



## The Evolutionary Dynamics of the Lexical Matrix

NATALIA L. KOMAROVA\* AND MARTIN A. NOWAK

Institute for Advanced Study,  
Einstein Drive,  
Princeton,  
NJ 08540, U.S.A.

The lexical matrix is an integral part of the human language system. It provides the link between word form and word meaning. A simple lexical matrix is also at the center of any animal communication system, where it defines the associations between form and meaning of animal signals. We study the evolution and population dynamics of the lexical matrix. We assume that children learn the lexical matrix of their parents. This learning process is subject to mistakes: (i) children may not acquire all lexical items of their parents (incomplete learning); and (ii) children might acquire associations between word forms and word meanings that differ from their parents' lexical items (incorrect learning). We derive an analytic framework that deals with incomplete learning. We calculate the maximum error rate that is compatible with a population maintaining a coherent lexical matrix of a given size. We calculate the equilibrium distribution of the number of lexical items known to individuals. Our analytic investigations are supplemented by numerical simulations that describe both incomplete and incorrect learning, and other extensions.

© 2001 Society for Mathematical Biology

### 1. INTRODUCTION

Humans use words as a basic unit of communication. To a first approximation, words are arbitrary symbols with conventionally attached meanings. Knowing a word means remembering both its sound and its meaning; language can be viewed as a code between the two (Sperber and Wilson, 1995). The lexical matrix,  $A$ , specifies the association between word meaning and word form (Hurford, 1989; Miller, 1996). Each column of the lexical matrix corresponds to a particular word meaning (or concept), each row corresponds to a particular word form (or word image). In the Saussurean terminology of arbitrary sign, the lexical matrix provides the link between signifié and signifiant (Saussure, 1983).

A lexical matrix is a convenient description of arbitrary relations between discrete forms and discrete concepts. It is a central component of all human languages, as well as protolanguages (Bickerton, 1990), holistic protolanguage (Wray, 1998, 2000) and non-syntactic forms of communication (Hallowell, 1960). In one form or another, a lexical matrix has been a part of hominid/human language for the past

---

\*Also at: Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, U.K.

four million years (Brandon and Hornstein, 1986; Lieberman, 1992; Pinker, 1995; Deacon, 1997). Furthermore, a simple lexical matrix is at the basis of any animal communication system, where it defines the relation between animal signals and their specific meanings (Cheney and Seyfarth, 1990; Macedonia and Evans, 1993; Hauser, 1997; Smith, 1977).

The goal of this paper is to study the evolutionary dynamics of the lexical matrix. We will calculate the conditions for the evolution and maintenance of a lexical matrix in a population of individuals. Evolving a coherent lexicon is not an easy task, because the correspondence prescribed by the lexical matrix is entirely arbitrary in the sense that the word meaning normally cannot be derived from the word form by any rule-based procedure. This arbitrariness gives rise to the problem of coherence. Namely, if different individuals happen to assign different word forms to the same word meaning (or vice versa), then how can any useful information be transferred?

Other researchers have worked on similar questions. Steels (1996) models a set of individuals who in the beginning use random associations between word meanings and word forms. Pairs of individuals get in contact and play a 'language game' (i.e., exchange signals to find out if they 'match'). If the game fails, a new (random) association is made. As a result, after a number of time-steps, the population converges to a unique association matrix, i.e., individuals understand each other perfectly. Steels and Vogt (1997) report on experiments with physically embodied robots which develop a shared vocabulary through interaction. After each unsuccessful interaction the robots improve their vocabulary. A selection mechanism is embedded in the dynamics of associations. There is a positive reinforcement if the association is used by many agents. If the association is not shared by many others it will eventually be replaced by a more successful one. This leads to a disappearance of shared meanings, or ambiguities, in the language, and a convergence to a unique lexicon. Cangelosi and Parisi (1998) have studied a computer model of syntax development in communicating neural networks, using an evolutionary framework. There, individuals who did not form 'working' associations were considered 'less fit', whereas those who formed coherent associations were rewarded. The score was calculated at the end of a discrete generation's life-span, which resulted in a reproductive advantage of more fit individuals. Since associations were transferred 'genetically' to following generations, a coherent communication system developed after several iterations.

In the present paper we consider the dynamics of lexical matrices in an evolutionary framework (Aoki and Feldman, 1987, 1989; Nowak and Krakauer, 1999; Nowak *et al.*, 1999). Individuals talk to each other. Whenever they succeed at transferring information, they receive a payoff. The payoff of this evolutionary language game is interpreted as fitness. Individuals with a higher payoff produce more offspring, who will learn their language. The assumption that language performance affects biological fitness is crucial in this model. Otherwise language cannot evolve as an evolutionarily stable strategy (Nowak, 2000).

The communicative ability of individuals which translates into their biological fitness is determined by the lexical matrices. Successful lexical matrices are characterized by the following two factors: (i) they are more informative, i.e., more concepts are uniquely paired with specific words, and (ii) they are shared by a larger fraction of individuals, which makes communication possible with a larger number of people in the group. More successful matrices have a higher probability to be learned by others.

The learning process is probabilistic and subject to mistakes. There are two kinds of mistakes. Children may not acquire all the lexical entries of their parents. We call this ‘incomplete learning’. Furthermore, children may mis-hear certain words or misinterpret their meanings. As a consequence, they form entries in their lexical matrix that differ from their parents’ entries. We call this ‘incorrect learning’.

In this paper we develop an analytic framework that deals with incomplete learning. We perform computer simulations that also include incorrect learning, but we do not have a complete analytic understanding for this type of mistake. This is a challenge for subsequent work.

The main analytical results for the incomplete learning model are as follows. Non-ambiguous lexical matrices, which are defined by a one-to-one correspondence between word meanings and word forms, are evolutionarily stable. Because of incomplete learning, some individuals only acquire a subset of the entries of the lexical matrix. The learning accuracy, which is the ability to copy an entry of the teacher’s matrix correctly, determines the number of words that can be kept stably in a population. We find the minimum requirements of the lexicon acquisition device that are compatible with a population of individuals evolving a coherent lexical matrix of a certain size. Similarly, given the learning accuracy of individuals, we find the lexical capacity of the population, i.e., the maximum number of lexical items in the collective vocabulary, and describe the distribution of the number of word-meaning associations that people know.

The model that we can study analytically has the following simplifying assumptions: (i) it contains incomplete learning, but no incorrect learning; (ii) the lexical matrix has binary entries, which means that associations between word forms and word meanings do not vary in strength, they are either there or not; and (iii) each individual learns the lexical matrix from one other individual, the parent. We include computer simulations to relax some of these assumptions. In particular, we show that if the probability to create an incorrect association while learning the lexicon is small, then the attractors of the system can be described using the analytical prediction of the simple incomplete learning model. Another extension of the model that we consider is viewing the dynamics as a stochastic process. It turns out that the stochastic system performs transitions between stable fixed points of the corresponding deterministic system, thus leading to spontaneous changes in the lexicon (see also Steels and Kaplan, 1998). Between transitions, the system is found in a quasi-stationary state which (under certain assumptions) is well described by our analytical results.

This paper is organized as follows. In Section 2 we formulate the basic incomplete learning model. In Section 3 we consider a reduced system (language in the absence of synonyms and homonyms), find stable equilibria and study the bifurcation diagram. Section 4 extends this result to the full system: the stable fixed points of the reduced system remain stable against perturbations by a general lexical matrix. Computer simulations for extended models which include incorrect learning and other more realistic assumptions are reported in Section 5. A discussion is presented in Section 6.

## 2. MODEL DESCRIPTION

Here we describe the basic incomplete learning model which will be solved analytically. It contains many simplifying assumptions; the effects of relaxing some of these assumptions are discussed in Section 5.

**2.1. A binary matrix and a fitness function.** Each individual is characterized by a lexical matrix,  $A$ , which links referents to signals. If there are  $n$  referents and  $m$  signals, the  $A$  is an  $n \times m$  matrix. We assume that the entries,  $a_{ij}$ , are either 0 or 1. If  $a_{ij} = 1$  then referent  $i$  is linked to signal  $j$ . If  $a_{ij} = 0$  then referent  $i$  is not linked to signal  $j$ . Each referent can be linked to several signals, and in turn each signal may denote several referents. There can also be referents not linked to any signal, and there can be signals not denoting any referent. The total number of  $A$  matrices of size  $n \times m$  is  $2^{nm}$ . More generally, non-negative integer-valued  $a_{ij}$  could denote the strength of association between referent  $i$  and signal  $j$ . We sacrifice this possibility in the present section, but gain in return a framework which is amenable to detailed mathematical analysis.

Next, we calculate fitness associated with communication. Let  $\tilde{p}_{ij}$  denote the probability that an individual will use signal  $j$  when wanting to communicate about referent  $i$ . Conversely, let  $\tilde{q}_{ji}$  denote the probability that an individual will interpret signal  $j$  as referring to referent  $i$ . We have

$$\tilde{p}_{ij} = a_{ij} / \sum_{j=1}^m a_{ij}; \quad \tilde{q}_{ji} = a_{ij} / \sum_{i=1}^n a_{ij}. \quad (1)$$

The denominators have to be greater than 0, otherwise simply take  $\tilde{p}_{ij}$  or  $\tilde{q}_{ji}$  as 0. A language is completely defined by its association matrix,  $A_I$ , or by the matrices  $\tilde{p}^{(I)}$  and  $\tilde{q}^{(I)}$ .

Let us now consider two individuals  $I$  and  $J$  with languages  $A_I$  and  $A_J$ . We define the payoff, or fitness function, for  $I$  communicating with  $J$  as

$$F(A_I, A_J) = (1/2) \sum_{i=1}^n \sum_{j=1}^m (\tilde{p}_{ij}^{(I)} \tilde{q}_{ji}^{(J)} + \tilde{p}_{ij}^{(J)} \tilde{q}_{ji}^{(I)}). \quad (2)$$

The term  $\sum_{j=1}^m \tilde{p}_{ij}^{(I)} \tilde{q}_{ji}^{(J)}$  gives the probability that individual  $I$  will successfully communicate ‘referent  $i$ ’ to individual  $J$ . This probability is then summed over all referents and averaged over the situation where individual  $I$  signals to individual  $J$  and vice versa [see also Hurford (1989), Nowak and Krakauer (1999)].

There is a natural correspondence between all binary matrices and the binary numbers from 0 to  $2^{mn} - 1$ , obtained by reading each matrix row by row from left to right. These numbers can be recast in the decimal form as natural numbers (we can get rid of the zero by shifting everything by 1). The function  $F$  can then be viewed as a mapping from  $\mathbf{N}^2$  to  $\mathbf{Z}$ , i.e., a rational-valued matrix. For example, in the case of  $n = m = 2$ , we have 16 possible  $A$  matrices and find nine discrete payoff values. For  $n = m = 3$  there are 512 different  $A$  matrices that give rise to 93 discrete payoff values.

**2.2. Deterministic modeling.** Let us define the population dynamics for the evolution of the lexical matrix. Denote by  $x_I$  the frequency of individuals with association matrix  $A_I$ . Consider the following system of ordinary differential equations:

$$\dot{x}_I = \sum_J f_J x_J Q_{JI} - \phi x_I. \tag{3}$$

The summation runs over all possible  $A$  matrices, that is  $1 \leq J \leq 2^{mn}$ . The fitness of individuals  $J$  is given by

$$f_J = \sum_I F(A_J, A_I) x_I. \tag{4}$$

This assumes that individual  $J$  talks to individuals  $I$  with probability  $x_I$ . The quantity  $f_J$  denotes the expected payoff of all interactions of individual  $J$ . The average fitness of the population is given by

$$\phi = \sum_I f_I x_I. \tag{5}$$

For equation (3), the total population size is constant by definition of  $\phi$ . We set  $\sum_I x_I = 1$ . The parameter  $Q_{JI}$  denotes the probability that someone learning from an individual with  $A_J$  will end up with  $A_I$ . Thus  $Q_{II}$  denotes the probability of correct learning, while  $Q_{JI}$  with  $I \neq J$  denotes learning with mistakes. We will at first assume that the learner can miss certain associations, but cannot form new associations, i.e., the incomplete learning scenario. If the teacher has  $a_{ij} = 1$ , then the learner will have  $a_{ij} = 1$  with probability  $q$  and  $a_{ij} = 0$  with probability  $1 - q$ . If the teacher has  $a_{ij} = 0$ , then the learner will always have  $a_{ij} = 0$ . All entries are learned independently from each other. The parameter  $q$  is called the *learning fidelity*, or the *learning accuracy*, and is the only free parameter of the system.

