_1i_1} - c_{j_1}) > 0$. So $u(y,x) = u(x,x)$, and $u(x,y) < u(y,y)$, hence x is not neutrally stable. □

Proof of Lemma 6. If x^* is not an ESS, Q^x is not the unique best response to P^x. Hence P^x must contain columns with multiple column maxima. As x is neutrally stable, these must be zero-columns.

The minimal difference between a column maximum and the second largest entry in some column in P^x is e_n (the smallest positive entry in P^x; the non-maximal entries all equal 0). Hence if $\varepsilon < e_n/2$, the configuration of column maxima in P is preserved in y provided $\|x - y\| < \varepsilon$.

Let d_j be the difference between the column maximum in Q^x_j and the second largest entry within this column:

$$d_j \doteq \min_{i'}\left(\arg\max_i q^x_{ji} - q^x_{ji'}\right).$$

We assume that $\varepsilon < \min(\min_j d_j, e_n)/2$. In this way we ensure that for each strategy z within the ε-environment of x, each column maximum in P^z and Q^z corresponds to a column maximum in P^x or Q^x, respectively. Since P^x contains zero-columns, not each column maximum of P^x must correspond to a column maximum of P^z though.

Let y be a Nash equilibrium within the ε-environment of x. Suppose the j^*th column of P^x is a zero-column. Then none of the entries in the j^*th row of Q^x is a column maximum. Accordingly, the j^*th row of Q^y does not contain any column maxima either. Because y is Nash, the j^*th column of P^y must also be a zero-column. This in fact entails that $P^x = P^y$. Since x and y have the same configuration of column maxima, $BR(x) = BR(y)$.

Now suppose z is a best response to y. Then it is also a best response to x:

$$u(z,y) = u(z,x) = u(y,y) = u(x,x).$$

Because x is neutrally stable, $u(x,x) \geqslant u(z,z)$. Hence $u(y,y) \geqslant u(z,z)$. This establishes that y is neutrally stable.

Suppose w is a convex combination of x and y: $w = \alpha x + (1-\alpha)y$. w thus has zero-entries wherever both x and y have zero entries. Therefore w has the same configuration of column maxima as x and y. Therefore $BR(w) = BR(x)$. Due to the linearity of the utility function, $w \in BR(w)$, so w is a Nash equilibrium. By the same argument used for y, we can prove that w is also neutrally stable. □

Proof of Theorem 5. Suppose x is not a Nash equilibrium. Then either P^x is not a best response to Q^x or vice versa. Let us assume the former is the case. Then there is some pure sender strategy P with

$$\mathrm{tr}(PQ^x) > \mathrm{tr}(P^xQ^x).$$

For one pure receiver strategy Q in the support of x, it must be the case that

$$\mathrm{tr}(PQ) \geqslant \mathrm{tr}(PQ^x).$$

If Q^x is not a best response to P^x, the argumentation is analogous. So in sum we see that from each non-equilibrium strategy x we can reach a pure strategy y with $u(y,y) > u(x,x)$ with at most two better-response steps. If y is not a Nash equilibrium, we can continue this construction. This procedure will necessarily lead to a Nash equilibrium after finitely many steps because (a) the number of pure strategies is finite and (b) the fitness always strictly increases in each double-step.

So suppose x is a Nash equilibrium that does not belong to any ESSet. It is shown in Cressman (2003) that in a symmetrized asymmetric game, a set of Nash equilibria is an ESSet if and only if it is closed under mixed-strategy best replies. So if x does not belong to any ESSet, there must be a finite number of best-reply steps starting from x that lead to some point y that is not a Nash equilibrium.

So starting from some arbitrary point, I will always reach a pure strategy equilibrium with finitely many better-response steps. A pure strategy equilibrium x either belongs to some ESSet or I can reach another pure strategy equilibrium y from x with $u(y,y) > u(x,x)$. Since there are only finitely many pure strategies, it follows that I can reach some ESSet from any arbitrary point in finitely many steps. □

(Note that the argument does not depend on the specific properties of signaling games but applies to each symmetrized asymmetric partnership game.)

References

Akin, E., Hofbauer, J., 1982. Recurrence of the unfit. Math. Biosci. 61, 51–62.

Blume, A., Kim, Y.-G., Sobel, J., 1993. Evolutionary stability in games of communication. Games Econ. Behav. 5, 547–575.

Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50 (6), 1431–1451.

Cressman, R., 2003. Evolutionary Dynamics and Extensive Form Games. MIT Press, Cambridge, MA.

Darwin, C., 1871. The Descent of Man, and Selection in Relation to Sex. John Murray, London.

Grafen, A., 1990. Biological signals as handicaps. J. Theor. Biol. 144 (4), 517–546.

Greenberg, J., 1966. Universals of Language. MIT Press, Cambridge, MA.

Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J. (Eds.), Syntax and Semantics 3: Speech Acts. Academic Press, New York, pp. 41–58.

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

Hurd, P.L., 1995. Communication in discrete action-response games. J. Theor. Biol. 174 (2), 217–222.

Hurford, J.R., 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. Lingua 77, 187–222.

Huttegger, S.H., 2007. Evolution and the explanation of meaning. Philos. Sci. 74, 1–27.

Jäger, G., 2007. Evolutionary game theory and typology: a case study. Language 83 (1), 74–109.

Lewis, D., 1969. Convention. Harvard University Press, Cambridge, MA.

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

Nowak, M.A., Krakauer, D.C., 1999. The evolution of language. Proc. Nat. Acad. Sci. 96 (14), 8028–8033.

Pawlowitsch, C., 2007. Why evolution does not always lead to an optimal signaling system. Games Econ. Behav., in press, doi:10.1016/j.geb.2007.08.009.

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

Skyrms, B., 1996. Evolution of the Social Contract. Cambridge University Press, Cambridge, UK.

Spence, M., 1973. Job market signaling. Q. J. Econ. 87 (3), 355–374.

Swinkels, J., 1992. Stability and evolutionary stability: from Maynard Smith to Kohlberg and Mertens. J. Econ. Theory 57, 333–342.

Thomas, B., 1985. On evolutionarily stable sets. J. Math. Biol. 22, 105–115.

Trapa, P., Nowak, M., 2000. Nash equilibria for an evolutionary language game. J. Math. Biol. 41, 172–188.

van Rooij, R., 2004. Signalling games select Horn strategies. Linguist. Philos. 27, 493–527.

Wärneryd, K., 1993. Cheap talk, coordination and evolutionary stability. Games Econ. Behav. 5, 532–546.

Zahavi, A., 1975. Mate selection—a selection for a handicap. J. Theor. Biol. 53, 205–213.

Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge.