NASH EQUILIBRIA FOR AN EVOLUTIONARY LANGUAGE GAME

PETER E. TRAPA
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540

MARTIN A. NOWAK
Program in Theoretical Biology
Institute for Advanced Study
Princeton, NJ 08540

ABSTRACT. We study an evolutionary language game that describes how signals become associated with meaning. In our context, a language, L, is described by two matrices: the P matrix contains the probabilities that for a speaker certain objects are associated with certain signals, while the Q matrix contains the probabilities that for a listener certain signals are associated with certain objects. We define the payoff in our evolutionary language game as the total amount of information exchanged between two individuals. We give a formal classification of all languages, $L(P, Q)$, describing the conditions for Nash equilibria and evolutionarily stable strategies (ESS). We describe an algorithm for generating all languages that are Nash equilibria. Finally, we show that starting from any random language, there exists an evolutionary trajectory using selection and neutral drift that ends up with a strategy that is a strict Nash equilibrium (or very close to a strict Nash equilibrium).

1. INTRODUCTION

Theories for the evolution of language (Pinker & Bloom 1990, Bickerton 1995, Dunbar 1995, Pinker 1995, Deacon 1997, Hurford et al 1998) should address three points. First, they have to study the evolution of the simplest possible communication systems. Second, they should explore how natural selection can guide the transition from animal communication to human language and thereby explain the evolution of the very simplest properties of human language that are absent in animal communication. Third, they have to show how complex features of modern human language can evolve by natural selection (Niyogi & Berwick 1997).

The present paper is concerned with the first point. We present a formal analysis of an evolutionary game that was designed to capture the very first step of evolution toward a communication system (Hurford 1989, Nowak & Krakauer 1999, Nowak et al 1999). We explore how specific signals can evolve to become associated with specific objects. In the simplest evolutionary language game, we imagine a group of individuals (early hominids or other animals) that can produce a variety of signals. Signals can either be vocal, based on signs or a combination of both. We are interested in how such signals can evolve to become associated with specific objects. We use "object" in an extended sense to include other animals, inanimate objects, actions, events or concepts. In our context, an "object" is everything that can be referred to.

Date: October 25, 1999.
Support from the Leon Levy and Shelby White Initiatives Fund, the Florence Gould Foundation, the Alfred P Sloan Foundation and NSF grant DMS 97-29995 is gratefully acknowledged.

Denote by P the $n \times m$ matrix whose entries p_{ij} represent the probabilities that an individual will send signal j when wanting to transmit the information "object i". Let Q be the $m \times n$ matrix whose entries q_{ji} denote the probabilities that an individual will conceive of "object i" when receiving signal j. Hence, there are n objects and m signals. We have

$$\sum_{j=1}^{m} p_{ij} = 1 \qquad \text{and} \qquad \sum_{i=1}^{n} q_{ji} = 1.$$

We could also relax these constraints and allow for the possibility that an individual may not always signal when seeing an object or, vice-versa, may not always conceive of an object when receiving a signal. In this case, we would have

$$\sum_{j=1}^{m} p_{ij} \leq 1 \qquad \text{and} \qquad \sum_{i=1}^{n} q_{ji} \leq 1.$$

Consider two individuals that use languages $L(P,Q)$ and $L'(P',Q')$. We define the payoff as

$$F(L,L') = (1/2)\sum_{i=1}^{n}\sum_{j=1}^{m}(p_{ij}q'_{ji} + p'_{ij}q_{ji}).$$

The probability of transmitting object i from L to L' is given by $\sum_j p_{ij}q'_{ji}$. The payoff function sums these probabilities for all objects and then takes the average over the two situations, L signals to L' and L' signals to L. The specific assumptions of the payoff function are that sending and receiving yield equal payoffs and that all objects contribute the same amount to the payoff. Note that the payoff function is symmetric, $F(L,L') = F(L',L)$. A similar model was used by Hurford (1989) to study the evolution of the Saussyrian sign.

Following the notions of classical game theory, we can define a language L as a *strict Nash equilibrium* if $F(L,L) > F(L',L)$ for all languages $L' \neq L$. Furthermore, a language L is a *Nash equilibrium* if $F(L,L) \geq F(L',L)$ holds for all languages L'. In terms of evolutionary game theory (Maynard Smith 1982, Hofbauer & Sigmund 1998), a language L is an evolutionarily stable strategy (ESS) if $F(L,L) > F(L',L)$ holds for all $L' \neq L$, or if $F(L,L) = F(L',L)$ then we must have $F(L',L') > F(L,L')$.

Nash equilibria or ESS are fixed points of evolutionary dynamics: if a whole population uses a language that is either Nash or ESS then evolution will normally not change this situation. In other words, mutant strategies — as long as they are rare — cannot invade the population by natural selection. Strict Nash or ESS strategies are also protected against invasion by random drift. This is not the case for general Nash equilibria: if $F(L,L) = F(L',L) = F(L',L')$ then L and L' are neutral variants. There is no selection, but random drift can replace L by L'.

Two points are of general interest. First, while strict Nash equilibria or ESS are fixed points of evolutionary dynamics, there is no necessity that they are evolutionary attractors. Evolutionary dynamics can lead away from them (Nowak 1990, Hofbauer & Sigmund 1998). Second, strictly speaking they are only stable against invasion by mutants in arbitrarily low frequencies. Many reasonable phenomena, such as clustering, spatial effects, group selection or kin selection, may help mutants to overcome the *invasion barrier*. Despite this limitations, however, studying a new game always starts with an attempt to characterize all Nash or ESS strategies. And this is what we will do here.

Section 2 makes some preliminary remarks. In Section 3, we show that a language is a strict Nash equilibrium if and only if $n = m$, P is a permutation matrix and Q is the transpose of P. A permutation matrix has exactly one 1 entry in every row and column, while all other entries are zero. (The transpose of a matrix is the matrix that has rows and columns interchanged. Thus Q is the transpose of P, $Q = P^{tr}$, if $q_{ji} = p_{ij}$.)