Equation (3) is an extension of the quasispecies equation (Eigen and Schuster, 1979). Standard quasispecies theory has constant fitness values, whereas equation (3) has frequency dependent fitness values. Thus equation (3) can be considered to be at the interface between quasispecies theory and evolutionary game dynamics (Nowak, 2000; Nowak *et al.*, 2001). Individuals that communicate well receive a high payoff which translates into reproductive success: successful individuals produce more children who learn their lexical matrix. It is essential that language performance contributes to biological fitness. Otherwise mutants who do not communicate at all will not be selected against.

### 3. NON-AMBIGUOUS LANGUAGES

Let us assume that there are  $n$  referents and  $n$  signals that can be used in the language. In this section, we will restrict our analysis to *non-ambiguous* languages where only one or zero signals can be used for each referent, and each signal refers to only one referent. This means that there are no synonyms or homonyms in the language. We will further assume that *all* the signals (and their referents) used in all the languages, form a non-ambiguous language (that is, the union of all the languages is a non-ambiguous language). Note that since no erroneous entries can be made in the process of learning, the dynamics cannot create new synonyms or homonyms. Therefore, if the union of all languages is non-ambiguous, it will remain so for all generations in the future.

The assumption of this section (that the meanings are pre-given) is not a limitation. It is merely a way to find some solutions of the full system. We reduce equation (3) to a simpler set of ODEs whose phase space we can analyse. In Section 4 we will prove that the stable fixed points found by this method remain stable in the full, unrestricted, system.

Furthermore, we emphasize that even though in the next subsections the signals are uniquely paired with referents, it does not violate the concept of words being ‘arbitrary signs’. The pairing we assume here is absolutely arbitrary and the solutions we find for a particular set of associations will hold for *any* other non-ambiguous set of associations, exactly because of this fundamental arbitrariness (see Section 3.4).

**3.1. Fixed points.** The lexical matrix of the perfect language is an  $n \times n$  permutation matrix. By proper renumbering of signals and referents it can be written as the identity matrix. The rest of the possible non-ambiguous  $A$  matrices can have  $n - 1$  or fewer diagonal entries. Therefore, we can characterize every language  $I$  by a sequence of  $n$  zeros and ones which are the diagonal entries of the corresponding  $A$  matrix,  $S^I = (s_1^I, \dots, s_n^I)$ ,  $s_j^I \in \{0, 1\}$ , where  $1 \leq j \leq n$  and  $1 \leq I \leq 2^n$ . The fitness function in the case of non-ambiguous languages is just the conventional inner product of the corresponding sequences,  $F(A_I, A_J) = (S^I, S^J)$ . We will say

that language  $I$  has rank  $k$  if  $R(I) \equiv \sum_{j=1}^n s_j^I = k$ . All languages can be divided into  $n + 1$  non-intersecting classes, each class has a different rank  $k$  and contains  $C_n^k = n!/(k!(n - k)!)$  languages. We will use lower-case subscripts to enumerate classes, and capital letters are reserved for numbering all the languages. The transition probabilities  $Q_{IJ}$  are defined as

$$Q_{IJ} = \begin{cases} (1 - q)^{d(S^I, S^J)} q^{R(J)}, & \text{if } s_m^I - s_m^J \geq 0, \quad 1 \leq m \leq n, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where  $d(S^I, S^J) = \sum_{i=1}^n |s_i^I - s_i^J|$  is the Hamming distance between the two sequences. Here  $q$  is learning fidelity, i.e., the probability to memorize one association correctly. The condition in rule (6) simply means incomplete learning: the new language,  $J$ , does not contain new associations with respect to the language  $I$ ; learners can only lose associations. Let us enumerate the languages inside class  $k$  by the index  $\alpha_k$ ,  $1 \leq \alpha_k \leq C_n^k$ . We denote the fraction of people who speak language  $\alpha_k$  of class  $k$  as  $x_k^{(\alpha_k)}$ . The total number of people whose language belongs to class  $k$  is  $x_k = \sum_{\alpha_k=1}^{C_n^k} x_k^{(\alpha_k)}$ .

Our goal is to find some steady-state solutions of system (3) explicitly, and then study their stability. Let us assume that within each class, every language has an equal share, i.e.,

$$x_k^{(1)} = x_k^{(2)} = \dots = x_k^{(C_n^k)}. \quad (7)$$

Now we can define ‘coarse’ transition and fitness matrices. The probability to go from class  $j$  to class  $m$  is given by

$$Q_{jm} = \begin{cases} C_j^m (1 - q)^{j-m} q^m, & j \geq m, \\ 0, & j < m, \end{cases} \quad (8)$$

and the mutual fitness of two classes  $m$  and  $j$  is found to be

$$F_{mj} = mj/n. \quad (9)$$

This formula guarantees that, if (7) is satisfied, then

$$F_{mj} x_m x_j = \sum_{\alpha_m, \alpha_j} F(A_m^{\alpha_m}, A_j^{\alpha_j}) x_m^{(\alpha_m)} x_j^{(\alpha_j)}, \quad (10)$$

where languages  $A_m^{(\alpha_m)}$  ( $A_j^{(\alpha_j)}$ ) belong to class  $m$  (class  $j$ ), and the summation is performed over all languages within the corresponding classes. Formula (9) is derived from equation (10) by means of some combinatorics. Now we can write down system (3) under assumption (7):

$$\dot{x}_m = \left( q^m \sum_{j=m}^n C_j^m (1 - q)^{j-m} \frac{j}{n} x_j - \frac{\langle m \rangle}{n} x_m \right) \langle m \rangle, \quad 1 \leq m \leq n, \quad (11)$$

where we have introduced the notation

$$\langle m \rangle \equiv \sum_{k=1}^n kx_k \quad (12)$$

for the average number of associations used by individuals. In the derivation of (11) we used the fact that  $\phi = \sum_{i,j} F_{ij}x_ix_j = \langle m \rangle^2/n$ , which follows from equation (9). We do not include the equation for class 0 because it follows from the other  $n$  equations and the conservation of the number of people. Fixed points of system (11) can be found. If  $\langle m \rangle \neq 0$ , we can multiply system (11) by  $n/\langle m \rangle$ , set the left-hand side to zero and obtain an eigen-system with a triangular matrix. The eigenvalues are:

$$\langle m \rangle = (n-l)q^{n-l}, \quad 0 \leq l \leq n-1. \quad (13)$$

We will refer to the fixed point with  $l = 0$  as the *optimal solution*, and the fixed points with  $1 \leq l \leq n$  as *sub-optimal solutions*. Note that the case  $l = n$  corresponds to the situation with  $\langle m \rangle = 0$ , i.e., nobody has any associations. It is shown in Appendix A that sub-optimal solutions are unstable for all values of  $q$ . Here we will analyse the optimal solution corresponding to

$$\langle m \rangle = nq^n. \quad (14)$$

Let us denote the optimal solution as  $x_i = x_i^{\text{opt}}$ ,  $1 \leq i \leq n$ . It can be defined recursively as

$$x_n^{\text{opt}} = n^n/n! \prod_{j=0}^{n-1} (q^{n-j} - j/n), \quad (15)$$

$$x_m^{\text{opt}} = (q^{n-m} - m/n)^{-1} \sum_{j=m+1}^n C_j^m (1-q)^{j-m} j/n x_j^{\text{opt}}, \quad 1 \leq m \leq n-1, \quad (16)$$

$$x_0^{\text{opt}} = 1 - \sum_{j=1}^n x_j^{\text{opt}}. \quad (17)$$

The requirement  $0 \leq x_j \leq 1$  implies that

$$q \geq 1 - 1/n, \quad (18)$$

which is the existence condition for the optimal solution. Solution (15), (16) for  $n = 10$  is shown in Fig. 1. Note that for  $q \approx 1$ ,  $x_j^{\text{opt}} \propto (1-q)^{(n-j)}$ , which means that at  $q = 1$ ,  $x_n^{\text{opt}} = 1$ , and near  $q = 1$ , the fraction of people speaking language  $j$  decreases rapidly as  $j$  decreases. The average fitness,  $\phi$ , corresponding to this solution, is

$$\phi = nq^{2n}. \quad (19)$$

This quantity has the meaning of the effective lexicon size of the population. It gives the expected number of signals that any two people will have in common.

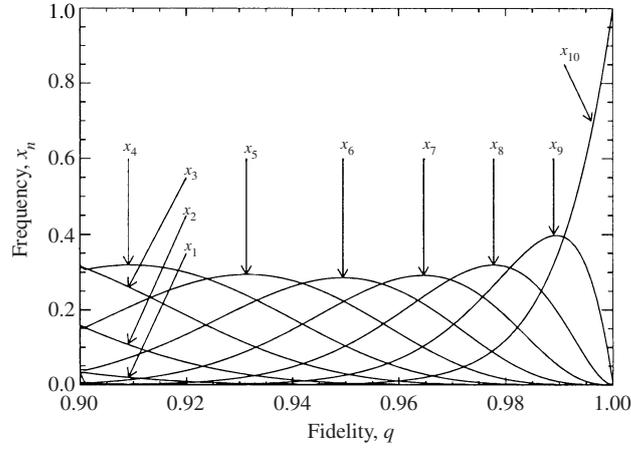


Figure 1. The components of the optimal solution for  $n = 10$ , as functions of the learning fidelity,  $q$ , in the region of its existence ( $0.9 \geq q \geq 1$ ). Each  $x_i^{\text{opt}}$  denotes the fraction of people who know  $i$  signals,  $1 \leq i \leq n$ .

**3.2. The optimal distribution of the number of associations.** Recursive formulas (15)–(17) give little insight into what the optimal distribution looks like. In this section we find an analytical expression for this distribution in the limit of large  $n$ .

Solution (15)–(17) has a maximum at  $m = \langle m \rangle$  (by construction, the average number of associations that people know is  $\langle m \rangle = nq^n$ ). The standard deviation can be found in the following way. From (11) we have

$$x_m^{\text{opt}} = q^{m-n} \sum_{j=m}^n C_j^m (1-q)^{j-m} j/n x_j^{\text{opt}}. \quad (20)$$

By multiplying both sides by  $m$  and performing the summation over all  $0 \leq m \leq n$ , we have  $\langle m \rangle = nq^n$  on the left-hand side. In the expression on the right-hand side we change the order of summation in  $j$  and  $m$  to obtain

$$q^{-n} \sum_{j=0}^n \left( \sum_{m=0}^j m C_j^m q^m (1-q)^{j-m} \right) j/n x_j^{\text{opt}} = \sum_{j=0}^n j^2 x_j^{\text{opt}} q^{-(n-1)}/n, \quad (21)$$

where the expression in brackets is the expectation of a binomial distribution. Equating both sides we find that  $\langle m^2 \rangle \equiv \sum_{m=0}^n m^2 x_m^{\text{opt}} = n^2 q^{2n-1}$ . The standard deviation is given by  $\sigma^2 = \langle m^2 \rangle - \langle m \rangle^2$ :

$$\sigma^2 = n^2 q^{2n-1} (1-q). \quad (22)$$

Recall that in the equilibrium solution,  $(x_0^{\text{opt}}, x_1^{\text{opt}}, \dots, x_n^{\text{opt}})$ ,  $x_k^{\text{opt}}$  stands for the fraction of people who have exactly  $k$  associations (that is, any associations). Let us introduce the variable  $z_m$  which is the fraction of people who know  $m$  given

associations. At the optimal solution, this quantity does not depend on which  $m$  associations are considered. In order to find the relation between  $x_k^{\text{opt}}$  and  $z_m$ , we note that in the class  $x_k$ , there are  $C_n^k$  different ‘configurations’ (vocabularies), each of them has the fraction  $x_k^{\text{opt}}/C_n^k$ . If  $m > k$ , there are  $C_{n-m}^{k-m}$  distinct vocabularies that contain the  $m$  given associations. Summing over all  $k$ , we obtain

$$z_m = \frac{(n-m)!}{n!} \sum_{k=m}^n x_k^{\text{opt}} k(k-1) \cdots (k-m+1). \tag{23}$$

