(YD) epoch in the North Atlantic region and to the warming interval after the ACR in Antarctica.

Data from Vostok suggest an important role of the Southern Ocean in regulating the glacial-interglacial $CO_2$ changes (*5*). This role is confirmed by measurements from Taylor Dome for shorter time intervals in the last glaciation (*16*). The $CO_2$ increase in interval I, which occurred before any substantial warming in the Northern Hemisphere, is consistent with the present view of the role of the Southern Hemisphere for causing the $CO_2$ increase.

Methane starts to increase parallel to $CO_2$ in interval I. The methane increase is in agreement with the Greenland Ice Core Project (GRIP) record (*28*). The parallelism of the methane and $CO_2$ increase in interval I is somewhat surprising because the causes for methane variations are certainly different from those for $CO_2$. It is assumed that methane concentration changes were mainly due to changes of the extent and activity of wetlands in northern latitudes and the tropics (*29*). No substantial variations can be seen in the GRIP stable isotope record during this time period, but a small change of the methane production in low and mid-latitudes is not necessarily recorded in a Greenland temperature record. There is no obvious cause of the reduced rates of growth in $CO_2$ and methane between intervals I and II visible in the stable isotope records of Dome C or of GRIP.

The fast increases of $CO_2$ and methane concentrations between intervals II and III, at ~13.8 ky B.P. according to the Dome C time scale, correspond to the fast warming in the Northern Hemisphere observed at 14.5 ky B.P. on the GRIP time scale. This warming was probably caused by enhanced formation of North Atlantic Deep Water (NADW) (*30*), suggesting that the sudden $CO_2$ increase could have been caused by changes in thermohaline circulation. The methane increase, on the other hand, is thought to have been caused by an intensified hydrological cycle during the B/A warm phase, which led to an expansion of wetlands in the tropics and northern latitudes.

$CO_2$ decreased slightly during interval III and then increased during interval IV. The methane concentration follows the temperature evolution of the Northern Hemisphere in intervals III and IV as expected. The accelerated $CO_2$ increase at the end of interval IV probably is connected to the fast warming in the Northern Hemisphere rather than to any climate or environmental evolution in the Southern Hemisphere, because it is synchronous with the methane increase.

These data support the idea that the Southern Ocean was an important factor in regulating the $CO_2$ concentration during the last transition. However, the fast increases between intervals II

and III and at the end of interval IV show that additional mechanisms in the Northern Hemisphere influenced $CO_2$, presumably through changes in NADW formation.

### References and Notes

1. C. D. Keeling, T. P. Whorf, *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, 2000); an update is available at http://cdiac.esd.ornl.gov/trends/co2/sio-mlo.htm.
2. A. Neftel, E. Moor, H. Oeschger, B. Stauffer, *Nature* **315**, 45 (1985).
3. J.-M. Barnola *et al.*, *Tellus* **47B**, 264 (1995).
4. D. M. Etheridge *et al.*, *J. Geophys. Res.* **101**, 4115 (1996).
5. J. R. Petit *et al.*, *Nature* **399**, 429 (1999).
6. A. Neftel, H. Oeschger, T. Staffelbach, B. Stauffer, *Nature* **331**, 609 (1988).
7. T. Staffelbach, B. Stauffer, A. Sigg, H. Oeschger, *Tellus* **43B**, 91 (1991).
8. M. Anklin *et al.*, *J. Geophys. Res.* **102**, 26539 (1997).
9. O. Marchal *et al.*, *Clim. Dyn.* **15**, 341 (1999).
10. H. Fischer, M. Wahlen, J. Smith, D. Mastroianni, B. Deck, *Science* **283**, 1712 (1999).
11. M. Anklin, J.-M. Barnola, J. Schwander, B. Stauffer, D. Raynaud, *Tellus* **47B**, 461 (1995).
12. H. J. Smith, M. Wahlen, D. Mastroianni, K. C. Taylor, *Geophys. Res. Lett.* **24**, 1 (1997).
13. B. Stauffer *et al.*, *Nature* **392**, 59 (1998).
14. A. Indermühle *et al.*, *Nature* **398**, 121 (1999).
15. J. Tschumi, B. Stauffer, *J. Glaciol.* **46**, 45 (2000).
16. A. Indermühle, E. Monnin, B. Stauffer, T. F. Stocker, M. Wahlen, *Geophys. Res. Lett.* **27**, 735 (2000).
17. Supplementary information concerning procedures or assumptions is available at www.sciencemag.org/cgi/content/full/291/5501/112/DC1.
18. J. Schwander *et al.*, in preparation.
19. Throughout the remainder of the report, the precision of the ages is given in decimals of ky B.P. in order to facilitate the identification in Fig. 1. Nevertheless, the uncertainty in the absolute time scale remains as described in the text.
20. The available $CO_2$ data also suggest a small jump at the transition between intervals I and II, but this must be confirmed by further measurements.
21. H. J. Smith, M. Wahlen, D. Mastroianni, K. Taylor, P. Mayewski, *J. Geophys. Res.* **102**, 26577 (1997).
22. R. Röthlisberger, M. A. Hutterli, S. Sommer, E. W. Wolff, R. Mulvaney, *J. Geophys. Res.* **105**, 20565 (2000).
23. J. Jouzel *et al.*, in preparation.
24. W. Dansgaard *et al.*, *Nature* **364**, 218 (1993).
25. T. Sowers, M. Bender, *Science* **269**, 210 (1995).
26. J. Jouzel *et al.*, *Clim. Dyn.* **11**, 151 (1995).
27. T. Blunier *et al.*, *Geophys. Res. Lett.* **24**, 2683 (1997).
28. T. Blunier *et al.*, *Nature* **394**, 739 (1998).
29. J. Chappellaz *et al.*, *J. Geophys. Res.* **102**, 15987 (1997).
30. W. S. Broecker, G. H. Denton, *Geochim. Cosmochim. Acta* **53**, 2465 (1989).
31. We thank O. Marchal, F. Joos, J. Schwander, J. Chappellaz, and N. Shackleton for helpful comments. This work is a contribution to EPICA, a joint European Science Foundation/European Commission (EC) scientific program, funded by the EC under the Environment and Climate Programme (1994–98) contract ENV4-CT95-0074 and by national contributions from Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland, and the United Kingdom. This is EPICA publication no. 23. The measurements were supported by the Swiss NSF, the University of Bern, and the "Bundesamt für Energie."