Our result shows that the conditions for a strict Nash equilibrium are quite restrictive. Specifically we need that each object must be associated with one unique signal and vice versa. Hence, the number of objects and signals must be identical. For $n = 2$ there are two strict Nash equilibria:

$$P = Q^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad P = Q^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

For $n = 3$ there are 6 strict Nash equilibria. For a given number, n, of objects and signals there are $n!$ many languages which are strict Nash equilibria.

In Section 4, we show that a language, L, is ESS if and only if it is strict Nash. Hence, in the evolutionary language game, strict Nash and ESS languages are identical.

In Section 5,we show that a language, $L(P, Q)$ is a Nash equilibrium if the following, somewhat unexpected, conditions are fulfilled:

(i) in each column of P, entries are either zero or a specific number between 0 and 1;
(ii) in each column of Q, entries are either zero or a specific number between 0 and 1;
(iii) Q^T must be on the support of P

Conditions (i) and (ii) mean that in any one column of P (or Q) all non-zero entries must be identical. Thus, the entries of column j of the P matrix must be drawn from the set $\{0, p_j\}$, where p_j is any number less than or equal to $1/n$. The number p_j is fixed for column j, but different columns can have different numbers p_j. In complete analogy, the entries of column i of the Q matrix must be drawn from the set $\{0, q_i\}$. Condition (iii) means that q_{ji} is positive if and only if p_{ij} is positive.

Let us consider an example of a Nash equilibrium language (which is not strict Nash) for $n = m = 3$:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-x & x \end{pmatrix} \qquad Q = \begin{pmatrix} 1-y & y & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Here x and y are arbitrary numbers between 0 and 1. For both P and Q all rows sum to one. In this example, signal 1 refers to both object 1 and object 2, while signals 2 and 3 both refer to object 3. Thus we have homonymy (one signal has two meanings) and synonymy (two signals have the same meaning.) Let us check that all Nash equilibrium conditions hold: In column 1 of P, entries are either 0 or 1, in column 2 they are either 0 or $1 - x$, in column 3 they are either 0 or x. Similarly in column 1 of Q, entries are either 0 or $1 - y$, in column 2 they are either 0 or y, in column 3 they are either 0 or 1. Finally, we confirm that p_{ij} is positive if and only if q_{ji} is positive.

Here is an example for $n = 4$ and $m = 3$

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-x & x \\ 0 & 1-x & x \end{pmatrix} \qquad Q = \begin{pmatrix} 1-y & y & 0 & 0 \\ 0 & 0 & 1-z & z \\ 0 & 0 & 1-z & z \end{pmatrix}$$

Here x, y and z are arbitrary numbers in $(0, 1)$. Again we have homonymy: signal 1 denotes objects 1 and 2. There is also a case of "syno-homonymy": signals 2 and 3 both denote objects 3 and 4, but they have to do so in equal proportions. Thus, if object 3 induces signal

3 with probability x, then object 4 must also induce signal 3 with probability x. This is quite an unnatural constraint.

Note that the following P matrix cannot be part of a Nash equilibrium:

$$P = \begin{pmatrix} x & 1-x & 0 \\ 0 & 1-x & x \end{pmatrix}$$

While this P matrix fulfills the Nash criteria, it is not possible to construct a Q matrix that would fulfill the Nash criteria.

In summary, therefore, it seems that Nash equilibria have the following interesting properties (i) there can be homonymy: two or more objects are associated with the same signal, but none of the objects is also associated with another signal. (ii) there can be synonymy: one object is associated with two or more signals, but none of the signals are also associated with other objects (unless in the unlikely situations where *all* of the signals are associated in exactly the same proportion with another object) Section 5 also describes an algorithm for constructing all languages which are Nash equilibria.

In Section 6, we show that starting from any arbitrary language, $L_1(P_1, Q_1)$, there exists a sequence of languages, $L_1, L_2, ..., L_N$, such that for each i, L_i can replace L_{i-1} via selection or random drift, and so that the terminal language L_N is either: strict Nash, if $n = m$; or, if $n \neq m$, L_N is of the form $(P_N, Q_N = P_N^{\mathrm{tr}})$, where P_N is an *extended* permutation matrix. (Such a matrix is obtained by adding rows or columns of zeros to a permutation matrix of size $\min\{m, n\}$.) This latter situation can be understood in very simple terms. Suppose for definiteness that there are more objects than signals ($n > m$). The terminal language L_N then associates each signal to a unique object, and to such objects associates the unique signal describing it. The remaining objects are neither associated to a signal nor are there signals describing these objects. Said differently, the remaining objects are completely ignored. The benefit of neglecting them is that there is no ambiguity (either synonymy or homonymy) in the resulting language.

2. Preliminaries

We begin by defining a function (the *fitness function*)

$$F : (\{m \times n \text{ real matrices}\} \times \{n \times m \text{ real matrices}\})^2 \longrightarrow \mathbb{R}$$

via

$$F((P,Q),(P',Q')) = \frac{1}{2}(\mathrm{Tr}(PQ') + \mathrm{Tr}(P'Q)).$$

Given an $m \times n$ matrix A, we say that it is *weak row stochastic* if each of its rows consist of positive real numbers that sum to a numerical less than or equal to 1; if the value is exactly 1, the matrix is called *row stochastic*. An $m \times n$ matrix is called *binary* if all of its entries are either 0 or 1. A binary $m \times m$ matrix is called a *permutation matrix* if each row and column contains a unique nonzero entry. An $m \times n$ matrix is an *extended permutation matrix* if it is obtained from a permutation matrix by adding rows and columns consisting entirely of zeros. Note that any permutation matrix is row stochastic and that the transpose of a permutation matrix is again a permutation matrix. Similarly any extended permutation matrix is weak row stochastic, and its transpose is an extended permutation matrix.

