THE ROYAL SOCIETY

doi 10.1098/rspb.2001.1629

Natural selection of the critical period for language acquisition

Natalia L. Komarova[1,2] **and Martin A. Nowak**[1]

[1]*Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA*
[2]*Department of Mathematics, University of Leeds, Leeds LS2 9JT, UK*

The language acquisition period in humans lasts about 13 years. After puberty it becomes increasingly difficult to learn a language. We explain this phenomenon by using an evolutionary framework. We present a dynamical system describing competition between language acquisition devices, which differ in the length of the learning period. There are two selective forces that play a role in determining the critical learning period: (i) having a longer learning period increases the accuracy of language acquisition; (ii) learning is associated with certain costs that affect fitness. As a result, there exists a limited learning period which is evolutionarily stable. This result is obtained analytically by means of a Nash equilibrium analysis of language acquisition devices. Interestingly, the evolutionarily stable learning period does not maximize the average fitness of the population.

Keywords: universal grammar; language evolution; game theory; evolutionary dynamics; learning; fitness

1. INTRODUCTION

The ability of language acquisition in humans lasts roughly until the onset of reproduction, which happens approximately at the age of 13 (Lenneberg 1967). During this period children can learn a language (or several languages) with relative ease; after the age of 13 it becomes increasingly hard to acquire a language, and the result of learning becomes less and less perfect (Ingram 1989). Various explanations of this fact have been suggested, including the maturation of language circuits during a child's early years (Huttenlocher 1990; Bates *et al.* 1992; Locke 1993; Pinker 1994). It is important, however, to develop an evolutionary description of this phenomenon. Several recent studies have addressed this problem.

Hurford (1991) developed a numerical model where each individual was characterized by a certain language acquisition profile. A genetic algorithm was employed to evolve the most efficient profile. Knowing a language had a positive correlation with the individual's reproductive success. The language was assumed to have a certain 'size', and if an individual had acquired all of the language by a certain age, no further language acquisition was possible. As knowing 'more' of the language increased the person's fitness, learning all of the language during the first (non-reproductive) stage of the individual's life was selected for. Keeping up the theoretical ability to learn the language after the whole of the language had been learned did not change the person's lingual abilities. It was therefore evolutionarily neutral and could be eliminated by a random drift. As a result, most individuals learned their language by puberty as a result of this evolutionary process.

In the study by Hurford & Kirby (1999), the same approach was modified to include the possibility of innovation, i.e. expanding the language size by some individuals in the process of evolution. As a result, the amount of language available to individuals grew, and so did the speed of language acquisition. Forced by natural selection, the age of full language acquisition (defined as size/speed) reached puberty and then remained constant throughout generations. In these two papers no costs of learning were included in the simulations, which led to a somewhat puzzling result of an unbounded growth of both the language size and the speed of acquisition.

The role of learning costs has been studied by many authors (see, for example, Mayley (1996)). In Cecconi *et al.* (1996), the costs of learning were incorporated in a genetic algorithm which modelled the evolution of learning behaviour in a group of agents (neural networks). The age of reproduction of the agents was assumed to coincide with the end of their learning period. The gene for the age of reproduction onset was passed down the generations, and the learning cycle of immature individuals was modelled explicitly. The agents' ability to learn, and some other features of their phenotype such as characteristics of the neural network architecture, were also inherited genetically. An empirical 'energy' parameter was introduced, and all functions of agents, e.g. reproduction, parenting and learning, were assigned a cost which was subtracted from their initial energy level throughout their lives. As a result of non-zero learning costs, there was an evolutionary pressure to keep the maturation period short. Because the morphology of the neural nets was allowed to change via mutations, after a certain (large) number of generations the ability to learn was outcompeted by the corresponding in-built characteristics of the organisms; i.e. the Baldwin effect, or assimilation, was taking place. In the present study we shall not include the possibility of replacing adaptive characteristics by inherited features, assuming that this process takes much longer than the time-scale considered in our model.

In this paper we introduce an evolutionary model that incorporates some of the above ideas in a very simple way (see also Cavalli-Sforza & Feldman (1981)).

© 2001 The Royal Society
Received 27 October 2000 *Accepted* 14 February 2001

Namely, we assume that:

(i) successful communication increases an individual's biological fitness;
(ii) the ability to learn is genetically inherited;
(iii) learning is costly;
(iv) there is no assimilation of learnable characteristics on the time-scales of interest.

The model presented below is amenable to an analytical treatment. The main result is the existence of a critical learning period that is an evolutionarily stable strategy (ESS), or Nash equilibrium (Nash 1950; Maynard Smith 1982). It is defined by two competing forces: if the learning period is too short, then the result of learning is too far from perfect, which reduces the individual's fitness; if the learning period is too long, it reduces the reproduction rate because keeping up the ability to learn is very costly (for the learning individual, its parents or both). There is a learning period that optimizes the interplay between these two factors, yielding an evolutionary equilibrium which cannot be invaded by any other learning period. The evolutionarily stable learning period does not optimize the average fitness of the population.

In the next section we shall describe the model, in § 3 we present the results, while § 4 is reserved for conclusions and discussion. Most of the details of the analysis are presented in Appendix A.