Note that  $z_0 = 1$  and  $z_1 = q^n = \langle m \rangle/n$ . In order to derive equations for  $z_m$ , we use equation (11). Let us set the left-hand side to zero, multiply these equations by  $\frac{m!}{(m-k)!} \frac{(n-k)!}{n!}$  and perform a summation over all  $m$ . Rearranging the order of summation, we obtain

$$\left( \frac{(n-k)!}{n!} \sum_{j=0}^n \frac{j}{n} x_j^{\text{opt}} \left[ \sum_{m=0}^j q^m (1-q)^{j-m} C_j^m \frac{m!}{(m-k)!} \right] - z_1 z_k \right) n z_1 = 0. \tag{24}$$

The expression in square brackets is equal to  $q^k j!/(j-k)!$ , and we can evaluate the summation in  $j$ :

$$(q^k z_{k+1}(n-k) + kq^k z_k - n z_1 z_k) n z_1 = 0, \quad 1 \leq k \leq n-1. \tag{25}$$

Using the expression for  $z_1$  we obtain the following recursive relationship for  $z_k$ :

$$z_{k+1} = z_k \frac{nq^{n-k} - k}{n-k}. \tag{26}$$

This quantity,  $z_k$ , will help us find explicitly the distribution  $\{x_m^{\text{opt}}\}$  as  $n \rightarrow \infty$ . We define the generating function of the distribution  $\{x_m^{\text{opt}}\}$  by

$$f(t) = \sum_{m=0}^n x_m^{\text{opt}} t^m \tag{27}$$

and note that  $z_k = \frac{d^k f}{dt^k} \frac{(n-k)!}{n!}$ . Then we can expand the function  $f$  in the Taylor series around  $t = 1$  which gives

$$f(t) = \sum_{k=0}^n z_k \frac{n!}{(n-k)!k!} (t-1)^k. \tag{28}$$

The first term in this series is 1. Differentiating both sides with respect to  $t$  we get

$$f'(t) = \sum_{k=1}^m z_k C_n^k k (t-1)^{k-1} = \sum_{k=0}^{n-1} z_{k+1} C_n^{k+1} (k+1) (t-1)^k. \tag{29}$$

Now we can use relation (26) to obtain

$$\sum_{k=0}^{n-1} z_k (nq^{n-k} - k) \frac{n!}{k!(n-k)!} (t-1)^k = nq^n \sum_{k=0}^{n-1} z_k C_n^k \left(\frac{t-1}{q}\right)^k - (t-1) \frac{d}{dt} \sum_{k=0}^{n-1} z_k C_n^k (t-1)^k. \quad (30)$$

Note that the summation index can be taken up to  $n$  (instead of  $n - 1$ ) because the  $n$ th term has a zero contribution. Therefore we have:

$$tf'(t) = nq^n f(t + (1 - q)(t - 1)/q). \quad (31)$$

This equation describes the generating function of the distribution  $x_0^{\text{opt}}, x_1^{\text{opt}}, \dots, x_n^{\text{opt}}$ . Note that the fidelity  $q$  varies in the range  $1 - 1/n \leq q \leq 1$ , so that the argument of the function  $f$  in the right-hand side is not too different from  $t$  if  $n \gg 1$ . Let us set  $(1 - q)n \equiv \alpha$ . In the zeroth approximation, we have  $tf'(t) = nq^n f(t)$  with  $f(1) = 1$ , which gives  $f(t) = t^{nq^n}$ . In order to obtain the distribution  $\{x_m^{\text{opt}}\}$  we replace the argument of  $f(t)$  by  $e^{ip}$ , i.e.,  $f(p) = e^{ipnq^n}$ . The reverse Fourier transform of the function  $f(p)$  is  $x_m^{\text{opt}}$ :

$$x_m^{\text{opt}} = \delta(m - nq^n). \quad (32)$$

Indeed, in the limit  $n \rightarrow \infty$  the distribution function becomes sharper and sharper, tending to the delta-function centered at  $m = \langle m \rangle = nq^n$ . In the next approximation we have

$$tf'(t) = nq^n (f(t) + f'(t)(1 - q)(t - 1)/q). \quad (33)$$

The solution of this equation is

$$f(t) = [t(1 - \alpha q^{n-1}) + \alpha q^{n-1}]^{\frac{nq^n}{1 - \alpha q^{n-1}}}. \quad (34)$$

Again, we can replace  $t \rightarrow e^{ip}$  and perform the back Fourier transformation. Note that since  $n$  is large, the quantity  $N \equiv \frac{nq^n}{1 - \alpha q^{n-1}}$  can be taken to be an integer number. Then we can expand the power using binomial coefficients. The result is

$$x_m^{\text{opt}} = C_N^m (1 - \alpha q^{n-1})^m (\alpha q^{n-1})^{N-m}. \quad (35)$$

This is a binomial distribution centered at  $m = nq^n$ . Its standard deviation is given by formula (22). In Fig. 2, both the exact solution (the diamonds) and the distribution (35) (solid line) are shown.

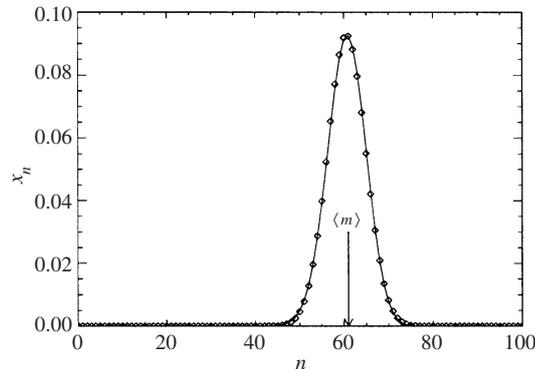


Figure 2. The components of the optimal solution for  $n = 100$ ,  $q = 1 - 1/(2n)$ . The diamonds denote  $x_0^{\text{opt}}, x_1^{\text{opt}}, \dots, x_{100}^{\text{opt}}$  [solution (15), (16)]. The continuous line is the binomial distribution found in formula (35).

**3.3. Stability analysis.** The stability of solution (15)–(17) can be easily examined in the framework of system (11), by standard methods of a linear analysis. Namely, we perturb all the components of the optimal solution by adding  $\tilde{y}_j$  to each  $x_j^{\text{opt}}$  and assuming that the norm of the vector  $\tilde{\mathbf{y}}$  is small. Then system (11) is linearized with respect to components of  $\tilde{\mathbf{y}}$ , and an exponential time behaviour is assumed, i.e.,  $\tilde{y}_j = y_j e^{\Gamma t}$ . The resulting system is a homogeneous set of linear algebraic equations, which only has non-trivial solutions if the determinant of the corresponding matrix is equal to zero. This gives an equation for the growth rate,  $\Gamma$ , with  $n$  solutions:

$$\Gamma_j = q^{n+j} n(j/n - q^{n-j}), \quad 0 \leq j \leq n - 1. \quad (36)$$

In order to guarantee the stability of the optimal solution, we need to require that all the values of  $\Gamma$  are non-positive. If inequality (18) holds, then all  $\Gamma_j \leq 0$ , which implies that the optimal solution is stable with respect to perturbations allowed by system (11). However, such analysis only deals with a restricted class of perturbations. Namely, it only includes functions that satisfy condition (7). It turns out that general perturbations require a more strict stability condition. This can be explained in the following intuitive way.

Let us consider the situation when nobody in the whole population knows the  $n$ th association. This means that the last entry in all the language sequences will remain zero for all generations. In order to find solutions of the corresponding system, we can simply forget about the last association and treat the system as an  $n - 1$  system. The optimal solution can be found again and it exists for  $q \geq 1 - 1/(n - 1)$ . Its average fitness can be found from formula (19) with  $n$  replaced by  $n - 1$ , and it is strictly smaller than the fitness of solution (15)–(17) as long as  $q^2 > 1 - 1/n$ . Now if we go back and try to characterize the solution with one association missing in terms of the original variables of the  $n$ -system, we will find that this is impossible

because condition (7) is violated. It turns out that it is towards such solutions (the loss of one association) that the original solution (15)–(17) loses stability.

In order to find the stability criterion, we will introduce new variables. Let us separate all languages into those which have the  $n$ th association (the corresponding fraction of people is denoted by  $v_k$ ) and those which do not (the corresponding fraction of people is denoted by  $u_k$ ). The index  $k$  is the number of associations that exist in the language *besides* the  $n$ th association. The last entry is zero in all  $u$ -languages and 1 in all  $v$ -languages, so that language  $u_k$  has rank  $k$  and language  $v_k$  has rank  $k + 1$ . We will assume that

$$u_k^{(i)} = u_k^{(j)}, \quad v_k^{(i)} = v_k^{(j)}, \tag{37}$$

which means that languages  $u$  and  $v$  are uniformly distributed within their classes. Note that condition (7) implies (37) but not vice versa. It is easy to check that solution (15)–(17) can be rewritten in the new variables as

$$v_j = x_{j+1}^{\text{opt}}(j + 1)/n, \quad u_j = x_j^{\text{opt}}(n - j)/n. \tag{38}$$

We will now use system (3) together with assumption (37) to find the stability of the fixed point (38). The fitness and transition matrices have to be refined to accommodate the new variables. We have

$$Q_{mj}^{vv} = Q_{mj}q, \quad Q_{mj}^{uu} = Q_{mj}, \quad Q_{mj}^{vu} = Q_{mj}(1 - q), \quad Q_{mj}^{uv} = 0, \tag{39}$$

$$F_{mj}^{vv} = mj/(n - 1) + 1, \quad F_{mj}^{uu} = F_{mj}^{uv} = F_{mj}^{vu} = mj/(n - 1), \tag{40}$$

where the superscript  $vv(uu)$  indicates that we make transitions between two  $u$ -languages ( $v$ -languages), and the superscripts  $uv$  and  $vu$  imply a transition between groups  $u$  and  $v$ . Let us define  $\mathcal{A} = \sum_{l=1}^{n-1} l(u_l + v_l)$ ,  $\mathcal{B} = \sum_{l=0}^{n-1} v_l$ . Then the system that the new variables satisfy is given by

$$\dot{v}_k = q^{k+1} \sum_{j=k}^{n-1} v_j C_j^k (1 - q)^{j-k} \left( \frac{j}{n-1} \mathcal{A} + \mathcal{B} \right) - v_k \left( \frac{\mathcal{A}^2}{n-1} + \mathcal{B}^2 \right), \tag{41}$$

$$\begin{aligned} \dot{u}_m = & q^m \sum_{j=m}^{n-1} u_j C_j^m (1 - q)^{j-m} \frac{j}{n-1} \mathcal{A} \\ & + q^m \sum_{j=m}^{n-1} v_j C_j^m (1 - q)^{j-m+1} \left( \frac{j}{n-1} \mathcal{A} + \mathcal{B} \right) - u_m \left( \frac{\mathcal{A}^2}{n-1} + \mathcal{B}^2 \right) \end{aligned} \tag{42}$$

with  $0 \leq k \leq n - 1$ ,  $1 \leq m \leq n - 1$ . One can check that solution (38), (15), (16) is a fixed point of this system. To study its stability, we perturb the optimal solution:

$$v_j = x_{j+1}^{\text{opt}}(j + 1)/n + V_j e^{\Gamma t}, \quad u_j = x_j^{\text{opt}}(n - j)/n + U_j e^{\Gamma t}. \tag{43}$$

Note that the only perturbations of interest are those with

$$V_{j-1} + U_j = 0, \quad 1 \leq j \leq n-2, \quad \text{and} \quad V_{n-1} = 0. \quad (44)$$