A *weak* (m,n)*-language* $L = (P, Q)$ is a pair of weak row stochastic matrices, one of size $m \times n$, the other of size $n \times m$. Denote the set of (m, n)-languages by $\mathbf{L}_w^{m,n}$. If P and Q are actually row stochastic, then L is simply called a language. The set of all (m, n) languages is denoted $\mathbf{L}^{m,n}$. To avoid confusion with the modifier "weak" defined below, we refer to a weak language as a w-language

We say that $L \in \mathbf{L}_w^{m,n}$ is *strict Nash* if $F(L,L) > F(L,L')$ for all $L' \in \mathbf{L}_w^{m,n}$ with $L' \neq L$. L is called *Nash* if $F(L,L) \geq F(L,L')$ for all $L' \in \mathbf{L}_w^{m,n}$. Finally L is called an *evolutionary stable state* (briefly *ESS*) if L is Nash and for every L' with $F(L,L) = F(L,L')$, we have $F(L,L) > F(L',L')$. If the final strict inequality is relaxed to a weak one, L is called a *weak ESS*. ¿From the definitions, the following chain of implications is clear:

$$\text{strict Nash} \Longrightarrow \text{ESS} \Longrightarrow \text{weak ESS} \Longrightarrow \text{Nash}.$$

(The asymmetry in the usage of the modifiers "strict" and "weak" is unfortunate, but conventional.)

The obvious definitions apply to languages instead of w-languages. For instance we say that $L \in \mathbf{L}^{m,n}$ is Nash if $F(L,L) \geq F(L,L')$ for all languages $L' \in \mathbf{L}^{m,n}$; similar definitions apply for strict Nash, ESS, and weak ESS. But now a potential ambiguity arises. Conceivable there could exist a language $L \in \mathbf{L}^{m,n}$ which is Nash as a language, but *not* Nash as a w-language. A little checking of the definitions shows that this kind of example never arises:

Lemma 2.1. *A language $L \in \mathbf{L}^{m,n}$ is strict Nash (respectively, Nash, ESS, or weak ESS) as a language if and only if it is strict Nash (respectively, Nash, ESS, or weak ESS) as a w-language.*

It is helpful to introduce a little more notation. Given a weak $m \times n$ weak row stochastic matrix A, consider the set of real numbers obtained by multiplying A by some weak $n \times m$ row stochastic matrix and then taking the trace; more precisely, set

$$S_A = \{\mathrm{Tr}(AB) \mid B \text{ is any } n \times m \text{ weak row stochastic matrix}\}.$$

Note that S_A is bounded and, moreover, that S_A actually achieves a maximum. Indeed, for a fixed A, $\mathrm{Tr}(AB)$ is maximized by choosing B to be a binary row stochastic matrix satisfying

$$B_{ij} = 1 \implies A_{ji} \text{ is a maximal entry in the } i\text{th column of } A.$$

We can thus define the set $\mathbf{Max}_w(A)$ to consist of all $n \times m$ weak row stochastic matrices B so that $Tr(AB)$ is indeed maximal. Similarly we define $\mathbf{Max}(A)$ to consist of all $n \times m$ row stochastic matrices B so that $Tr(AB)$ is maximal. By the above discussion, $\mathbf{Max}(A)$ (and hence $\mathbf{Max}_w(A)$) is never empty.

The next lemma follows directly from the definitions, and quantifies the sense in which they are symmetric.

Lemma 2.2. *Suppose $L = (P,Q) \in \mathbf{L}_w^{m,n}$ is strict Nash (respectively, ESS, weak ESS, or Nash). Then $L^{\mathrm{op}} = (Q,P) \in \mathbf{L}_w^{n,m}$ is strict Nash (respectively, ESS, weak ESS, or Nash).*

We find it useful to recast some of the definitions.

Lemma 2.3. 1. *Suppose $L = (P,Q) \in \mathbf{L}_w^{m,n}$. Then*
   (a) *L is Nash, if and only if $P \in \mathbf{Max}_w(Q)$ and $Q \in \mathbf{Max}_w(P)$.*
   (b) *L is strict Nash if and only if $\{P\} = \mathbf{Max}_w(Q)$ and $\{Q\} = \mathbf{Max}_w(P)$; i.e. P is the unique element in $\mathbf{Max}_w(Q)$ and Q is the unique element in $\mathbf{Max}_w(P)$.*

2. *Suppose $L = (P,Q) \in \mathbf{L}^{m,n}$. Then*
   (a) *L is Nash, if and only if $P \in \mathbf{Max}(Q)$ and $Q \in \mathbf{Max}(P)$.*

(b) *L is strict Nash if and only if $\{P\} = \mathbf{Max}(Q)$ and $\{Q\} = \mathbf{Max}(P)$; i.e. P is the unique element in $\mathbf{Max}(Q)$ and Q is the unique element in $\mathbf{Max}(P)$.*

Proof. We prove only part (1), with part (2) being identical. Consider the "only if" assertion of part (a) of (1). Let $L = (P, Q)$ be Nash and suppose $P \notin \mathbf{Max}(Q)$. Choose $P' \in \mathbf{Max}(Q)$ and define $L' = (P', Q)$. Necessarily $L \neq L'$ and it is easy to check that $F(L, L') > F(L, L)$. This contradicts the Nash hypothesis on L, and hence we conclude $P \in \mathbf{Max}(Q)$. Lemma 2.2 implies that $Q \in \mathbf{Max}(P)$, and thus the "only if" part of (a) follows. For the converse in (a), suppose $P \in \mathbf{Max}(Q)$ and $Q \in \mathbf{Max}(P)$. Then it follows directly from the definitions that L is Nash. Part (b) is proved in the same way. □

3. Strict Nash Languages

Here is a characterization of strict Nash languages.

Theorem 3.1 (Strict Nash Languages). *A w-language $L \in \mathbf{L}_w^{m,n}$ is strict Nash if and only if $m = n$, P is a permutation matrix, and $Q = P^{tr}$.*

The rest of this subsection is devoted to proving the proposition. We start with a couple of easy lemmas.