2. LANGUAGE ACQUISITION DEVICE AND EVOLUTIONARY DYNAMICS

We consider a group of individuals that has a constant size. The individuals reproduce according to their language-related fitness, and the children learn the language of their parent (for simplicity, asexual reproduction is assumed). The learning procedure is defined by means of a language acquisition device (LAD), which is genetically inherited.

The framework developed here can be adopted for studying the critical period for acquisition of various aspects of human language, such as phonology, lexicon and grammar, as well as animal communication systems. In this work, we shall concentrate on grammar acquisition. For our purposes the terms 'language acquisition device' and 'universal grammar' (Chomsky 1980, 1993) can be used synonymously.

Universal grammar (UG) specifies the range of grammatical hypotheses that children entertain during language acquisition (see Wexler & Culicover (1980); Lightfoot (1982)). Many linguists believe that universal grammar is the consequence of specific genetically encoded structures within the human brain (Hornstein & Lightfoot 1981; Pinker & Bloom 1990; Jackendoff 1997). It is important to note the difference between universal grammar and the grammar of the spoken language. The former is a hardwired property of the child's brain, whereas the latter is the specific grammar that the child learns during its maturation phase.

Let us denote all possible spoken grammars by $G_1, \ldots, G_n$, where n is some finite integer. During the language acquisition phase, each child has to 'decide' which grammar is the actual grammar of its parent, based on a finite number of the input sentences, b, that the child receives during the language acquisition period (see Osherson *et al.* (1986); Lightfoot (1991, 1999); Niyogi & Berwick (1996, 1997); Niyogi (1998)). Note that the number of candidate grammars can also be infinite, provided that children have a prior probability distribution specifying that some grammars are more likely than others. In this paper, however, we shall restrict our analyses to the case of a finite search space, where all candidate grammars are equally likely at the beginning of the learning process.

Mistakes in learning take place. Denote by $Q_{ij}(b)$ the probability that a child learning from a parent with grammar G_i will end up speaking grammar G_j. The matrix $Q(b) \equiv [Q_{ij}(b)]$ is equivalent to the mutation matrix in quasi-species theory or population genetics (Eigen & Schuster 1979; Aoki & Feldman 1987), except here it is not connected with the genetic inheritance, but rather with the learning (copying) precision, or *learning accuracy*. The value of b is a convenient measure of the length of the learning period, if we assume that the input sentences are delivered at a roughly constant rate. The existence of an evolutionarily stable value of b will suggest that there is a natural selection for a critical period of language acquisition.

Formally, each LAD is characterized by:

(i) a search space of a fixed size, i.e. the sets $G_1, \ldots, G_n$; each of the grammars can be the spoken grammar of the language;
(ii) the number of learning events, b;
(iii) and a learning mechanism, which gives the matrix $Q(b)$.

The value of b has two important consequences for the evolution of learning. Firstly, the matrix $Q(b)$ can be explicitly calculated if the grammars $G_1, \ldots, G_n$ and the learning mechanism are specified (Komarova *et al.* 2001). For the present study we do not need to specify the precise form of $Q(b)$; the only important property that we are going to use is the following: $\lim_{b\to\infty} Q_{ii}(b) = 1$ for all $1 \leqslant i \leqslant n$. This means that the larger the number of input sentences, the better are the chances of learning the grammar perfectly.

The second effect of b is the *reproductive rate*, $r(b)$. This quantity includes implicitly all the costs of learning that depend on the length of the learning period (Mayley 1996). It may consist of several components. We can conceive that while the individual is concentrated on learning, the resources (which are always limited) are in some sense taken away from the reproductive function. More precisely, (i) time and energy get directly invested in learning; (ii) the ability to memorize linguistic items requires a sophisticated memory storage system, which needs constant maintenance; and (iii) the brain has a limited capacity for processing information, i.e. intensive learning may decrease the individual's performance in other areas of life (see Dukas (1998)). Also, if an individual needs a lot of 'help' in learning from its parent, this means that the parent cannot go on reproducing while its child is still in the maturation stage. In any case, the longer the learning period lasts, the more energy it uses, which could otherwise be spent on reproduction. The corresponding reproductive rate, $r(b)$, reflects these mechanisms. Again, the exact form of this function is not important for the qualitative results provided it decays with b. An example of a simple empirical parameterization will be considered in the next section.

Because the purpose of this work is to understand the selection of a critical language acquisition period, we shall assume that all the LADs present in the population only differ in the number of learning events, b, and they are identical otherwise. This means that all of the LADs consist of the n grammars $G_1, \ldots, G_n$. The strategy we are going to use is as follows. First we shall assume that there are only *two* different LADs present in the population, and perform the analysis of the corresponding system. The result of this analysis will reveal which one of the two LADs will be selected (in other words, which of the values of b will invade when the two of them are present). Then it will be possible to find the LAD that cannot be invaded by *any* other LAD, i.e. the evolutionarily stable b. Note that the size of the search space (i.e. the parameter n) and the grammars $G_1, \ldots, G_n$ are assumed to be fixed in this model and do not evolve.

We shall denote the two LADs present in the population by U_1 and U_2. Let the vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ stand for the fraction of people who speak grammars G_1 through G_n of U_1, and the vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ denote the fraction of people who speak grammars G_1 through G_n of U_2. The total population size is scaled to unity: $\sum_{i=1}^{n} (x_i + y_i) = 1$. The language acquisition devices are inherited genetically. The system of $2n$ equations describing the coexistence of U_1 and U_2 is as follows (Nowak *et al.* 2001):

$$\dot{x}_i = r(b_1) \sum_{j=1}^{n} x_j f_j^{(1)} Q_{ji}(b_1) - \phi x_i, \tag{1}$$

$$\dot{y}_i = r(b_2) \sum_{j=1}^{n} y_j f_j^{(2)} Q_{ji}(b_2) - \phi y_i, \qquad i = 1, \ldots, n. \tag{2}$$

The left-hand side of these equations contains the time-derivative of the frequency of each grammar. We shall now explain the terms in the right-hand side of equations (1) and (2).

First we note that the grammars $G_1, \ldots, G_n$ do not have to be specified precisely for the purposes of the present study. However, we do need to use the information about the *pairwise intersections* of the grammars. Denote by a_{ij} the probability that a speaker who uses grammar G_i formulates a sentence that is compatible with grammar G_j. Hence, the matrix $A = [a_{ij}]$ contains the pairwise relationship among the n grammars. We have $0 \leqslant a_{ij} \leqslant 1$ and $a_{ii} = 1$.

We assume that there is a reward for mutual understanding. The pay-off for an individual using G_i communicating with an individual using G_j is given by $F(G_i, G_j) = (1/2)(a_{ij} + a_{ji})$, which is just the average probability of mutual understanding. Note that $F(G_i, G_i) = 1$ (hence, all n grammars are equally powerful and allow the same level of communication). The average fitness of individuals who use grammar G_j (of U_1 and U_2 respectively) is found by

$$f_j^{(1)} = 1/2\left(\sum_{k=1}^{n} (a_{jk} + a_{kj}) x_k + \sum_{m=1}^{n} (a_{j,n+m} + a_{n+m,j}) y_m\right), \tag{3}$$

$$f_j^{(2)} = 1/2\left(\sum_{k=1}^{n} (a_{n+j,k} + a_{k,n+j}) x_k + \sum_{m=1}^{n_2} (a_{n+j,n+m} + a_{n+m,n+j}) y_m\right). \tag{4}$$

Here the pairwise relationship matrix A has the size $2n \times 2n$ because we have two sets of n grammars which belong to U_1 and U_2. In the present study, the two sets of grammars are identical, and we can set

$$f_j^{(1)} = f_j^{(2)} \equiv f_j = 1/2 \sum_{k=1}^{n} (a_{jk} + a_{kj})(x_k + y_k), \tag{5}$$

for all j. Someone who uses a grammar that is understood by others has a better performance during life history in terms of survival probability or reproductive success. Individuals who communicate successfully leave more offspring (thus the factors f_j in equations (1) and (2)), who in turn learn their language. This puts the problem of grammar acquisition in an evolutionary context (Nowak & Krakauer 1999; Nowak *et al.* 1999, 2000).

Finally, the quantity ϕ in equations (1) and (2) is the average fitness of the population. It is the average probability that a sentence said by one person is understood by another person (or the *grammatical coherence*), weighted by the individuals' reproductive rate:

$$\phi = \sum_{j=1}^{n} (r(b_1) f_j^{(1)} x_j + r(b_2) f_j^{(2)} y_j). \tag{6}$$

The analysis of system (1)–(2) is presented in Appendix A. In § 3, we shall outline the main results and present some examples.

3. AN EVOLUTIONARILY STABLE LEARNING PERIOD

As was argued before, the reproduction rate, $r(b)$, decays with b, and the learning accuracy for each grammar, $Q_{ii}(b)$, grows with b, so the functions $r(b)Q_{ii}(b)$ have a maximum. Let us define b_i^* as the value of b that corresponds to the maximum of the function $r(b)Q_{ii}(b)$. The following result holds: if b_m^* corresponds to the highest among the maxima of functions $r(b)Q_{ii}(b)$, then the LAD with $b = b_m^*$ is evolutionarily stable with respect to any other LAD with a different number of sampling events. The spoken grammar in this case is G_m. More generally, if G_i is the spoken grammar, then under some mild conditions the evolutionarily stable LAD has $b = b_i^*$. These results are exact in the limit of large values of n.

Intuitively speaking, as b increases, two things happen: (i) the learning accuracy increases, and (ii) the reproductive rate decreases. These are the natural requirements that should guarantee that there is a selection for intermediate values of b.

To illustrate this we chose the memoryless learner algorithm of grammar acquisition to define the learning accuracy function (Niyogi 1998). This algorithm describes the interaction between a learner and a teacher. Suppose the teacher uses grammar G_k. The learner starts with a randomly chosen hypothesis, G_i. The teacher generates sentences consistent with G_k. Provided that these sentences are also consistent with G_i, the learner maintains his hypothesis. If a sentence occurs that is not

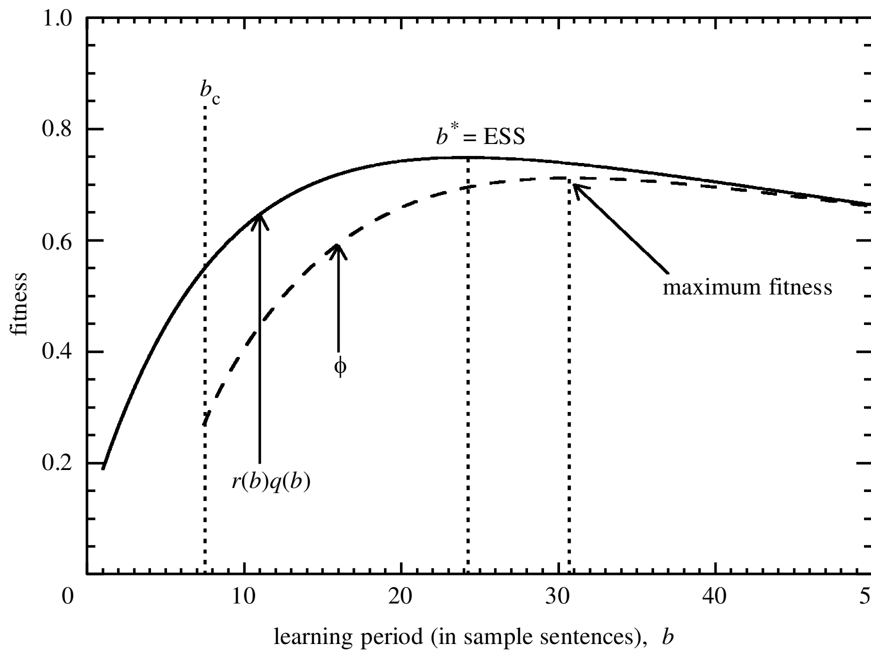

Figure 1. The evolutionary selection of the number of sampling events in the case of a fully symmetrical system; $a = 0.5$, $n = 40$, $c_1 = 1$, $c_2 = 0.001$. The value of b which corresponds to the ESS coincides with b^*.

consistent with G_i, the learner picks at random a different hypothesis, G_j. After b sample sentences, the process stops, and the learner remains with his current hypothesis. This learning algorithm defines a Markov process. The transition probabilities depend on the teacher's grammar and on the a_{ij} values.

At equilibrium, all the individuals in the population have the same LAD (the one with $b = b_k^*$). The corresponding stable solution of system (1)–(2) is a fixed point which can be described as a *one-grammar solution* (Komarova *et al.* 2001). This means that the majority of the population have the same grammar (say, G_k, which is called the *dominant grammar*). Furthermore, there are small fractions of people who speak any of the other grammars (*secondary grammars*), $G_1, \ldots, G_{k-1}, G_{k+1}, \ldots, G_n$. The frequencies of these grammars are small in comparison with the frequency of the dominant grammar, G_k. The exact proportion of the dominant grammar is defined by the learning accuracy, $Q(b_k^*)$. The higher the learning accuracy, the closer the frequency of G_k is to unity. If $Q(b_k^*)$ is an identity matrix, the entire population has exactly the same grammar, G_k.

(a) *Fully symmetrical systems*

Let us impose the following symmetry condition on the pairwise intersection matrix A: $A_{ij} = a$ for all $i \neq j$. The Q matrix in this case is also symmetrical: we have $Q_{ii} = q$ for all i, and in the case of a memoryless learner algorithm we obtain (Komarova *et al.* 2001)

$$q(b) = 1 - \left(1 - \frac{1-a}{n-1}\right)^b \frac{n-1}{n}. \quad (7)$$

This is a monotonically growing function of b. Our choice of the reproduction rate dependence on b is $r(b) = (c_1 + c_2 b)^{-1}$, where $c_{1,2}$ are some positive constants. This function monotonically decreases with b. The function $q(b)r(b)$ is a one-humped function (see figure 1, solid line). The maximum of this function corresponds to the value b^*, the number of sampling events that cannot be invaded by any other LAD.

Let ϕ_0 denote the average fitness of the population at equilibrium. It is equal to the grammatical coherence of a one-grammar solution multiplied by the reproductive rate. The function ϕ_0 is plotted in figure 1 with a dashed line. The maximum of ϕ_0 gives the number of sampling events, b, which leads to the maximum average fitness. Note that the maxima of the functions qr and ϕ_0 do not coincide. This means that the number of sampling events that guarantees the evolutionary stability of the corresponding LAD, does not in general lead to the maximum possible fitness of the population.

Now we ask the question: does the value b^* defined above always give an evolutionarily stable LAD? It turns out that a further restriction must be imposed. In the example of figure 2 we changed the parameters c_1 and c_2, so that the value of b that maximizes the function qr decreased below the *coherence threshold*, b_c. It can be shown that if $b < b_c$, the only equilibrium solution of system (1)–(2) is a *uniform solution*, where all grammars are spoken with similar frequencies. In contrast to one-grammar solutions, this solution does not correspond to any degree of coherence in the population. The value of b^* in this case optimizes the learning accuracy–reproduction curve, but it is too small to support coherence. The evolutionarily stable duration of the learning period (the optimal b) in the case when $b^* < b_c$ is given by b_c.