Indeed, if the total perturbation within each rank is non-zero, the corresponding perturbation in system (11) is also non-zero, and the appropriate analysis has already been performed and resulted in instability criterion (18). The only perturbations that the old system ‘overlooked’ are the ones that sum up to zero within the corresponding class. Therefore, we can use equations (44) to reduce the number of variables from  $2n-1$  to  $n-1$ . Another simplification resulting from (44) is that the total perturbation of the average fitness is zero, because

$$\phi = \mathcal{A}^2/(n-1) + \mathcal{B}^2 = nq^{2n}, \quad (45)$$

where  $u_j$  and  $v_j$  are defined by equations (43) and the perturbations satisfy condition (44). We obtain the linear system

$$\Gamma V_m = q^{n+m+1} \sum_{j=m}^{n-2} V_j C_j^m (1-q)^{j-m} (j+1) + W_m \sum_{j=0}^{n-2} V_j - nq^{2n} V_m, \\ 0 \leq m \leq n-2, \quad (46)$$

where  $W_m \equiv q^{m+1} \sum_{l=m}^{n-1} x_{l+1}^{\text{opt}} \frac{(l+1)(n-1-l)}{n(n-1)} C_l^m (1-q)^{l-m}$  and  $x_{l+1}^{\text{opt}}$  comes from the equilibrium solution (15), (16). The condition that the determinant of the corresponding  $(n-1) \times (n-1)$  matrix is equal to zero leads to  $(n-1)$  expressions for the growth rate,  $\Gamma$ , as functions of  $q$ . It turns out that all of the growth rates are negative if  $q$  is sufficiently big. In other words, there exists a value  $q = q_c$  such that for all  $q \geq q_c$ , all the growth rates  $\Gamma$  are non-positive. The value  $q_c$  is the error threshold, i.e., the minimum fidelity required for the population to maintain all the  $n$  associations in the vocabulary. The threshold fidelity values can be found for all  $n$ , the result is presented in Fig. 3 by the crosses. Note that the threshold value is always bigger than  $1 - 1/n$  (the solid line in Fig. 3), i.e., there is a region in  $q$  where the optimal solution exists but is unstable. The existence condition is given by formula (18) and the stability criterion is specified by the  $q_c$  values calculated from system (41), (42).

**3.4. Lexical capacity and error threshold.** We can find the limiting behavior of the stability criterion in the limit of large  $n$ . We know that the most unstable perturbation,  $\delta \mathbf{x}$ , does not change the value of the average fitness, see equation (45). Therefore, we can write:

$$\phi\{\mathbf{x}^{\text{opt}}(n) + \delta \mathbf{x}\} = \phi\{\mathbf{x}^{\text{opt}}(n)\}. \quad (47)$$

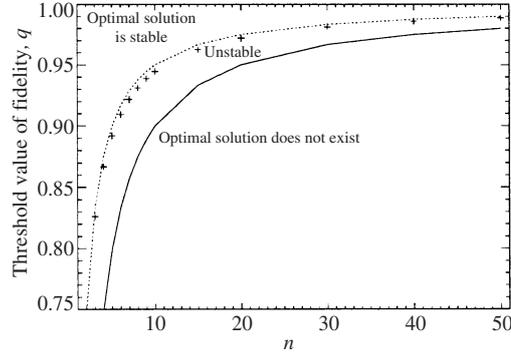


Figure 3. The stability diagram for the optimal solution for different values of  $n$ . Crosses are the values of  $q_c$  calculated from system (46). The solid line is  $q = 1 - 1/n$ , the existence threshold, and the dotted line is  $q = 1 - 1/(2n)$ .

In addition, we know that the optimal solution becomes unstable towards losing an association. This means that in the space of vectors  $\mathbf{x}$ , the corresponding perturbation points in the direction of  $\mathbf{x}^{\text{opt}}(n - 1)$ , the solution where one association is missing from the vocabulary. Therefore, we have

$$\delta \mathbf{x} = C(\mathbf{x}^{\text{opt}}(n - 1) - \mathbf{x}^{\text{opt}}(n)), \tag{48}$$

where  $C$  is some constant. For large values of  $n$  we can rewrite this as  $\delta \mathbf{x} = \partial \mathbf{x}^{\text{opt}}(n) / \partial n \, dn$ . If we plug this into equation (47) and use the Taylor expansion we obtain:

$$\frac{\partial \phi}{\partial \mathbf{x}^{\text{opt}}(n)} \frac{\partial \mathbf{x}^{\text{opt}}(n)}{\partial n} = \frac{\partial \phi}{\partial n} = 0. \tag{49}$$

Let us introduce the term *lexical capacity* for the maximum number of associations that can be maintained in the population. Formula (49) shows that for a given learning accuracy, the capacity of the system is equal to the number  $n$  which maximizes the average fitness. This is a well-known principle (Fisher, 1930) which, however, does not always hold for frequency dependent payoff. Using  $\phi = nq^{2n}$ , we obtain for capacity,  $n_{\text{max}}(q) = (-2 \log q)^{-1}$ . For  $q$  close to 1, the capacity is simply given by

$$n_{\text{max}}(q) = 1/[2(1 - q)]. \tag{50}$$

We can also use expression (49) to find the error threshold compatible with the population maintaining a given number,  $n$ , of associations. We have  $q_c = e^{-1/2n}$ , or, for large values of  $n$ , the stability threshold is given by

$$q_c = 1 - \frac{1}{2n}. \tag{51}$$

The dotted line in Fig. 3 represents the function  $1 - 1/(2n)$ . Note that it is always just inside the stability region. This means that at the transition point, the solution

with  $n$  associations loses stability to the solution with  $n - 1$  associations with a slight *increase* in the average fitness.

The full transition diagram is shown in Fig. 4. As  $q$  gets larger, more and more associations can be maintained in the population. For simplicity let us use the large  $n$  estimate for  $q_c$  given by equation (51). Then for a given  $q$ , at most  $n = n_{\max}(q)$  associations can be maintained in the population, where the capacity  $n_{\max}(q)$  is the integer between  $1/(2(1 - q)) - 1$  and  $1/(2(1 - q))$ . The interval  $0 \leq q \leq 1$  consists of an infinite number of sub-intervals with increasing lexical capacity (eight of them are shown in Fig. 4). In each of the sub-intervals,  $n_{\max}(q)$  stable optimal solutions with  $n = 1, n = 2, \dots, n = n_{\max}$  coexist, and the solution with  $n = n_{\max}$  corresponds to the maximum possible average fitness. Note that the optimal solution corresponding to  $n_{\max}(q)$  also has the largest basin of attraction. The basin of attraction of the other optimal solutions with  $n < n_{\max}$  is of the order of  $n(1 - q)$ , i.e., it shrinks to zero for smaller  $n$  and larger  $q$ , see Appendix B. This means that starting from a vast majority of initial conditions, the system will relax to the stable optimal solution corresponding to  $n_{\max}(q)$ , i.e., reach its full capacity. The average number of associations known by people is then given by  $n_{\max}q^{n_{\max}}$ . As  $q$  approaches 1 (and  $n \rightarrow \infty$ ), this can be simplified to

$$\langle m \rangle = \frac{1}{2\sqrt{e}(1 - q)}. \quad (52)$$

**3.5. Words are arbitrary signs.** To conclude this section we will note that the total number of stable fixed points of system (3) that we have found is very large. Each optimal solution of size  $n$  represents a *family* of similar solutions of size  $n$ , each of them can be obtained from the diagonal one by interchanging rows and columns. The number of solutions of size  $n$  is given by  $n!$ . For each given value of  $q$  the solutions which have maximum fitness contain  $n_{\max}(q)$  associations, where  $n_{\max}$  is the integer between  $1/(2(1 - q)) - 1$  and  $1/(2(1 - q))$ . Thus there are  $[n_{\max}(q)]!$  coexisting stable solutions maximizing the fitness. This is a direct consequence of the fact that the associations between signals and referents are arbitrary. For each non-ambiguous system of associations, the corresponding stable solution can be found. Exactly which set of associations will become adopted in the language depends entirely on the initial conditions of the system.

#### 4. THE FIXED POINTS ARE STABLE AGAINST PERTURBATIONS BY GENERAL LEXICAL MATRICES

Here we will prove that the stable fixed points found in Section 3 remain stable attractors in the full system. The method we are going to use is as follows. We will take one of the  $n!$  stable fixed points found in the previous section and write it down in terms of general variables of the full system (3). This means that only the

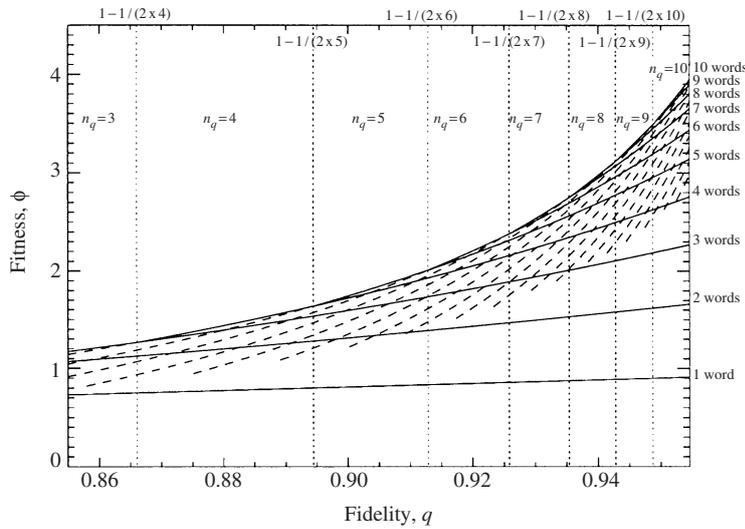


Figure 4. The full transition diagram. Solid lines denote stable solutions (the maximum number of associations that can be contained in the corresponding population’s vocabulary is marked on the right). The dashed lines represent unstable solutions. The dotted vertical lines are  $q = 1 - 1/(2n)$ , they separate different regions of the diagram. In each of the regions,  $n_q$  denotes the maximum number of associations,  $n_{\max}(q)$ , that can be stably maintained in the population.

non-ambiguous matrices will have a non-zero share in the population, and the share of the others will be set to zero. Next, we will carry out a linear stability analysis of such a solution in the general system and prove that it is stable with respect to all possible perturbations. Note that it is enough to only consider one of the family of optimal solutions because they all are equivalent up to permutations of rows and columns. Our analysis will show that the optimal solutions are stable in the system where all binary matrices are allowed, including those with more than one positive entry in the same row/column. In other words, the optimal solutions will turn out to be stable with respect to the invasion of synonyms and homonyms.

There are  $N = 2^{n^2}$  different  $A$  matrices. The dynamics of language acquisition is described by the general system (3). Let us consider the subset  $\mathcal{L}$  of all lexical matrices defined by the following rule: take any permutation matrix and form all matrices that can be obtained from the permutation matrix by removing positive entries. By the appropriate renumbering of referents and signals, this set can be made identical to the set considered in the previous section (i.e., the matrices whose positive entries are situated on the main diagonal). The set  $\mathcal{L}$  only includes non-ambiguous languages. We will call all languages that do not belong to  $\mathcal{L}$  *competing* languages. This is because every such language will contain at least one entry which will compete with the entries of the unambiguous  $\mathcal{L}$  languages (it could be a signal that shares its meaning with another signal from an  $\mathcal{L}$  language, or/and a second referent assigned to the existing signal).

We will use lower case subscripts to denote the rank of  $\mathcal{L}$  languages and capital subscripts for matrices/languages. Let us write down the optimal solution found in the previous section in terms of the general variables. It corresponds to the vector  $(x_1, \dots, x_N)$ , such that

$$x_J = \begin{cases} 0, & A_J \notin \mathcal{L}, \\ x_k^{\text{opt}}/C_n^k, & A_J \in \mathcal{L}, \quad R(A_J) = k, \end{cases} \quad (53)$$

where  $x_k^{\text{opt}}$  is given in (15)–(17). For this solution, in the whole population only one signal can be used for each referent, only one referent corresponds to each signal, and the distribution is the one found before. A straightforward substitution suggests that this is a fixed point of system (3). Note that there are  $n!$  of such fixed points (as many as there are permutation matrices, see Section 3.4), all of them can be reduced to solution (53) by renumbering referents and signals.