Lemma 3.2. *Suppose $L \in \mathbf{L}_w^{m,n}$ is strict Nash. Then in fact $L \in \mathbf{L}^{m,n}$.*

Proof. This is obvious. □

Hence, for the remainder of this section, we can restrict our attention exclusively to languages (and not the more general setting of w-languages).

Lemma 3.3. *Suppose $L = (P, Q) \in \mathbf{L}^{m,m}$. Then $F(L, L) \leq m$, and equality occurs if and only if $P = Q^{\mathrm{tr}}$ is a permutation matrix.*

Proof. The first assertion is obvious, since P and Q are weak row stochastic. Now

$$\begin{aligned} F(L, L) = m &\iff \sum_j P_{ij} Q_{ji} = 1, \text{ for all } i \\ &\iff Q_{ji} = 1 \text{ for every } j \text{ such that } P_{ij} \neq 0, \end{aligned}$$

with the latter condition holding since P is row stochastic. But since Q is also row stochastic, the final condition is a contradiction unless P is a permutation matrix and $Q = P^{\mathrm{tr}}$. □

Next we prove the easy half of Theorem 3.1.

Lemma 3.4. *If $L = (P, Q)$ is an (m, m)-language with $P = Q^{\mathrm{tr}}$ a permutation matrix, then L is strict Nash.*

Proof. If $P = Q^{\mathrm{tr}}$ is a permutation matrix, then clearly we can assume that $P = Q = I_m$, the $m \times m$ identity matrix. (This amounts to rearranging the order that we assign the "sounds" in our language.) We want to show that if $L' = (P', Q') \in \mathbf{L}^{m,m}$ such that $F(L, L') \geq F(L, L)$, then $L = L'$. The condition $F(L, L') \geq F(L, L)$ together with Lemma 3.3 implies that we must have

$$\frac{1}{2}(\mathrm{Tr}(Q') + \mathrm{Tr}(P')) \geq m.$$

Since P' and Q' are row stochastic, their traces are bounded by m; so the last displayed equation implies that $\mathrm{Tr}(P') = \mathrm{Tr}(Q') = m$. Hence $P' = Q' = I_m$ and $L = L'$, as we wished to show. $\square$

Now we start to investigate the converse statement in Theorem 3.1. Here is a first step.

Lemma 3.5. *Suppose $L = (P, Q) \in \mathbf{L}^{m,n}$ is strict Nash. Then P is binary; i.e. each row of P contains a single nonzero entry (which is necessarily 1). Moreover fix i and let j be the unique index such that $P_{ij} = 1$; then Q_{ji} is the unique maximal entry in the ith column of Q.*

Proof. Assume $L = (P, Q) \in \mathbf{L}^{m,n}$ is strict Nash, and suppose P is not binary. Since P cannot be zero, there is a row — say the ith row — with more than one nonzero entry. Consider the ith column of Q, and select an index j so that Q_{ji} is maximal among the entries in the ith column of Q. (At this point, we do not know j is unique — below we will see it is however.) Define a new $m \times n$ row stochastic matrix $P' \neq P$ via

$$P'_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j \\ 0 & \text{if } k = i \text{ and } l \neq j \\ P_{kl} & \text{else} \end{cases}.$$

By construction $\mathrm{Tr}(P'Q) \geq \mathrm{Tr}(PQ)$, so we conclude that both P and P' are in $\mathbf{Max}(Q)$. This contradiction Lemma 2.3(b), and hence we conclude that P has at most one nonzero entry in each row.

Now consider the second assertion of the lemma. By the previous paragraph we may assume P is binary. Fix a row i, and let j be the unique index such that $P_{ij} = 1$. The analysis of the previous paragraph implies that Q_{ji} is a maximal entry in the jth column of Q. Suppose this maximal entry is not unique; i.e. suppose there is an $j' \neq j$ such that $Q_{ji} = Q_{j'i}$. Define a new matrix $P' \neq P$ by moving the nonzero entry in the ith row from the j slot to the j' slot; i.e. define

$$P'_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j' \\ 0 & \text{if } k = i \text{ and } l \neq j \\ P_{kl} & \text{else.} \end{cases}$$

By construction, $\mathrm{Tr}(PQ) = \mathrm{Tr}(P'Q)$ so both P and P' are elements of $\mathbf{Max}(Q)$. Lemma 2.3(b) again gives a contradiction. The proof is complete. $\square$

Lemma 3.6. *If $L = (P, Q) \in \mathbf{L}^{m,n}$ is strict Nash, then P is a binary matrix with no two nonzero entries in the same column. In particular, $m = n$.*

Proof. Suppose $L = (P, Q) \in \mathbf{L}^{m,n}$ is strict Nash. Lemma 3.5 implies that P is a binary matrix. Suppose that there are two nonzero entries in some column of P; i.e. suppose we can find indices i, i' and j such that $P_{ij} = P_{i'j} = 1$. Consider the entries Q_{ji} and $Q_{ji'}$. These are nonzero by Lemma 3.5. Form a new matrix Q' obtained from Q by adding a sufficiently small positive number to Q_{ji}, subtracting the same amount from $Q_{ji'}$, but otherwise leaving Q unchanged. Then $Q \neq Q'$ and $\mathrm{Tr}(PQ) = \mathrm{Tr}(PQ')$ by construction, so both Q' and Q are elements of $\mathbf{Max}(P)$. This contradicts Lemma 2.3(b) and the current lemma follows. $\square$

Lemmas 3.4 and 3.6 complete the proof of Theorem 3.1, except for the assertion that $Q = P^{\mathrm{tr}}$. This latter statement now follows from the last assertion of Lemma 3.5; so the proof of the theorem is complete. $\square$.

4. Evolutionary stable states

First we start with an easy observation.