(b) *Asymmetric systems*

Now let us consider a system where the coefficients a_{ij} are arbitrary numbers between zero and unity. The functions $r(b)Q_{kk}(b)$ for an $n = 7$ case are shown in figure 3. We used the memoryless learner algorithm to calculate $Q_{ii}(b)$, and set $r(b) = (c_1 + c_2 \exp(\gamma b))^{-1}$. The values b_k^* are found by maximizing functions $r(b)Q_{kk}(b)$. As in the fully symmetrical case, the values b_k^* may fall below the

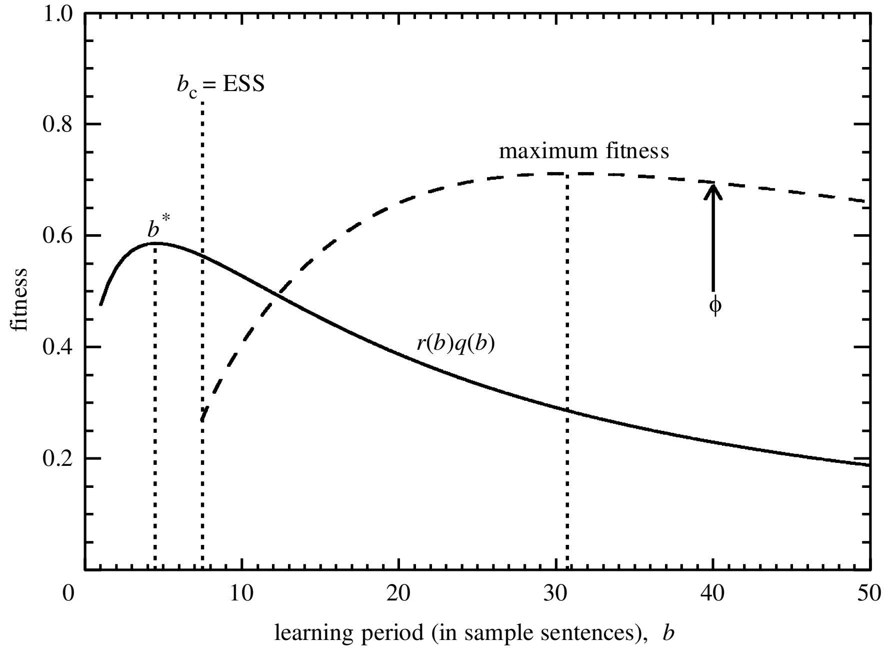

Figure 2. The evolutionary selection of the number of sampling events in the case of a fully symmetrical system. The parameters are as in figure 1 except $c_1 = 0.3$ and $c_2 = 0.1$. The value b^* is below the error threshold, so the ESS is the value b_c.

coherence threshold. Let us denote as b_k the threshold value of b such that for $b \geqslant b_k$, the one-grammar solution with G_k as the dominant grammar exists. It is convenient to introduce the functions

$$\alpha_k(b) = \begin{cases} r(b)Q_{kk}(b), & b \geqslant b_k \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

In general, the values b_k^* are defined to maximize the functions $\alpha_k(b)$.

For the choice of coefficients in figure 3, b_1^* corresponds to the highest of the maxima of the functions $\alpha_k(b)$. This means that the one-grammar solution with G_1 as the dominant grammar and $b = b_1^*$ cannot be invaded by any other LAD. Thus b_1^* defines the selected period of grammar acquisition for grammar G_1.

For G_k that correspond to other (lower) maxima of functions $\alpha_k(b)$, the criterion of stability is as follows. If the inequality

$$f_k(b_k^*)\alpha_k(b_k^*) \geqslant f_j(b_k^*)\alpha_j(b_j^*) \quad (9)$$

holds for all $j \neq k$, then the grammar G_k with $b = b_k^*$ is an ESS. This condition involves knowing the fitness of the spoken grammar G_k and the secondary grammar G_j evaluated at the one-grammar solution with G_k as the dominant grammar. If the values b_i^* for all i are well beyond the coherence threshold, i.e. $b_i^* \gg b_i$ for all i, then we have a much simpler condition for the stability of grammar G_k with $b = b_k^*$:

$$\alpha_k(b_k^*) \geqslant 1/2(a_{kj} + a_{jk})\alpha_j(b_j^*), \qquad \forall j. \quad (10)$$

This means that the one-grammar solution with G_k as the dominant grammar and $b = b_k^*$ may only be unstable towards a grammar that is very similar to G_k but is more efficient, i.e. it has a higher reproduction–accuracy curve. In the example of figure 3, condition (9) is satisfied for all b_i^*, i.e. all the grammars can be ESS if the optimal b is used.

4. CONCLUSIONS

It has been shown that an evolutionarily stable value of the number of sampling events, b, can be found for the evolutionary language game where the LADs have the same finite search space and learning mechanism but differ by the parameter b. The optimal b can be obtained by maximizing the functions $r(b)Q_{ii}(b)$ (the product of learning accuracy and reproductive rate), where each function $r(b)Q_{ii}(b)$ is restricted to the existence domain of the one-grammar solution with the dominant grammar G_i.

The main result of the paper, i.e. the existence of the number of learning events, which is an ESS, can be reinterpreted more generally. So far we have used the number of sampling events, b, as a convenient way to parameterize the family of LADs that we considered. Higher b corresponded to higher learning accuracy and lower reproductive rate. The optimization problem of the accuracy of learning versus its costs can be considered in a more general setting. It is instructive to contrast the following two strategies (LADs). The first assumes that a lot of energy gets invested in learning (this might depend on the time of learning, intensity of learning, the brain size or the existence of some sophisticated hard-wired imitation machinery). As a result, the learning accuracy is very high. The price to be paid for this learning precision is increased learning costs, which reduce the reproduction rate, r. The second strategy invests less energy in learning, and more energy is used for reproduction. The question that arises is still the same: is there an ESS, and if so, how can we find it?