Let us perform a stability analysis of solution (53). As usual, we will perturb each component of  $x_L$  with  $y_L e^{\Gamma t}$  and linearize around the fixed point. First of all we note that the equations for  $\dot{x}_J$  where  $A_J \notin \mathcal{L}$  decouple from the equations for the  $\mathcal{L}$  languages. This is a consequence of the fact that the (unperturbed) share of the competing languages is zero. Therefore, equations for competing languages will not contain perturbations of the  $\mathcal{L}$  languages. This simplifies the analysis because it is sufficient to consider only the equations for competing languages. The analysis for the  $\mathcal{L}$  languages has already been performed. Thus, the system of linear equations is

$$\Gamma y_I = \sum_{A_J \notin \mathcal{L}} f_J y_J Q_{JI} - y_I \phi, \quad A_I \notin \mathcal{L} \quad (54)$$

(the shorthand subscript for the sum means that the summation is performed over all matrices  $J$  such that  $A_J \notin \mathcal{L}$ ). In (54), the unperturbed fitness of the language  $A_J$  is  $f_J = \sum_{A_K \in \mathcal{L}} F(A_J, A_K) x_K$ , and the average fitness is given by equation (19). Since we do not allow for errors which lead to forming new entries in the  $A$  matrix (only to losing old ones),  $Q_{JI} = 0$  whenever the language  $A_I$  has positive entries which are not present in the language  $A_J$ . Therefore, we have  $Q_{JI} = 0$  whenever  $I > J$  for all competing languages; this is the consequence of our numbering procedure. Thus, the matrix of linear system (54) is triangular, with all the entries above the main diagonal being identically zero. To ensure the existence of non-trivial solutions, the determinant of this matrix has to be zero, which gives

$$\prod_{A_J \notin \mathcal{L}} (-\Gamma - \phi - f_J Q_{JJ}) = 0. \quad (55)$$

This equation has to be solved for  $\Gamma$ . The condition which guarantees stability of solution (53) is that all the values of the growth rate,  $\Gamma$ , are non-positive. This leads to the following inequality:

$$f_J Q_{JJ} \leq \phi, \quad \forall J \text{ such that } A_J \notin \mathcal{L}. \quad (56)$$

Our task is to show that the fitness of each competing language cannot exceed the average fitness of the population with the optimal distribution of languages. To show that this is indeed the case, we will need to use the fact that the fitness of each of the  $\mathcal{L}$  languages is defined by

$$f_K = kq^n, \quad A_K \in \mathcal{L}, \quad R(A_K) = k. \quad (57)$$

To prove the above equality, let us consider the language  $A_K \in \mathcal{L}$  which has rank  $k$  (i.e., it has exactly  $k$  positive elements). Its fitness is given by

$$f_K = \sum_{A_M \in \mathcal{L}} F(A_K, A_M)x_M. \quad (58)$$

It is convenient to rewrite this as

$$f_K = \sum_{m=0}^n \sum_{\alpha_m=1}^{C_n^m} F(A_K, A_m^{\alpha_m})x_m^{(\alpha_m)}, \quad (59)$$

where the first summation is performed over all ranks, and the second summation goes through all the  $\mathcal{L}$  languages of the current rank  $m$ . Using  $x_m^{(\alpha_m)} = x_m^{\text{opt}}/C_n^m$  for all  $\alpha_m$ , we can perform the inner summation:

$$\sum_{\alpha_m=1}^{C_n^m} F(A_K, A_m^{\alpha_m}) = \sum_{l=0}^{\min(m,k)} l C_k^l C_{n-k}^{m-l} = k C_{n-1}^{m-1}, \quad (60)$$

Then we have

$$f_K = \sum_{m=0}^n x_m^{\text{opt}} k \frac{C_{n-1}^{m-1}}{C_n^m} = \frac{k}{n} \sum_{m=0}^n m x_m^{\text{opt}} = kq^n, \quad (61)$$

which completes the proof of statement (57).

Let us consider the matrices of competing languages and count the number of their positive diagonal elements. If language  $A_J$  has  $j$  diagonal elements, we will say that it has rank  $j$ , i.e.,  $R(A_J) = j$ . We will now show that if  $A_J \notin \mathcal{L}$ ,  $A_K \in \mathcal{L}$  and both languages have rank  $j$  then

$$f_J \leq f_K. \quad (62)$$

From equation (58) it follows that it is sufficient to show that  $F(A_J, A_M) \leq F(A_K, A_M)$  for all  $A_M \in \mathcal{L}$ . In order to obtain  $F(A_J, A_M)$  with  $A_M \in \mathcal{L}$ , we need to form string vectors out of diagonal entries of matrices  $\tilde{p}^{(J)}$  and  $\tilde{q}^{(J)}$  and then take a conventional inner product of these vectors with the diagonal of the  $\mathcal{L}$  language  $M$ . Then  $F(A_J, A_M)$  is the average of these two inner products. Now let us compare  $F(A_J, A_M)$  with  $F(A_K, A_M)$ , where the matrix of the language  $A_K \in \mathcal{L}$

consists of all the diagonal entries of the language  $A_J$ , but has no off-diagonal entries. Obviously,  $F(A_J, A_M) \leq F(A_K, A_M)$ , because the presence of off-diagonal elements can only make the positive diagonal elements of the corresponding  $\tilde{p}$  and  $\tilde{q}$  matrices smaller than 1 (and thus reduce the fitness). Inequality (62) follows immediately.

Next, we note that  $Q_{JJ}$  for  $A_J \notin \mathcal{L}$  satisfies  $Q_{JJ} \leq q^j$  where  $R(A_J) = j$ . Indeed, language  $A_J$  has at least  $j$  elements which all have to be memorized with probability  $q$ . Now it is clear that  $f_J Q_{JJ} \leq f_K q^k$ , and if we show that

$$kq^{n+k} \leq nq^{2n}, \quad (63)$$

then inequality (56) holds automatically [in expression (63) we used formula (57)]. Inequality (63) holds as long as  $q \geq \bar{q}$ , where

$$\bar{q} = \left(\frac{k}{n}\right)^{\frac{1}{n-k}}. \quad (64)$$

The function  $\bar{q}(k)$  increases with  $k$ , and  $\lim_{k \rightarrow n} \bar{q} = e^{-1/n}$ , i.e.,  $\bar{q} \leq e^{-1/n}$ . For large  $n$  we have  $\bar{q} \leq 1 - 1/n$ . Therefore, inequality (56) holds as long as (18) is satisfied. This means that solution (53) is stable with respect to competing languages in the whole domain of its existence.

We can conclude that when all languages are allowed in the system, those which are described by any permutation matrix  $A$ , and all the matrices obtained from  $A$  by removing positive entries, are stable. This can be viewed as an extension of the result of Trapa and Nowak (2000), who have shown that permutation matrices are the only strict Nash equilibria of a language system with the fitness defined by (2). We have made a connection with an evolutionary dynamics and proved that these languages form evolutionary stable states even if the learning ability is not perfect ( $q < 1$ ).

**Remark.** The above analysis proves the stability of optimal solutions which contain no homonyms or synonyms. It does not show that those are the *only* stable fixed points of the full system. However, we believe that the latter statement is nevertheless true because no numerical simulations (see the next section) revealed any other stable solutions of the system.

## 5. STOCHASTIC SIMULATIONS OF EXTENDED MODELS

In this section we will demonstrate how our analytical results can contribute to the understanding of more complicated and more realistic models. In many cases we do not have exact analytical tools to obtain solutions of such models. Therefore, we need to use computer simulations. Below we will describe some model

extensions and the simulation results, but first we would like to emphasize the importance of stochastic modeling when studying language evolution.

There are several reasons why stochastic, rather than deterministic, models should be considered. First, the number of deterministic ODEs [system (3)] grows like  $2^{n^2}$  where  $n$  is the matrix size. Therefore, it is not conceivable to simulate deterministic equations for  $n$  larger than, say, 3. Besides this mundane reason, we note that deterministic equations of the type (3) cannot be written down in the case of positive real-valued  $A$  matrices. On the other hand, stochastic modeling can handle this important case. Finally, stochastic language learning is what in fact happens in reality, and the deterministic equations only approximate this process in the limit of a very large population size. Finite population size effects can be significant and lead to some interesting features which are suppressed in the deterministic model, such as spontaneous language changes.

In this section we will present a brief report of preliminary numerical results that we have obtained and show that the analysis given above is an important first step towards our understanding of the full model.

**5.1. Incomplete learning.** In this subsection we set up a stochastic model and test it in the case of incomplete learning, where its behavior can be directly compared with our analytical results. Let us consider a population of size  $N$ . Each individual is characterized by an  $A$  matrix. The fitness is evaluated according to equation (2). Every individual talks with equal probability to every other individual and their fitness is evaluated. For the next generation, individuals produce children proportional to their payoff, i.e., successful communication is rewarded. Children learn the language of their parents. The average payoff of the population is given by

$$\phi = \frac{1}{N(N-1)} \sum_{I,J} F(A_I, A_J). \quad (65)$$

The summation is over all individuals so that  $I = 1, \dots, N$  and  $J = 1, \dots, N$  as long as  $I \neq J$ . The quantity  $\phi$  describes the expected number of referents that a random individual can successfully communicate to another random individual. Therefore,  $\phi$  can be interpreted as the *effective lexicon size of the population*. In the limit of  $N \rightarrow \infty$  this quantity coincides with the average fitness defined in equation (5).

We describe lexicon learning as a probabilistic process. The learner starts off with an  $A$  matrix that has all entries set to zero. Then the teacher's  $A$  matrix is copied into the learner's matrix in such a way that all the positive entries have the probability  $q$  to be copied as 'ones', and the probability  $1 - q$  to be copied as 'zeros'.

This stochastic model corresponds exactly to the previous model in the limit  $N \rightarrow \infty$ . Finite population effects play a role, but the analytical results obtained in the previous sections can still be used to describe the behavior of the system.

The long-time dynamics of the stochastic process can be viewed as a series of transitions between stable fixed points of the deterministic system. For instance, let us suppose that the system starts off with  $n_{\max}(q)$  signals. For a number of generations the population will remain in a vicinity of the optimal solution corresponding to  $n_{\max}(q)$ , and then the system will jump into a state which corresponds to the optimal solution with  $n = n_{\max}(q) - 1$ . The (time-averaged) fitness will lower from  $n_{\max}q^{2n_{\max}}$  to  $(n_{\max} - 1)q^{2(n_{\max}-1)}$ , and one association will be lost entirely from the population. The system will spend some time in this state until another jump will happen with a loss of another signal. Between transitions the system exhibits a quasi-stationary behavior. On average, the time spent at each of the fixed points gets longer and longer as  $n$  decreases. Eventually (after  $n_{\max}$  jumps), all lexical entries will be lost. An all-zero  $A$  matrix is the only absorbing state of this system, because there is no chance to create new lexical entries. The time it takes to reach this absorbing state can of course be extremely long.

Figure 5 presents an example of an evolutionary process for a population of  $N = 40$  individuals with a learning accuracy of  $q = 0.99$ . As the initial condition we chose a state close to the  $n = 6$  optimal solution. We can see that as time goes by, signals are lost from the vocabulary of the population. Between the transitions, the system oscillates near the optimal solution of the deterministic system and has fitness near  $\phi = nq^{2n}$  (the dotted horizontal lines). The time-evolution of the lexical matrix depends crucially on the number of people in the population. The larger the population size, the longer is the average time spent at each of the quasi-stationary states.

We have carried out a number of runs with random initial conditions, so that in the beginning, the individuals had different randomly chosen binary matrices. The system always converged to a non-ambiguous quasi-stationary state. Note that no quasi-stationary solutions with synonyms or homonyms have been observed in the stochastic system. This confirms our hypothesis that if no associations can be created by mistake, then the optimal non-ambiguous solutions are the only stable equilibria of the system.