Theorem 4.1 (Evolutionary Stable States). *A language $L = (P, Q) \in \mathbf{L}_w^{m,n}$ is an ESS if and only if it is strict Nash (and hence satisfies the conditions of Theorem 3.1).*

Proof. The "if" part of the assertion is immediate from the definitions. So, according to Lemma 2.3, to prove the proposition, we need only show that if L is an ESS, $\mathbf{Max}_w(P) = \{Q\}$ and $\mathbf{Max}_w(Q) = \{P\}$. Since L is an ESS, it is Nash, so Lemma 2.3(1) implies that Q is in $\mathbf{Max}_w(P)$ and P is in $\mathbf{Max}_w(Q)$. It remains only to show that there does not exist a $Q' \neq Q$ such that $Q' \in \mathbf{Max}_w(P)$. (The analogous statement for $\mathbf{Max}_w(Q)$ then follows from Lemma 2.2.)

Suppose there does indeed exist $Q' \neq Q$ such that $Q' \in \mathbf{Max}_w(P)$. Set $L' = (P, Q')$. Then $F(L, L) = F(L, L') = \mathrm{Tr}(PQ) = \mathrm{Tr}(PQ')$ and so $F(L, L) = F(L, L')$, contradicting the ESS assumption on L. Hence the theorem follows. □

The next example indicates that there exist Nash languages which are not weak ESS.

Example 4.2. Consider $L = (P, Q) \in \mathbf{L}^{2,2}$ specified by

$$P = Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} .$$

By Theorem 5.1 below (or by elementary computations), L is Nash. Consider $L' = (P', Q')$ defined by $P' = Q' = I_{2\times 2}$. Then $F(L, L) = F(L, L') = 1$, but $F(L', L') = 2$; so L is Nash, but not weak ESS.

As the example suggests, it is probably not terribly difficult to classify weak ESS languages, but the answer is probably too complicated to be of much interest.

5. Nash Languages

As a complement to the above analysis, we include a classification of Nash languages. The classification will eventually boil down to the following proposition. In its statement and proof, it is helpful to introduce a little terminology. Given any matrix A, we define the *support of* A, denoted $\mathrm{supp}(A)$, to be the set of those indices for which A has nonzero entries; more precisely,

$$\mathrm{supp}(A) := \{(i, j) \in \mathbb{N} \times \mathbb{N} \mid A_{ij} \neq 0\}.$$

Theorem 5.1 (Nash Languages). *Consider $L = (P, Q) \in \mathbf{L}^{m,n}$ and assume that no column of P (or Q) consist entirely of zero entries. Then L is Nash if and only if there exist real numbers $p_1, \ldots, p_n$ and $q_1, \ldots, q_m$ such that*

1. *For each j, the jth column of P has its entries drawn from $\{0, p_j\}$. Moreover, $P_{ij} = p_j$ if and only if $Q_{ji} = q_i$; i.e.* $\mathrm{supp}(P) = \mathrm{supp}(Q^{\mathrm{tr}})$.
2. *For each i, the ith column of Q has its entries drawn from $\{0, q_i\}$. Moreover, recall from (1) that* $\mathrm{supp}(Q) = \mathrm{supp}(P^{\mathrm{tr}})$.

Remark 5.2. The conditions on the columns of P and Q in the proposition may be a little confusing. We want each column to contain exactly one nonzero value, but this value may be repeated in several different entries. For instance

$$\begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$$

is the kind of column we allow, but

$$\begin{pmatrix} 3/4 \\ 1/2 \\ 0 \end{pmatrix}$$

is not.

Proof. Assume neither P nor Q has zero columns, and suppose that $L = (P, Q)$ is Nash. By Lemma 2.3, we have that $P \in \mathbf{Max}(Q)$ and $Q \in \mathbf{Max}(P)$. Fix i, and consider the ith column of Q, and let $j_1, \ldots, j_k$ denote those indices such that $Q_{ji} \neq 0$; i.e.

$$\{j_1, \ldots, j_k\} = \{j \mid Q_{ji} \neq 0\}.$$

Note that this set is nonempty, since we have assumed that Q has no zero columns. Now consider the ith row of P. We claim that

$$\{j \mid P_{ij} \neq 0\} \subset \{j_1, \ldots, j_k\}.$$

If not, any row stochastic matrix P' whose ith row satisfies this property (but whose other rows coincide with P) clearly has $\mathrm{Tr}(P'Q) > \mathrm{Tr}(PQ)$ thus contradicting that $P \in \mathbf{Max}(Q)$. Arguing column by column, we conclude

$$\mathrm{supp}(P) \subset \mathrm{supp}(Q^{\mathrm{tr}}).$$

Now reversing the roles of P and Q (or, alternatively, applying Lemma 2.2), we conclude that

$$\mathrm{supp}(Q) \subset \mathrm{supp}(P^{\mathrm{tr}}).$$

The last two displayed equations imply that $\mathrm{supp}(P) = \mathrm{supp}(Q^{\mathrm{tr}})$.

Hence if $L = (P, Q)$ is Nash (and if neither P nor Q contains zero columns), $\mathrm{supp}(P) = \mathrm{supp}(Q^{\mathrm{tr}})$. Next, we must now prove that under these hypothesis, each column of Q (and P) has its entries drawn from a two element set, one of whose elements is 0. Fix i and consider the ith column of Q. Suppose that there are indices $j_1, \ldots j_k$ such that each Q_{ij_l} is neither zero nor maximal amongst the elements in the first column of Q. (Here, of course, the indices j_l play a different role than they did in the preceding paragraph.) By the preceding paragraph $P_{ij_l} \neq 0$, for each index j_l. Consider any row stochastic matrix P' which is identical to P except in the i row; which satisfies $\mathrm{supp}(P') \subset \mathrm{supp}(P)$; but for which every $(i, j_l) \notin \mathrm{supp}(P')$. Then $\mathrm{Tr}(P'Q) > \mathrm{Tr}(PQ)$ contradicting the fact that $P \in \mathbf{Max}(Q)$. Hence we conclude that for each i, there exists a nonzero real number q_i such that the entries of the ith column of Q are drawn from $\{0, q_i\}$. The analogous statement for P follows from Lemma 2.2.