The answer is obtained directly by generalizing our previous results. For each of the strategies (out of a

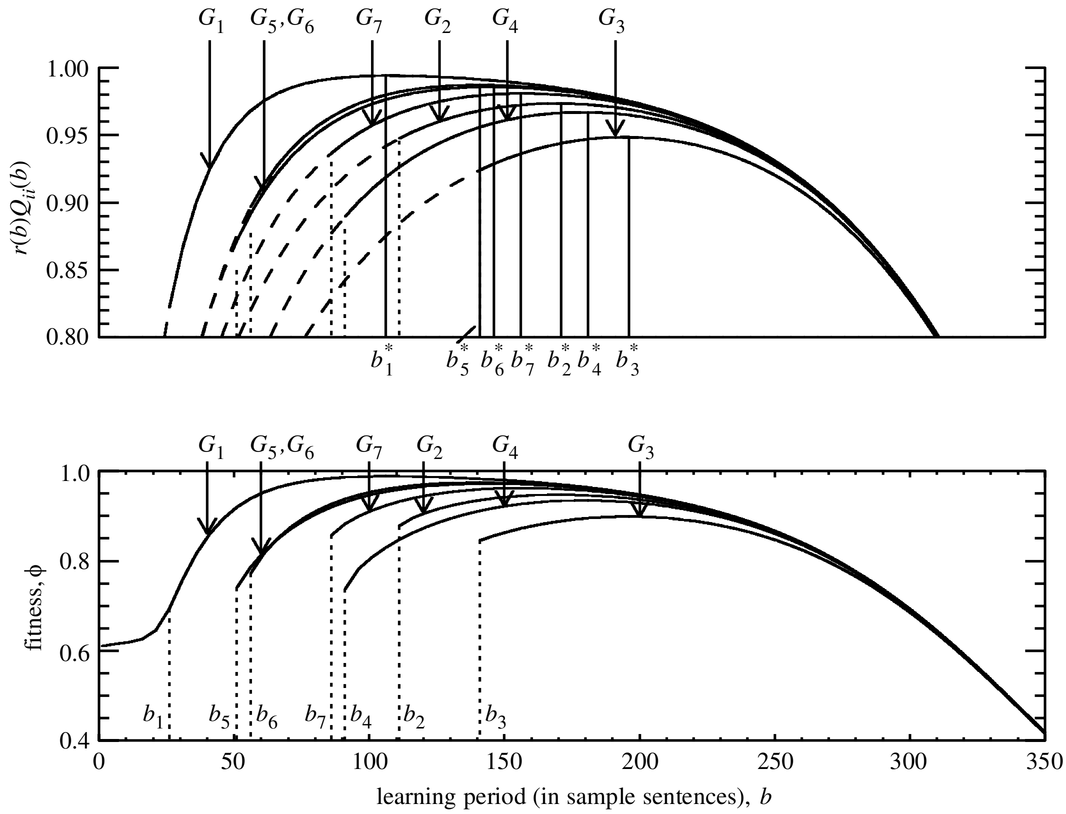

Figure 3. The evolutionary selection of the number of sampling events in the general case; a_{ij} are chosen from a uniform distribution between zero and unity, $n = 7$, and $r(b) = (c_1 + c_2 \exp(\gamma b))^{-1}$ with $c_1 = 1$, $c_2 = 0.0005$, $\gamma = 0.02$. The first plot presents the functions $r(b)Q_{ii}(b)$, the solid lines correspond to $\alpha_i(b)$. The second plot shows the average fitness corresponding to one-grammar solutions, for different dominant grammars. Each value b_i denotes the coherence threshold corresponding to the emergence of the one-grammar solution with the dominant grammar G_i. All the values b_i^* in this example correspond to ESS.

discreet or a continuous family), let us find the product of its learning accuracy and the reproduction rate. The winner is the strategy that maximizes this quantity.

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: ANALYSIS OF THE EVOLUTION EQUATIONS

Here we outline some details of the analysis of equations (1) and (2). Let us suppose that nobody in the population has U_2 (i.e. $y_j = 0$ for all j), and consider equation (1). Let us list some qualitative features of the dynamics for $n \gg 1$. Coherent communication can exist only if the the accuracy of learning is sufficiently high, that is if the matrix Q is not too far from the identity matrix. This means that we have $Q_{jj} \sim 1$ and $Q_{ji} \sim 1/n$ for $j \neq i$. There can be n stable one-grammar solutions. If the dominant grammar is G_k, then we have $x_k \sim 1$ and $x_j \sim 1/n$ for all the secondary grammars. This means that the shares of secondary grammars are very low, even though the *sum* of the shares of those grammars might be significant, i.e. $\sum_{j \neq k} x_j$ could be of the order of x_k. It is useful to rewrite equation (1) for x_k neglecting all the terms of order $1/n$:

$$\dot{x}_k = r(b_1)x_k f_k Q_{kk}(b_1) - \phi x_k. \quad \text{(A1)}$$

At equilibrium we have

$$\phi_0 = r(b_1) f_k Q_{kk}(b_1), \quad \text{(A2)}$$

where ϕ_0 is the average fitness corresponding to the one-grammar solution.