**5.2. Incorrect learning.** Now let us extend the model of the previous section to incorporate a possibility of a different kind of mistake. Before, when learning, the children could lose some of the associations of their parents' matrix. No new entries were allowed to be made by mistake. Now, let us suppose that erroneous associations can be memorized by children during the learning stage. If the teacher has  $a_{ij} = 0$  then the learner will have  $a_{ij} = 0$  with probability  $q_0$ , and  $a_{ij} = 1$  with probability  $p_0 \equiv 1 - q_0$ . Thus, the probability to create an association which was not there in the teacher's matrix, or the *error rate* of incorrect learning, is  $p_0$ . As before, the probability to keep a positive entry is  $q$ . We will call the probability to lose a positive entry the error rate of incomplete learning,  $p \equiv 1 - q$ . With  $p_0 = 0$  this new model is reduced to the incomplete learning model.

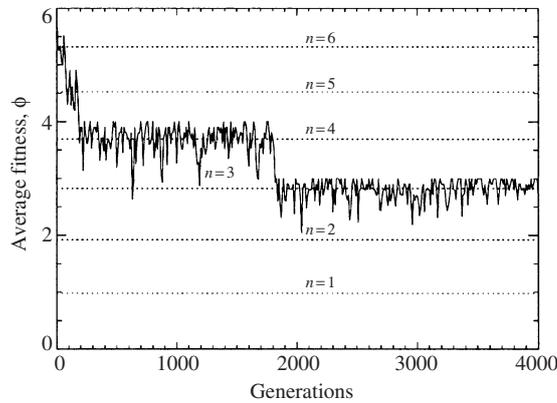


Figure 5. The average fitness of the population for a stochastic model as a function of time. The population size is  $N = 40$ , and  $q = 0.99$ . Quasi-stationary states are observed which correspond to optimal solutions of the deterministic model with  $n = 6, \dots, n = 3$ . At each of these states, the average fitness oscillates around the theoretically calculated value,  $\phi = nq^{2n}$ , plotted with dotted lines. Eventually, all signals will be lost, but this can take a very long time. For practical purposes, the population will maintain a certain set of lexical items.

If the error rate of incorrect learning is much smaller than the error rate of incomplete learning,

$$p_0 \ll p, \tag{66}$$

then the dynamics can still be described by using our results for the  $p_0 = 0$  case. As in the case of the incomplete learning stochastic process, the system performs transitions between stable fixed points of the corresponding deterministic system. Since the error rate  $p_0$  is very small, the quasi-stationary states can be well approximated by stable fixed points of the deterministic system with  $p_0 = 0$ . However, the fact that  $p_0$  is bigger than zero plays an important role in the long-term dynamics of the system. Recall that before, the transitions were always made in one direction, namely, towards a loss of an association. In the case of  $p_0 = 0$ , once an association is lost from the population, it cannot be gained back. If  $p_0 > 0$  (incorrect learning), the situation changes and with a finite probability, a state with  $n$  associations can give way to a state with  $n + 1$  associations. This is illustrated in Fig. 6, which shows three examples of evolutionary runs for a system with  $p = 10^{-2}$  and  $p_0 = 10^{-4}$  for three different population sizes:  $N = 10$ ,  $N = 40$  and  $N = 70$ . For each of the runs, we can see that as generations go by, the system attends the states with  $n = 6, \dots, n = 1$ , gaining or losing associations. At each of the states, the average fitness is well approximated by  $\phi = nq^{2n}$ , the formula derived under the assumption that  $p_0 = 0$ .

Let us now take a closer look at the behavior of the stochastic system once it is in a quasi-stationary state. For comparison with the deterministic system, it is convenient to introduce the average lexical matrix of the population,

$$\bar{A} = \sum_I x_I A_I. \tag{67}$$

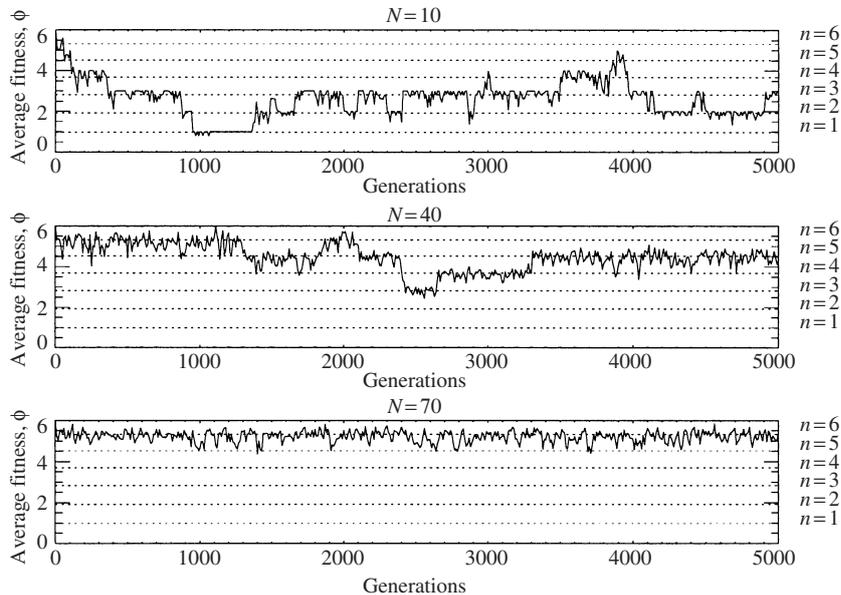


Figure 6. The average fitness of the population for a stochastic model as a function of time, for three different population sizes;  $p = 10^{-2}$ ,  $p_0 = 10^{-4}$ . The dotted lines are the levels  $nq^{2n}$ .

If the system is at an optimal solution, we have

$$\bar{A}^{\text{opt}} = q^n \mathbf{I}_n, \tag{68}$$

where  $q$  is the accuracy of learning,  $n$  is the maximum number of associations that people know; this quantity is defined by the fixed point the system is at, and  $\mathbf{I}_n$  is a permutation matrix with  $n$  positive entries. Not surprisingly, if condition (66) is satisfied, then the average lexical matrix of the stochastic process in a quasi-stationary state is very close to that of the deterministic  $p_0 = 0$  process. As an illustration, let us consider the average lexical matrix for the run in Fig. 6 with  $N = 40$ , taken for the quasi-stationary state between generations 2700 and 3300. The time-averaged  $\bar{A}$  and the corresponding  $\bar{A}^{\text{opt}}$  are, respectively,

$$\bar{A} = \begin{pmatrix} \cdot & \cdot & \cdot & \circ & \cdot & \cdot \\ \cdot & \cdot & \cdot & \bigcirc & \cdot & \cdot \\ \circ & \bigcirc & \cdot & \circ & \cdot & \cdot \\ \cdot & \circ & \circ & \cdot & \cdot & \bigcirc \\ \bigcirc & \cdot & \cdot & \circ & \cdot & \circ \\ \cdot & \cdot & \circ & \cdot & \cdot & \cdot \end{pmatrix}; \quad \bar{A}^{\text{opt}} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \bigcirc & \cdot & \cdot \\ \cdot & \bigcirc & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \bigcirc \\ \bigcirc & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \tag{69}$$

here the radius of each circle is proportional to the strength of the association. We can see that the average lexical matrix,  $\bar{A}$ , is largely dominated by a permutation

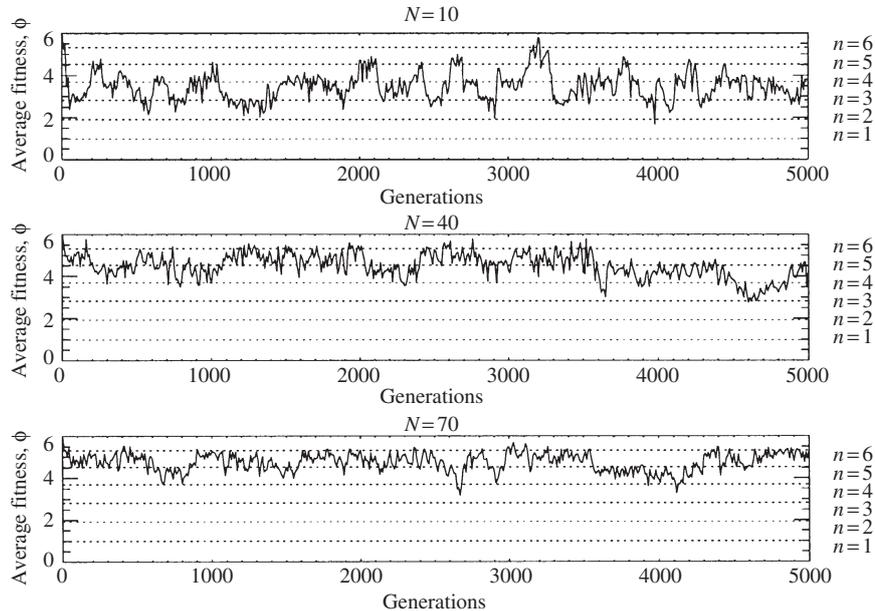


Figure 7. Same as Fig. 6, except  $p_0 = 10^{-3}$ .

matrix. In contrast with the  $p_0 = 0$  case, all matrices may have a non-zero share, thus  $\bar{A}$  has extra non-zero elements besides the dominating permutation ‘skeleton’. However, the fraction of matrices with ‘extra’ elements (what we called competing matrices in Section 4) is very small. Therefore, it still makes sense to talk about a lexicon size.

As long as condition (66) is satisfied, the average fitness of quasi-stationary states oscillates around the average fitness of the corresponding optimal solutions with  $p_0 = 0$ , as illustrated by Fig. 6. As  $p_0$  becomes larger, this is no longer the case, see Fig. 7. For the three runs here, the population sizes and the error rate of incomplete learning,  $p$ , were taken to be the same as in the previous figure, but the probability to create an erroneous association was increased, so that  $p_0 = 10^{-3}$ . In this intermediate regime, the average fitness of each of the quasi-stationary states is lower than the expected fitness of the fixed points of the deterministic system with  $p_0 = 0$ . The average lexical matrix is still close to a permutation matrix, but the share of other elements is larger than it was for runs with  $p_0 = 10^{-4} \ll p$ .

As the error rate of incorrect learning becomes significant, it gets harder and harder to determine the lexicon size by looking at the average association matrix. When  $p_0 \sim p$ , the competing matrices have a considerable share, which can in fact be larger than the share of non-ambiguous matrices. In Fig. 8 we performed stochastic runs with  $p_0 = p = 10^{-2}$ . As a result of a large probability to create new entries, the coherence is lowered significantly; too many erroneous entries are accumulated during the learning process. A more detailed description of the regimes of Figs 7 and 8 goes beyond the scope of the present paper and will be addressed elsewhere.

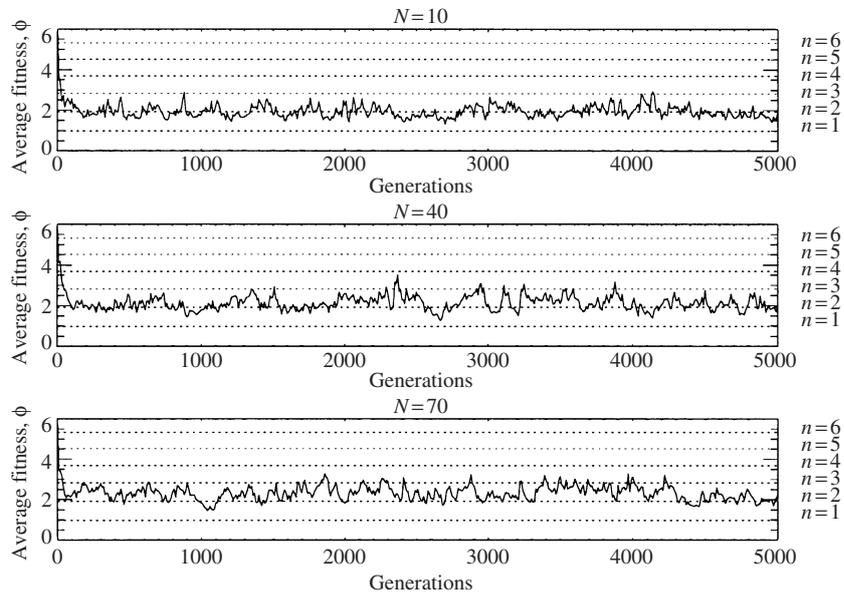


Figure 8. Same as Fig. 6, except  $p_0 = 10^{-2}$ .