We have thus proved that if $L = (P, Q)$ is Nash (and no column of P or Q is zero), then conditions (1) and (2) of the proposition hold. To prove the converse, we must show that any pair of matrices (P, Q) satisfying the conditions of the proposition define a Nash language. Using arguments similar to those above, this is easy and routine by now. We leave the details to the reader. $\square$

Remark 5.3. We return to the setting of w-languages. The argument above shows that any Nash w-language $L = (P, Q)$ satisfies the conditions of the theorem — but the argument for the converse (which we omitted above) breaks down. For instance, the following w-language satisfies the conditions of the theorem, but fails to be Nash (as the reader may easily verify):

$$P = \begin{pmatrix} x & 1-x & 0 \\ 0 & x & 1-x \end{pmatrix} \qquad Q = \begin{pmatrix} 1/2 & 0 \\ 1/2 & 1/2 \\ 0 & 1/2 \end{pmatrix} .$$

In fact, it is easy to see that *any Nash w-language is in fact a Nash language*, i.e. satisfies the row stochastic condition. To see this suppose that $L = (P, Q)$ is a Nash w-language. We know it satisfies the conditions of the theorem. Suppose L is not in fact a language in the strong sense, i.e. suppose that some row (of P, say) has entries which sum to a value strictly less than one. Consider the matrix P' obtained by modifying P by increasing one of the nonzero entries in this row (but so that the sum of the entries is still less than 1). Then the support condition of the theorem, together with the construction of P' imply that $\mathrm{Tr}(P'Q) > \mathrm{Tr}(PQ)$, contradicting the Nash hypothesis on L. We thus conclude that the only Nash w-languages (with no zero columns) are those appearing in Theorem 5.1.

Next we record the following easy and useful corollary of Theorem 5.1; its proof is easy.

Corollary 5.4. *Let P be any binary matrix with no zero columns, and let Q be any row stochastic matrix with* $\mathrm{supp}(Q) = \mathrm{supp}(P^{\mathrm{tr}})$. *Then $L = (P, Q)$ is Nash.*

The next question to address is how to build all pairs (P, Q) described in the proposition. They can all be described by a relatively simple algorithm, which we now explain. (The reader is advised to consult the example below.) Begin with a row stochastic matrix $P^{(0)}$ subject to the first condition in (1) above; i.e. assume that for each j, the jth column of $P^{(0)}$ has its entries drawn from $\{0, p_j^{(0)}\}$. Consider the set of all matrices $A^{(0)}$ such $\mathrm{supp}(A^{(0)}) = \mathrm{supp}(P^{\mathrm{tr}})$. Impose the conditions that $A^{(0)}$ be row stochastic and that there exist nonzero real numbers $a_1, \ldots, a_m$ such that the entries in the ith column of $A^{(0)}$ are drawn from $\{0, a_i\}$. It requires a little subtle checking of the definitions to see that such an $A^{(0)}$ always exists — the restrictions do not "overdetermine" $A^{(0)}$ — so choose a particular matrix, say $Q^{(0)}$. Now if $(P^{(0)}, Q^{(0)})$ satisfy the conditions of Theorem 5.1, we have produced a Nash language and the algorithm stops. If not, then repeat the above prescription, applied this time to $Q^{(0)}$ (instead of $P^{(0)}$) to obtain an $m \times n$ row stochastic matrix $P^{(1)}$. Then check whether $(P^{(1)}, Q^{(0)})$ satisfy the conditions of the proposition. If so, the algorithm terminates; if not, it continues.

Note that after a finite number of steps, we must end up with a Nash language. The reason is easy: if $(P^{(k)}, Q^{(k)})$ is not Nash, then the number of entries in $P^{(k+1)}$ equal to some $p_j^{(k+1)}$ is strictly less than the number of entries in $P^{(k)}$ equal to some $p_j^{(k)}$. (Similar statements of course hold if $(P^{(k)}, Q^{(k-1)})$ is not Nash.) Hence the algorithm must terminate after a finite number of steps. Note also that if we $L = (P, Q)$ is Nash to begin with, and we start the algorithm with $P^{(0)} = P$, then $Q^{(0)}$ can be taken to be Q, and the algorithm will recover the pair (P, Q). Hence the algorithm produces all Nash languages for which no column of P or Q consists entirely of zeros.

Example 5.5. Consider

$$P^{(0)} = \begin{pmatrix} 0 & 1/8 & 3/4 & 1/8 \\ 1/8 & 0 & 3/4 & 1/8 \\ 1/8 & 0 & 3/4 & 1/8 \\ 0 & 1/8 & 3/4 & 1/8 \end{pmatrix}$$

Then $A^{(0)}$ (described above) is of the form

$$\begin{pmatrix} 0 & \alpha_1 & \alpha_2 & 0 \\ \beta_1 & 0 & 0 & \beta_2 \\ \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ \delta_1 & \delta_2 & \delta_3 & \delta_4 \end{pmatrix}$$

Imposing the additional requirements, we see that $Q^{(0)}$ is any matrix of the form

$$\begin{pmatrix} 0 & \alpha_1 & \alpha_2 & 0 \\ \beta_1 & 0 & 0 & \beta_2 \\ 0 & \alpha_1 & \alpha_2 & 0 \\ 0 & \alpha_1 & \alpha_2 & 0 \end{pmatrix}, \qquad \text{with } \alpha_1 + \alpha_2 = \beta_1 + \beta_2 = 1,\ \alpha_1\alpha_2\beta_1\beta_2 \neq 0;$$

or

$$\begin{pmatrix} 0 & a_1 & a_2 & 0 \\ b_1 & 0 & 0 & b_2 \\ b_1 & 0 & 0 & b_2 \\ b_1 & 0 & 0 & b_2 \end{pmatrix}, \qquad \text{with } a_1 + a_2 = b_1 + b_2 = 1,\ a_1 a_2 b_1 b_2 \neq 0;$$