If only one LAD is allowed in the population, one-grammar solutions are stable given that b is higher than a threshold value. Let us consider the stability of one-grammar solutions with respect to general perturbations of system (1)–(2), including invasion of users of U_2. Initially, the share of users of U_2 is very low, so the equations for y_j decouple from the equations for x_j and can be considered separately. Neglecting all terms that are small in $1/n$, we can write the linearized equations:

$$\dot{y}_k = [r(b_2) f_k Q_{kk}(b_2) - \phi_0] y_k, \quad \text{(A3)}$$

$$\dot{y}_j = [r(b_2) f_j Q_{jj}(b_2) - \phi_0] y_j, \qquad j \neq k. \quad \text{(A4)}$$

Combining equations (A2), (A3) and (A4), we obtain the following conditions for stability of the one-grammar solution with G_k of U_1 against users of U_2:

$$r(b_1) f_k Q_{kk}(b_1) \geqslant r(b_2) f_j Q_{jj}(b_2), \qquad 1 \leqslant j \leqslant n. \quad \text{(A5)}$$

This allows us to find the conditions for the stability of the G_k with $b = b_1$ against the invasion of *any* other b. We have

$$f_k(b_1)r(b_1)Q_{kk}(b_1) \geqslant f_j(b_1)r(b)Q_{jj}(b) \quad 1 \leqslant j \leqslant n, \ \forall b. \quad \text{(A6)}$$

Here, the fitness f_i is evaluated from equation (5) at the one-grammar solution with G_k the dominant grammar and with $b = b_1$. The more general condition is given by inequality (9),

$$f_k(b_1)\alpha_k(b_1) \geqslant f_j(b_1)\alpha_j(b_1) \quad 1 \leqslant j \leqslant n, \ \forall b, \quad \text{(A7)}$$

where each function $r(b)Q_{ii}(b)$ is only considered within the domain of existence of the corresponding one-grammar solution. The average fitness of the users of the dominant and secondary grammars cannot be calculated explicitly; thus it is not easy to verify whether the stability condition holds. However, we can simplify this condition in certain cases.

(a) *Fully symmetrical systems*

Let us consider an example of a system where one-grammar solutions can be found explicitly. We assume that the intersection matrix, A, is fully symmetrical, i.e. $a_{ij} = a$ for all $i \neq j$. Then (for any reasonable learning mechanisms) the matrix Q also contains symmetries and is given by

$$Q_{ij}(b_1) = \begin{cases} q(b_1), & i = j, \\ (1 - q(b_1))/(n-1), & i \neq j, \end{cases}$$
$$Q_{ij}(b_2) = \begin{cases} q(b_2), & i = j, \\ (1 - q(b_2))/(n-1), & i \neq j. \end{cases} \quad \text{(A8)}$$

The concrete form of the function $q(b)$ is given by the learning mechanism.

If the entire population uses U_1, there are n identical one-grammar solutions corresponding to each of the grammars G_k (Komarova *et al.* 2001; Nowak *et al.* 2001). These solutions exist only if $q(b_1) \geqslant q_c$, where the threshold value of q is given by

$$q_c = \frac{2\sqrt{a}}{1 + \sqrt{a}} + O(1/n). \quad \text{(A9)}$$

The one-grammar solutions have the form

$$x_k = X, \quad x_i = (1 - X)/(n - 1), \quad \text{for } i \neq k; \quad y_j = 0 \quad \forall j, \quad \text{(A10)}$$

where X is given in terms of a, n and q. Also we can show that for a one-grammar solution with G_k as the dominant grammar,

$$f_k \geqslant f_j \quad \forall j. \quad \text{(A11)}$$

Using these inequalities and condition (A5), we can see that if

$$r(b_1)q(b_1) \geqslant r(b_2)q(b_2), \quad \text{(A12)}$$

then the one-grammar solution with $b = b_1$ is stable with respect to the invasion of users of U_2. Similarly, for the stability against *any* other LAD we need to have

$$r(b_1)q(b_1) \geqslant r(b)q(b) \quad \forall b. \quad \text{(A13)}$$

This is satisfied for $b_1 = b^*$, the value that maximizes the function $r(b)q(b)$. The LAD with $b = b^*$ is stable with respect to any other b. Note that because of the symmetries of this system, the optimal number of input sentences, b^*, does not depend on k, i.e. it is the same for all grammars.

(b) *General A matrices*

Now let us consider systems where the off-diagonal entries of the A matrix are arbitrary numbers between zero and unity. Again, let us suppose that there are two LADs, with the numbers of sampling events b_1 and b_2. The stability conditions for the one-grammar solution with the dominant grammar G_k of U_1 is given by inequalities (A7). We can use $f_k \geqslant f_j$ to write down a sufficient condition for stability:

$$\alpha_k(b_1) \geqslant \alpha_i(b_2), \quad 1 \leqslant i \leqslant n. \quad \text{(A14)}$$

From this condition it is possible to find a LAD that is stable with respect to the invasion of any other LAD. Let us consider the functions $\alpha_i(b)$ for all $1 \leqslant i \leqslant n$, and assume that each of these functions has one maximum. Let us find the value, b_m^*, which corresponds to the *highest* maximum among the maxima of the functions $\alpha_i(b)$. We have

$$\alpha_m(b_m^*) \geqslant \alpha_i(b), \quad 1 \leqslant i \leqslant n, \quad \forall b. \quad \text{(A15)}$$

For a LAD with $b = b_m^*$, the one-grammar solution with the dominant grammar G_m is stable with respect to invasion of any other b. Thus the value b_m^* gives the optimal length of the learning period for grammar G_m.