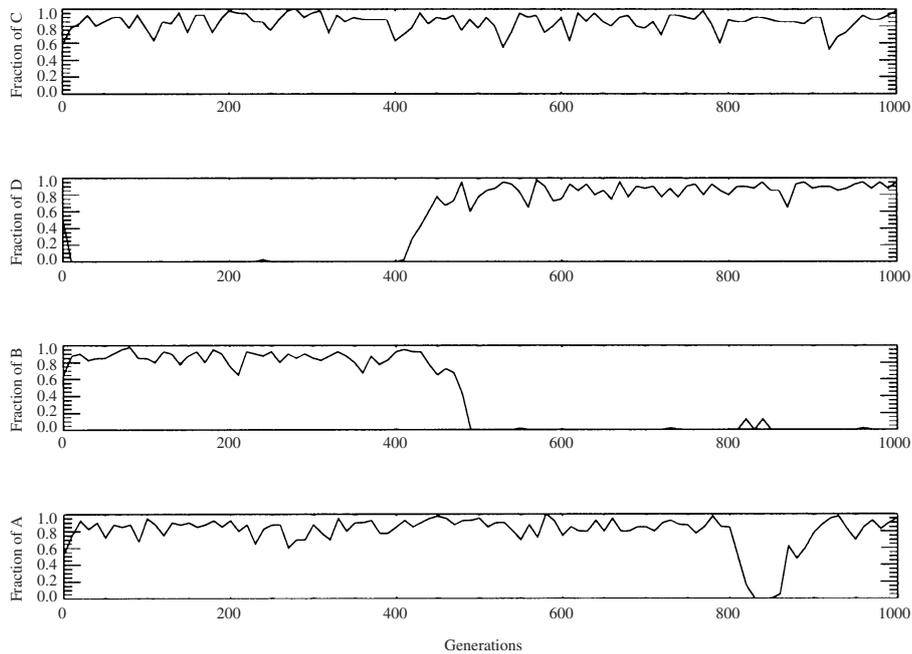


Figure 9. Spontaneous changes in lexicon. For this run,  $N = 40$ ,  $p = 4 \cdot 10^{-2}$  and  $p_0 = 10^{-3}$ . The time evolution of four entries in the average lexical matrix, formula (70), is presented. The rest of the associations remain very weak and are not shown.

**5.3. Spontaneous changes of lexicon.** The simple model described above gives rise to stochastic changes in lexicon over time. This is a phenomenon which is observed in historical linguistics; our model demonstrates that it may be a consequence of finite population size effects.

We know that stable equilibrium solutions of the deterministic model consist of non-ambiguous languages (and may contain a small share of competing languages if  $p_0$  is very small). There are  $n!$  different ways to associate  $n$  signals with  $n$  referents which gives the total number of non-ambiguous solutions. These solutions coexist in the phase space. In equations (3), once the system has reached one of the equilibrium solutions, no further changes are possible. However, a finite size plays a role of a finite perturbation that constantly acts in the system. Once in a while it may become sufficient to kick the system out of a stable equilibrium and bring it over to the attractor of another stable equilibrium. Then, a spontaneous change in lexicon takes place. The smaller the population size, the more likely such transitions are to occur in a given length of time.

It is interesting to follow the time evolution of the average lexical matrix. As an example we consider a system with  $N = 40$  people and  $p = 4 \cdot 10^{-2}$ ,  $p_0 = 10^{-3}$ . The average lexical matrix of this system for a particular run looked like

$$\bar{A} = \begin{pmatrix} \circ & D & \circ & \circ \\ A & \circ & \circ & \circ \\ \circ & B & \circ & \circ \\ \circ & \circ & C & \circ \end{pmatrix} \tag{70}$$

where by small circles we denoted entries whose values were less than 0.1 for generations 1 through 1000. Associations  $A$ ,  $B$ ,  $C$  and  $D$  are those of interest. Their evolution is plotted in Fig. 9. In the beginning, a population adopts a non-ambiguous language which consists of three associations,  $A$ ,  $B$ ,  $C$ . Then, around generation 400, one of the signals disappears from the language and gets replaced by another signal, which stands for the same referent. Thus the total number of associations is still three. Around generation 800, one of the associations is briefly lost from the vocabulary, so that for a while there are only two of them maintained. However, this association is gained back shortly afterwards. We can depict the spontaneous changes in lexicon observed here as

$$\begin{pmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \end{pmatrix} \rightarrow \begin{pmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \end{pmatrix} \rightarrow \begin{pmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \end{pmatrix} \rightarrow \begin{pmatrix} \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ \end{pmatrix}. \tag{71}$$

Note that in this simulation, a small fraction of synonyms and homonyms is always present because of a finite probability of incorrect learning, but we can still observe that the dominant language is non-ambiguous.

## 6. DISCUSSION

We studied the population dynamics of the acquisition and evolution of the lexical matrix. We assumed that information exchange between individuals leads to a payoff which contributes to fitness. Successful individuals are more likely to transfer their lexical matrix to others. We considered two different types of mistakes that children can make during the learning acquisition: (i) incomplete learning, where one or more of the teacher's associations are lost, and (ii) incorrect learning, where children create erroneous entries in their lexical matrices.

In the case of incomplete learning, we analysed how accurate the acquisition of the lexical matrix must be for a population to maintain a coherent (non-ambiguous) lexical matrix of a given size. If  $q$  denotes the probability that a learner acquires a specific entry of the lexical matrix of the teacher, and  $n$  is large, then  $q$  has to exceed the threshold value

$$q_c = 1 - \frac{1}{2n}. \quad (72)$$

For a given  $q$ , the lexical capacity of the system is defined by

$$n_{\max}(q) = [2(1 - q)]^{-1}. \quad (73)$$

This value of  $n$  maximizes the average fitness for a given  $q$ . Not everybody in the population knows all the associations. The average number of associations known by a person is given by  $\langle m \rangle = nq^n$ . For  $n = n_{\max}$ , we obtain

$$\langle m \rangle = n_{\max}/\sqrt{e}. \quad (74)$$

The standard deviation is  $\sigma^2 = n_{\max}^2 q^{2n_{\max}-1} (1 - q)$ . The distribution of the number of associations known to individuals is nearly binomial in the case of large  $n_{\max}$ . The effective lexicon size,  $\phi$ , is defined as the average number of associations that any two individuals have in common. We find that  $\phi = nq^{2n}$ . For  $n = n_{\max}(q)$  we obtain

$$\phi = n_{\max}/e. \quad (75)$$

For given values of  $q$ , equations (74) and (75) can be rewritten as

$$\langle m \rangle = [2\sqrt{e}(1 - q)]^{-1}, \quad \phi = [2e(1 - q)]^{-1}. \quad (76)$$

We can also rewrite the condition for  $q_c$ , formula (72), in terms of a minimum of learning events per learner. Let  $b$  denote the number of learning events that a learner has during its language acquisition period. Hence  $b$  specifies the number of times a correct entry is made in the lexical matrix of the learner. If all referents occur at the same frequency, we have

$$q = 1 - (1 - 1/n)^b. \quad (77)$$

There are  $b$  independent events. In any one event, one of  $n$  referents is chosen at random. Equations (72) and (77) lead to

$$b > b_c = n \ln(2n). \quad (78)$$

A more realistic model includes incorrect learning, where learners can by mistake form associations that are not present in their teacher's matrix. A full analysis of this situation is still missing. In the present paper we have considered the situation where the probability to create new associations is small. Numerical simulations have demonstrated a good quantitative agreement with our analytical predictions. The population tends to converge to a quasi-coherent state, where the averaged lexical matrix of all individuals contains dominant non-ambiguous entries, the share of synonyms and homonyms is rather small and the average fitness is well described by formula (76).

As the population size tends to infinity, the analytical results obtained for the deterministic system hold exactly. However, finite population effects may play a role. For instance, spontaneous changes of language have been observed in stochastic simulations for populations of a finite size. In the course of evolution, new associations have been observed to gradually replace old ones. The dynamics of the stochastic system can be viewed as a series of transitions between stable attractors of the corresponding deterministic system.

Throughout this paper we have assumed that each individual learns its lexical matrix from one teacher. We have also simulated the situation where individuals learn from more than one teacher [see also Nowak *et al.* (1999, 2000) and Nowak (2000)]. The teachers are chosen randomly but proportionally to their payoff in the population. In this case, we obtain similar results provided the lexical matrix has integer-valued entries that are not restricted to 0 and 1. Whenever a learner receives a referent–signal pair, the learner *increases* the corresponding entry in his association matrix by one point. With this extended mechanism, the population can evolve and maintain coherent lexical matrices. If, however, we keep the entries of the lexical matrix restricted to 0 and 1, then a population, where individuals learn from several teachers, does not maintain a coherent lexicon even for small error rates. We note that a population where individuals learn from one or more teachers that are randomly chosen irrespective of their payoff cannot evolve a coherent lexical matrix.

#### ACKNOWLEDGEMENTS

Support from the Packard Foundation, the Leon Levy and Shelby White Initiatives Fund, the Florence Gould Foundation, the Ambrose Monell Foundation, the Alfred P Sloan Foundation and the National Science Foundation is gratefully acknowledged.

### APPENDIX A: SUB-OPTIMAL SOLUTIONS

For the sake of completeness, we will consider fixed points of system (11), corresponding to the eigenvalues  $\langle m \rangle = (n-l)q^{n-l}$  with  $1 \leq l \leq n$ . We have  $x_n = x_{n-1} = \dots = x_{n-l+1} = 0$ , and the system for the rest of the variables is exactly (11) with  $n$  replaced by  $n-l$ . All the sub-optimal solutions (for each  $l$ ,  $1 \leq l \leq n$ ) can be written down:

$$x_m = 0, \quad n-l+1 \leq m \leq n, \quad (\text{A1})$$

$$x_{n-l} = l^l / l! \prod_{j=0}^{n-l-1} (q^{n-l-j} - j/n - l), \quad (\text{A2})$$

$$x_m = (q^{n-l-m} - m/n - l)^{-1} \sum_{j=m+1}^l C_j^m (1-q)^{j-m} j/n - lx_j, \quad (\text{A3})$$

$$1 \leq m \leq n-l-1,$$

$$x_0 = 1 - \sum_{j=1}^n x_j. \quad (\text{A4})$$

These solutions correspond to the situation where all the  $n$  signals are present in the language of the population, but no one knows all of them. There are  $n$  such non-optimal solutions. The value of  $n-l$  is the maximum number of associations known to a single person.

To check stability let us perturb these solutions by  $y_k e^{\Gamma t}$ ,  $0 \leq k \leq n$  and linearize equations (11). The equations for  $\dot{x}_m$  with  $n-l+1 \leq m \leq n$  do not contain terms  $y_k$  with  $0 \leq k \leq n-l$ , and can be considered separately. It is easy to show that these equations lead to the growth rate  $\Gamma = (n-l)/nq^{2(n-l)}[(n-l+1)q - (n-l)q]$ , which means that the solution is unstable for  $q > 1 - (n-l)/(n-l+1)$ . On the other hand, the equations for  $\dot{x}_m$  with  $0 \leq m \leq n-l$  can also be decoupled from the rest of the equations if we take  $y_k = 0$  for  $n-l+1 \leq k \leq n$ . The corresponding system is exactly the perturbed system (11) where  $n$  is replaced by  $n-l$ . It suggests instability for all  $q < 1 - 1/(n-l)$ . Moreover, for the correct instability criterion, the variables  $x_k$  with  $0 \leq k \leq n-l$  have to be perturbed in a more general way, see equation (37) and the argument below. Such perturbations lead to the instability for all  $q < q_c \approx 1 - 1/(2(n-l))$ . The two regions of instability intersect and cover the entire interval  $0 \leq q \leq 1$ , which means that solutions (A1)–(A4) are unstable.