In either case, the pair $(P^{(0)}, Q^{(0)})$ is not of the form required by Theorem 5.1, so we must iterate again. The result is that, in the first case, $P^{(1)}$ is any matrix of the form

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ \alpha_1' & 0 & \alpha_2' & \alpha_3' \\ \alpha_1' & 0 & \alpha_2' & \alpha_3' \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad \text{with } \alpha_1' + \alpha_2' + \alpha_3' = 1,\ \alpha_1'\alpha_2'\alpha_3' \neq 0;$$

and in the second case,

$$\begin{pmatrix} 0 & a_1' & a_2' & a_3' \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & a_1' & a_2' & a_3' \end{pmatrix}, \qquad \text{with } a_1' + a_2' + a_3' = 1,\ a_1'a_2'a_3' \neq 0;$$

Note that in either case, the number of nonzero entries in $P^{(0)}$ equal to some $p_j^{(0)}$ is strictly greater than the corresponding number for $P^{(1)}$. Finally, in each case, we check that $(P^{(1)}, Q^{(0)})$ satisfy the conditions of the proposition, and hence define a Nash language.

The final point we need to address is the possibility of P or Q having zero columns. This is clearly of practical interest, since such languages have unutilized sounds or indescribable objects. An easy result in this direction is Lemma 5.7 below, but first let us consider an example of how the conclusion of Theorem 5.1 can fail if zero columns are present.

Example 5.6. The (3,2)-language specified by

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ * & * \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

is Nash; here the $*$'s are arbitrary (subject to the row stochastic condition on P.) Note that the columns of P may have entries drawn from three values, not just two (as in the conclusion of Theorem 5.1). In fact, from this example generalizes easily: it isn't difficult to see that one can always obtain a Nash language by adding zero columns (and arbitrary rows) to a strict Nash language.

As another example, consider the following $(3, 2)$ language specified by

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Note that this example is obtained by adding a column of zeros to a Nash (not strict Nash) language; also see Example 4.2 below.

Lemma 5.7. *Suppose $L = (P, Q) \in \mathbf{L}^{m,n}$ is Nash, and assume further that the jth column of Q consists entirely of zero entries. Let P' be the $(m-1) \times n$ row stochastic matrix formed by removing the jth row of P, and let Q' be the $n \times (m-1)$ row stochastic matrix formed by removing the jth column of Q. Then $L' = (P', Q')$ is Nash. The analogous statement holds with the roles of P and Q interchanged.*

Proof. This is an easy consequence of Lemmas 2.2 and 2.3(1). □

Remark 5.8. The lemma holds verbatim for w-languages.

The above lemma, together with Theorem 5.1, reduces the classification of Nash languages to deciding when it is possible to *add* a column of zeros to a Nash language. Here is that answer — see Example 5.6 for illustrative examples.

Lemma 5.9. *Suppose $L = (P, Q) \in \mathbf{L}^{m,n}$ is a Nash language. Enumerate the maximal values of the columns of Q by $q_1, \dots, q_m$, and consider the $m \times (n+1)$ row stochastic matrix, say P', obtained by adding a column of zeros to P. Then there exist an $(n+1) \times m$ row stochastic matrix Q' such that $L' = (P', Q')$ is Nash if and only if $\sum_j q_j \geq 1$. The analogous statement holds if the roles of P and Q are interchanged.*

Proof. Consider the matrix Q' obtained from Q by adding a new row $Q'_{n+1,1}, \dots, Q'_{n+1,m}$. Clearly $Q' \in \mathbf{Max}(P')$, and the assertion of the lemma amounts to showing that we can choose the $Q'_{n+1,j}$ in such a way that $P' \in \mathbf{Max}(Q')$ if and only if $\sum_j q_j \geq 1$.

We claim that $Q' \in \mathbf{Max}(P')$ if and only if the $Q'_{n+1,j}$ are chosen so that each $Q'_{n+1,j} \leq q_j$. (Note that since Q' is row stochastic, the condition in the preceding sentence is possible if and only if $\sum_j q_j \geq 1$; so establishing the previous sentence proves the lemma.) In turn, this follows from an easy direct calculation. We omit the details. □

Remark 5.10. For w-languages the situation simplifies considerably, essentially because the parenthetical sentence in the proof above is not an issue. Retain the notations of the lemma, but this time assume only that (P, Q) is a w-language. Then there always exists a weak row stochastic matrix Q' so that the $(m, n+1)$ w-language (P', Q') is Nash: as in the proof above, we obtain Q' by adding a row to Q by such that each new entry $Q'_{n+1,j}$ is less than the corresponding q_j. (Since Q' is only weakly row stochastic, this is always possible, even if $\sum_j q_j < 1$.)

Lemmas 5.7 and 5.9, together with Theorem 5.1, thus complete the classification of Nash languages. Remarks 5.3, 5.8, and 5.10 complete the classification of Nash w-languages.

6. Evolutionary Dynamics

The point of this section is to show that through a combination of neutral drift and dominant invasion, any (m, n) w-language can be replaced by a w-language $L = (P, Q)$ where P and Q are extended permutation matrices.

We begin by recalling the notions of neutral drift and dominance. Suppose that we are given two w-languages L and L' satisfying one of the following conditions:

$$F(L', L') > F(L', L) > F(L, L); \text{ or}$$
$$F(L', L') = F(L', L) > F(L, L); \text{ or}$$
$$F(L', L') > F(L', L) = F(L, L).$$

Then we say that the language L' *dominates* L'. Similarly, if

$$F(L', L') = F(L', L) = F(L, L),$$

then we say that L can *neutrally drift* to L'. Here is the main result of this section.