For other grammars, the optimal number of sample sentences can be found explicitly if we assume that $b_i^* \gg b_i$ for all i, i.e. if the maxima of the functions $\alpha_i(b)$ are well within the domain of existence of one-grammar solutions. In this case, we have $f_k \approx 1$ and $f_j \approx 1/2(a_{kj} + a_{jk})$, which leads to condition (10) (see § 3).

REFERENCES

Aoki, K. & Feldman, M. W. 1987 Toward a theory for the evolution of cultural communication: coevolution of signal transmission and reception. *Proc. Natl Acad. Sci. USA* **84**, 7164–7168.

Bates, E., Thal, D. & Janowsky, J. S. 1992 Early language development and its neural correlates. In *Handbook of neuropsychology*, vol. 6. *Child neurology* (ed. I. Rapin & S. Segalowitz), pp. 69–110. Amsterdam: Elsevier.

Cavalli-Sforza, L. L. & Feldman, M. W. 1981 *Cultural transmission and evolution: a quantitative approach*. Princeton University Press.

Cecconi, F., Menczer, F. & Belew, R. K. 1996 Maturation and the evolution of imitative learning in artificial organisms. *Adaptive Behav.* **4**, 29–50.

Chomsky, N. 1980 *Rules and representations*. New York: Columbia University Press.

Chomsky, N. 1993 A minimalist program for linguistic theory. In *The view from building 20* (ed. K. Hale & S. J. Keyser), pp. 1–52. Cambridge, MA: MIT Press.

Dukas, R. 1998 Constraints on information processing and their effects on behavior. In *Cognitive ecology: the evolutionary ecology of information processing and decision making* (ed. R. Dukas), pp. 89–127. University of Chicago Press.

Eigen, M. & Schuster, P. 1979 *The hypercycle. A principle of natural self-organisation*. Berlin: Springer.

Hornstein, N. R. & Lightfoot, D. W. 1981 *Explanation in linguistics*. London: Longman.

Hurford, J. R. 1991 The evolution of the critical period for language acquisition. *Cognition* **40**, 159–201.

Hurford, J. R. & Kirby, S. 1999 Co-evolution of language-size and the critical period. In *New perspectives on the critical period hypothesis and second language acquisition* (ed. D. Birdsong), pp. 39–63. Hillsdale, NJ: Lawrence Erlbaum.

Huttenlocher, P. R. 1990 Morphometric study of human cerebral cortex development. *Neuropsychologia* **28**, 517–527.

Ingram, D. 1989 *First language acquisition: method, description, and explanation*. New York: Cambridge University Press.

Jackendoff, R. 1997 *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Komarova, N. L., Niyogi, P. & Nowak, M. A. 2001 Evolutionary dynamics of grammar acquisition. *J. Theor. Biol.* **209**, 43–59.

Lenneberg, E. H. 1967 *Biological foundations of language*. New York: Wiley.

Lightfoot, D. 1982 *The language lottery, towards a biology of grammars*. Cambridge, MA: MIT Press.

Lightfoot, D. 1991 *How to set parameters: arguments from language change*. Cambridge, MA: MIT Press.

Lightfoot, D. 1999 *The development of language: acquisition, changes and evolution*. Malden, MA: Blackwell.

Locke, J. L. 1993 *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Mayley, G. 1996 Landscapes, learning costs and genetic assimilation. In *Evolutionary computation, evolution, learning and instinct: 100 years of the Baldwin effect* (ed. P. Turney, D. Whitley & R. Anderson), pp. 213–234. Special issue of *Evol. Comput.* **4**.

Maynard Smith, J. 1982 *Evolution and the theory of games*. Cambridge University Press.

Nash, J. F. 1950 Equilibrium points in n-person games. *Proc. Natl Acad. Sci. USA* **36**, 48–49.

Niyogi, P. 1998 *The informational complexity of learning*. Boston: Kluwer Academic.

Niyogi, P. & Berwick, R. C. 1996 A language learning model for finite parameter spaces. *Cognition* **61**, 161–193.

Niyogi, P. & Berwick, R. C. 1997 Evolutionary consequences of language learning. *Linguist. Phil.* **20**, 697–719.

Nowak, M. A. & Krakauer, D. C. 1999 The evolution of language. *Proc. Natl Acad. Sci. USA* **96**, 8028–8033.

Nowak, M. A., Krakauer, D. C. & Dress, A. 1999 An error limit for the evolution of language. *Proc. R. Soc. Lond.* B **266**, 2131–2136.

Nowak, M. A., Plotkin, J. B. & Jansen, V. A. A. 2000 The evolution of syntactic communication. *Nature* **404**, 495–498.

Nowak, M. A., Komarova, N. L. & Niyogi, P. 2001 Evolution of universal grammar. *Science* **291**, 114–118.

Osherson, D., Stob, M. & Weinstein, S. 1986 *Systems that learn*. Cambridge, MA: MIT Press.

Pinker, S. 1994 *The language instinct*, ch. 9. New York: Morrow.

Pinker, S. & Bloom, A. 1990 Natural language and natural selection. *Behav. Brain Sci.* **13**, 707–784.

Wexler, K. & Culicover, P. 1980 *Formal principles of language acquisition*. Cambridge MA: MIT Press.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.