This proves that out of the  $n+1$  fixed points of system (11) corresponding to  $\langle m \rangle = (n-l)q^{n-l}$ ,  $0 \leq l \leq n$ , only one can be stable. The stable fixed point, or the optimal solution, corresponds to the largest eigenvalue  $\langle m \rangle = nq^n$ , i.e.,  $l = 0$ .

**APPENDIX B: THE BASIN OF ATTRACTION OF OPTIMAL SOLUTIONS**

Here we will show that if  $q$  is close to 1, then optimal solutions with low values of  $n$  have a very small domain of attraction. We will use the system (41), (42) with  $n - 1$  replaced by  $n$ :

$$\dot{v}_k = q^{k+1} \sum_{j=k}^n v_j C_j^k (1-q)^{j-k} \left( \frac{j}{n} \mathcal{A} + \mathcal{B} \right) - v_k \left( \frac{\mathcal{A}^2}{n} + \mathcal{B}^2 \right), \quad (\text{B1})$$

$$\begin{aligned} \dot{u}_m = & q^m \sum_{j=m}^n u_j C_j^m (1-q)^{j-m} \frac{j}{n} \mathcal{A} \\ & + q^m \sum_{j=m}^n v_j C_j^m (1-q)^{j-m+1} \left( \frac{j}{n} \mathcal{A} + \mathcal{B} \right) - u_m \left( \frac{\mathcal{A}^2}{n} + \mathcal{B}^2 \right) \end{aligned} \quad (\text{B2})$$

with  $0 \leq k \leq n$ ,  $1 \leq m \leq n$  to estimate the domains of attraction of various solutions. Here  $v_k$  is the fraction of people who know the  $(n+1)$ st signal plus any other  $k$  signals;  $u_m$  is the fraction of people who do not know the  $(n+1)$ st signal but know  $m$  other signals. System (B1), (B2) is convenient for testing the stability of the optimal solution (15)–(17) with respect to adding new signals to the population's vocabulary. Indeed, a fixed point of the system is the solution  $v_m = 0$ ,  $u_m = x_m^{\text{opt}}$ ,  $0 \leq m \leq n$ , where  $x_m^{\text{opt}}$  is solution (15)–(17). This means that the maximum number of signals people know is  $n$ , and we consider the possibility of the  $(n+1)$ st signal introduced in the language. Let us perturb the optimal solution:

$$v_m = 0 + \mathcal{V}_m, \quad u_m = x_m^{\text{opt}} + \mathcal{U}_m. \quad (\text{B3})$$

For the purposes of a linear stability analysis, we can set  $\mathcal{V}_m = v_m^{(1)} e^{\Gamma t}$ ,  $\mathcal{U}_m = u_m^{(1)} e^{\Gamma t}$ . This results in the stability criterion  $q \geq 1 - 1/n$ , i.e., for these values of  $q$  the solution with  $n$  signals is stable. Let us look at the growth rate of the optimal solution with  $n$  signals. From equation (B1) for  $k = n$  we obtain:

$$\Gamma v_n^{(1)} = q^{2n} v_n^{(1)} n (q - 1). \quad (\text{B4})$$

This is satisfied by  $v_n^{(1)} = 0$  or

$$\Gamma = q^{2n} n (q - 1). \quad (\text{B5})$$

It is the latter solution that we will concentrate on. The growth rate,  $\Gamma$ , is always non-negative. For  $|1 - q| \sim 1/n$  it is of order one,  $|\Gamma| \sim 1$ . However for  $|1 - q| \ll 1/n$ ,  $|\Gamma| \ll 1$ , and  $\Gamma = 0$  for  $q = 1$ . In the limit of absolute learning accuracy, this solution is neutrally stable, and its perturbations are only slightly damped when  $q$  is close to 1.

In order to obtain a crude estimate of the attraction domain let us assume that  $q$  is very close to 1, i.e.,

$$|1 - q| \ll 1/n. \tag{B6}$$

It is convenient to introduce a small parameter,

$$\beta = n(1 - q). \tag{B7}$$

Let us assume that  $\mathcal{V}_m \sim \beta$  and  $\mathcal{U}_m \sim \beta$  for all  $m$ . Then the expression on the right-hand side of equation (B4) is of order  $\beta^2$  and for consistency one must keep quadratic terms. Since the resulting equations are nonlinear, we cannot assume an exponential time dependence of the perturbation anymore. We have up to the second order in  $\beta$ :

$$\dot{\mathcal{V}}_n = -q^{2n}\mathcal{V}_n\beta + \mathcal{V}_n \left[ q^n \left( \sum_{l=0}^n l(\mathcal{U}_l + \mathcal{V}_l) + \sum_{l=0}^n \mathcal{V}_l \right) - 2q^{n-1} \sum_{l=0}^n l(\mathcal{U}_l + \mathcal{V}_l) \right]. \tag{B8}$$

The right-hand side is a function of  $q$ . Because of condition (B6), we can replace  $q$  by 1, and the error will be of order  $\beta$ , i.e., can be ignored in the current approximation. We have

$$\dot{\mathcal{V}}_n = -q^{2n}\mathcal{V}_n\beta + \mathcal{V}_n(\Delta\mathcal{B} - \Delta\mathcal{A}), \tag{B9}$$

where  $\Delta\mathcal{B}$  and  $\Delta\mathcal{A}$  are perturbations of the quantities  $\mathcal{B}$  and  $\mathcal{A}$  around their value at the optimal solution, i.e.,

$$\Delta\mathcal{B} = \sum_{l=0}^n \mathcal{V}_l, \quad \Delta\mathcal{A} = \sum_{l=0}^n l(\mathcal{U}_l + \mathcal{V}_l). \tag{B10}$$

Here  $(\mathcal{V}_1, \dots, \mathcal{V}_n, \mathcal{U}_1, \dots, \mathcal{U}_n)$  is the eigenvector of the linearized problem corresponding to the eigenvalue (B5). Once again, this eigenvector is a function of  $q$ . We can expand it around  $q = 1$  by  $\mathcal{V}_l = \mathcal{V}_l^{q=1} + O(\beta^2)$ ,  $\mathcal{U}_l = \mathcal{U}_l^{q=1} + O(\beta^2)$ . The correction will not contribute to equation (B9) because it gives rise to terms of the order of  $\beta^3$ . Hence, we only need to find the eigenvector of the linear problem with  $q = 1$ , corresponding to  $\Gamma = 0$ . In other words, we need to rewrite system (B1), (B2) for  $q = 1$ , substitute the solution in the form (B3) with  $\Gamma = 0$  and solve for the eigenvector. We have:

$$\dot{\mathcal{V}}_k^{q=1} = 0 = \mathcal{V}_k^{q=1}(k - n), \tag{B11}$$

$$\dot{\mathcal{U}}_k^{q=1} = 0 = \mathcal{U}_k^{q=1}(k - n) + x_k^{\text{opt}} \Delta\mathcal{A}^{q=1} [(k - n)/n - 1], \tag{B12}$$

$$1 \leq k \leq n.$$

From equations (B11) it follows that  $\mathcal{V}_k^{q=1} = 0$  for  $k < n$ . Therefore from (B10) we have  $\Delta\mathcal{B}^{q=1} = \mathcal{V}_n^{q=1}$ . Next, we remember that for  $q = 1$ ,  $x_k^{\text{opt}} = \delta_{kn}$ . From equations (B12) we obtain  $\Delta\mathcal{A}^{q=1} = 0$ . Finally, we can rewrite equation (B9) as

$$\dot{\mathcal{V}}_n = -\beta\mathcal{V}_n + \mathcal{V}_n^2. \quad (\text{B13})$$

It follows that if  $\mathcal{V}_n > \beta$  then the perturbation grows, that is the optimal solution with  $n$  signals becomes *nonlinearly* unstable. Therefore, the domain of attraction for the optimal solution with  $n$  signals is of the order of  $n(1 - q)$ , i.e., it shrinks to zero as  $q$  approaches one.

## REFERENCES

- Aoki, K. and M. W. Feldman (1987). Towards a theory for the evolution of cultural communication: Coevolution of signal transmission and reception. *Proc. Natl. Acad. Sci. USA* **84**, 7164–7168.
- Aoki, K. and M. W. Feldman (1989). Pleiotropy and preadaptation in the evolution of human language capacity. *Theor. Popul. Biol.* **35**, 181–194.
- Bickerton, D. (1990). *Species and Language*, Chicago: University of Chicago Press.
- Brandon, R. N. and N. Hornstein (1986). From icons to symbols: Some speculations on the origins of language. *Biol. Philos.* **1**, 169–189.
- Cangelosi, A. and D. Parisi (1998). The emergence of a “language” in an evolving population of neural networks. *Connection Science* **10**, 83–97.
- Cheney, D. L. and R. M. Seyfarth (1990). *How Monkeys See the World: Inside the Mind of Another Species*, Chicago: University of Chicago Press.
- Deacon, T. (1997). *The Symbolic Species*, New York: W. W. Norton.
- Eigen, M. and P. Schuster (1979). *The Hypercycle: A Principle of Natural Self-Organization*, Berlin: Springer-Verlag.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*, New York: Dover Publisher Inc.
- Hallowell, A. I. (1960). Self, society and culture in phylogenetic perspective, in *Evolution after Darwin*, Vol. II, The Evolution of Man, S. Tax (Ed.), Chicago: University of Chicago Press.
- Hauser, M. D. (1997). *The Evolution of Communication*, Cambridge, MA: MIT Press.
- Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* **77**, 187–222.
- Lieberman, P. (1992). On the evolution of human language, in *The Evolution of Human Languages, SFI Studies in the Science of Complexity*, J. A. Hawkins and M. Gell-Mann (Eds), Redwood City, CA: Addison-Wesley, pp. 21–47.
- Macedonia, J. M. and C. S. Evans (1993). Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethnol* **93**, 177–197.
- Miller, G. A. (1996). *The Science of Words*, New York: Scientific American Library.

- Nowak, M. A. (2000). The basic reproductive ratio of a word, the maximum size of a lexicon. *J. Theor. Biol.* **204**, 179–189.
- Nowak, M. A., N. L. Komarova and P. Niyogi (2001). Evolution of universal grammar. *Science* **291**, 114–118.
- Nowak, M. A. and D. C. Krakauer (1999). The evolution of language. *Proc. Natl. Acad. Sci. USA* **96**, 8028–8033.
- Nowak, M. A., J. B. Plotkin and V. A. A. Jansen (2000). The evolution of syntactic communication. *Nature* **404**, 495–498.
- Nowak, M. A., J. Plotkin and D. Krakauer (1999). The evolutionary language game. *J. Theor. Biol.* **200**, 147–162.
- Pinker, S. (1995). *The Language Instinct*, New York: Harper Perennial.
- Saussure, F. de (1983). *Course in General Linguistics*, C. Bally and A. Sechehaye (Eds), Translated and Annotated by Roy Harris, London: Duckworth.
- Smith, W. J. (1977). *The Behavior of Communicating*, Cambridge, MA: Harvard University Press.
- Smith, W. J. (1997). The Behavior of Communicating, after twenty years, in *Perspectives in Ethnology*, Vol. 10, D. H. Owings, M. D. Beecher and N. S. Thompson (Eds), New York: Plenum Press.
- Sperber, D. and D. Wilson (1995). *Relevance: Communication and Cognition*, 2nd edn, Oxford: Blackwell.
- Steels, L. (1996). Self-organizing vocabularies, in *Proceedings of Alife V*, C. Langston (Ed.), Nara, Japan.
- Steels, L. and F. Kaplan (1998). Spontaneous lexicon change, in *Proceedings of COLING-ACL*, Montreal. pp. 1243–1249.
- Steels, L. and P. Vogt (1997). Grounding adaptive language games in robotic agents, in *Proceedings of the Fourth European Conference on Artificial Life*, P. Husbands and I. Harvey (Eds), Cambridge, MA: MIT Press.
- Trapa, P. E. and M. A. Nowak (2000). Nash equilibria for an evolutionary language game. *J. Math. Biol.* , to appear.
- Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication* **18**, 47–67.
- Wray, A. (2000). Holistic utterances in protolanguage: the link from primates to humans, in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy and J. Hurford (Eds), Cambridge, New York: Cambridge University Press.

Received 27 June 2000 and accepted 13 December 2000