Theorem 6.1. *Given any w-language $L \in \mathbf{L}_w^{m,n}$, there is a finite sequence of w-languages*

$$L = L_0 = (P_0, Q_0), L_1 = (P_1, Q_1), \cdots, L_N = (P_N, Q_N),$$

such that

1. *$P_N = Q_N^{\mathrm{tr}}$ is an extended permutation matrix; and*
2. *At each stage, either L_{i-1} is dominated by L_i or L_{i-1} can neutrally drift to L_i.*

Proof. The proof we give is entirely constructive. Suppose we are given a w-language $L_0 = (P_0, Q_0)$. If it is of the final form described in the proposition, clearly there is nothing to prove. So we can assume without loss of generality that P_0 is not an extended permutation matrix. Construct an extended permutation matrix P_1 by defining its ith row to be identically zero if the ith column of Q_0 is identically zero; if the ith column of Q is not identically zero, choose a value j such that the entry Q_{ji} is maximal in the ith column of Q, and define $P_{ij} = 1$, while $P_{ik} = 0$ for $k \neq j$. These requirements define an extended permutation matrix P_1.

Now set $Q_1 = Q_0$ and $L_1 = (P_1, Q_1)$. It is a trivial matter to check that indeed $F(L_1, L_1) \geq F(L_1, L_0) \geq F(L_0, L_0)$; so either L_0 can neutrally drift to L_1 or in fact L_0 is dominated by L_1.

Next we repeat the above construction with the roles of P and Q interchanged. Namely, we construct an extended permutation matrix Q_2 by requiring its ith row to either be zero (if the ith column of P_1 is zero) or to consist of unique nonzero entry 1 in the slot corresponding to a maximal entry of the ith column of P_1. We define $P_2 = P_1$ and $L_2 = (P_2, Q_2)$ and again it it easy to see that either L_1 can neutrally drift to L_2 or in fact L_1 is dominated by L_2.

A moment's thought (see the illustrative example below) shows that for some choices this process must eventually converge; that is, at some stage we have $L_M = L_N$ for all $M \geq N$ and, moreover, that $P_N = Q_N^{\mathrm{tr}}$. This proves the theorem. $\square$

Example 6.2. Consider the $(3,3)$-language $L_0 = (P_0, Q_0)$ defined by

$$P_0 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} \qquad Q_0 = P_0^{\mathrm{tr}} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} \quad .$$

(In fact, this is a Nash language — see the appendix for details.) Now we apply the algorithm of the above proof. We have several choices for P_1; here is one of them:

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad .$$

We set $L_1 = (P_1, Q_1 = Q_0)$ Again applying the algorithm reversing the roles of P and Q, we get either

$$Q_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad Q_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad .$$

In either case, set $L_2 = (P_2 = P_1, Q_2)$. Finally (again in either case) we compute $P_3 = Q_2^{\mathrm{tr}}$, and the process stops at $L_3 = (P_3 = Q_2^{\mathrm{tr}}, Q_3 = Q_2)$.

Note that at the very first step we would have chosen

$$P_1 = I_{3\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad ,$$

and the algorithm would have terminated at the next step with $L_2' = (I_{3\times 3}, I_{3\times 3})$. This concludes the example.

Note that in the above example L_2' in fact dominates L_3. It is easy to see that this is a general phenomenon: if the number of nonzero entries in P_N (or Q_N) is strictly less than $\min\{m, n\}$, we can modify P_N and Q_N by adding appropriately placed 1's, and the resulting modification dominates $L_N = (P_N, Q_N)$. This leads to the following sharpening of the above theorem.

Theorem 6.3. *In the conclusion of Theorem 6.1, we may also impose the restriction that the number of nonzero entries in P_N (or Q_N) is equal to the minimum of m and n.*

For simulations of specific evolutionary dynamics in this language game we refer to Nowak et al (2000).

7. Conclusion

In this paper we have characterized Nash equilibria and ESS in the evolutionary language game. We have shown that a language $L(P, Q)$ is a strict Nash equilibrium or an ESS if and only if $n = m$, P is a permutation matrix and Q is the transpose of P. Furthermore, we have characterized all strategies which are Nash, but not strict Nash. Such strategies exist for $n \neq m$ and allow the same signal to be used for different objects (homonymy) and the same object being described by different signals (synonymy). We have given an algorithm to construct all Nash equilibria. Finally, we observe that starting from any language there exists an evolutionary trajectory (using selection and drift) that ends up with a strict Nash language (for $n = m$) or an extended permutation matrix for $n \neq m$.

The current analysis should be extended in many different directions. We assumed that correct communication about all objects has the same contribution to the payoff, but more realistically objects have different values. It may be important to have a word for 'lion' but less important to have a word for 'ant'. Furthermore, we assumed that only correct communication leads to a payoff, while mistaking one object for another yields zero payoff. Some mistakes may in fact be more costly than others. Thus a more general framework should contain a matrix that describes the (positive or negative) payoff that is associated with misunderstanding between individuals.

References

Bickerton, D. Language and Species. (University of Chicago Press, Chicago 1990)

Deacon, T. The Symbolic Species (Allen Lane, The Penguin Press, London, 1997).

Dunbar, R. Grooming, Gossip and the Evolution of Language (Harvard University Press, Cambridge, MA, 1997).

Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics, Cambridge Univ Press

Hurford J R. Biological evolution of the saussurean sign as a component of the language acquisition device. Lingua 77, 187-222 (1989)

Hurford J R, Studdert-Kennedy M, Knight C (1998) Approaches to the Evolution of Language, Cambridge University Press.

Maynard Smith, J (1982) Evolution and the Theory of Games (Cambridge University Press, Cambridge)

Nash J (1996) Essays on game theory. Cheltenham, UK: Elgar.

Niyogi P, Berwick R C (1997) A dynamical systems model for language change, Complex Systems 11, 161-204

Nowak M A (1990) An evolutionarily stable strategy may be inaccessible. J theor Biol 142, 237-241

Nowak M A, Krakauer D C (1999) The evolution of language, Proc Natl Acad Sci USA, 96. 8028-8033

Nowak M A, Plotkin J, Krakauer D C (1999) The evolutionary language game, J theor Biol, 200, 147-162

Pinker, S & Bloom, P (1990). Natural language and natural selection. Behavioural and Brain Sciences. 13, 707-784.

Pinker S (1995) The language instinct, Penguin