# Evolution of Universal Grammar

### Martin A. Nowak,[1]* Natalia L. Komarova,[1,2] Partha Niyogi[3]

Universal grammar specifies the mechanism of language acquisition. It determines the range of grammatical hypothesis that children entertain during language learning and the procedure they use for evaluating input sentences. How universal grammar arose is a major challenge for evolutionary biology. We present a mathematical framework for the evolutionary dynamics of grammar learning. The central result is a coherence threshold, which specifies the condition for a universal grammar to induce coherent communication within a population. We study selection of grammars within the same universal grammar and competition between different universal grammars. We calculate the condition under which natural selection favors the emergence of rule-based, generative grammars that underlie complex language.

Language consists of words and rules. The finite ensemble of memorized words is called the mental lexicon, whereas the set of rules is called the mental grammar of a person (*1, 2*). Grammar is the computational system (*3*) that is essential for creating the infinite expressibility of human language. Children acquire

their mental grammar spontaneously and without formal training. Children of the same speech community reliably learn the same grammar. Exactly how the mental grammar comes into a child's mind is a puzzle. Children have to deduce the rules of their native language from sample sentences they receive from their parents and others. This information is insufficient for uniquely determining the underlying grammatical principles (*4*). Linguists call this phenomenon the "poverty of stimulus" (*5*) or the "paradox of language acquisition" (*6*). The proposed solution is universal grammar (*7*).

Universal grammar consists of (i) a mech-

[1]Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA. [2]Department of Mathematics, University of Leeds, Leeds LS2 9JT, UK. [3]Department of Computer Science, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: nowak@ias.edu

anism to generate a search space for all candidate mental grammars and (ii) a learning procedure that specifies how to evaluate the sample sentences ($8$–$13$). Universal grammar is not learned but is required for language learning. It is innate ($14$–$16$). Most linguists claim that the acquisition of language is done by specific neuronal circuitry within the brain and not by the general purpose problem-solving ability of the brain ($17$). Therefore, it should be possible to observe genetic defects of language acquisition ($18$).

The emergence of language is a major transition in evolution ($19$). Language permits efficient transfer of nongenetic information between individuals and thus leads to a new mode of evolutionary change. We do not attempt to explain detailed linguistic properties of human language but general aspects that are relevant for its evolution and its distinction from animal communication ($20$–$26$).

A key issue in cognitive science and linguistics is the question of how much is innate and how much is learnt in human language. Most attempts to quantify the complexity of universal grammar come from the concept of learnability ($27$, $28$); the object of study is the ideal speaker-hearer pair in a homogeneous linguistic community. Here, we consider a heterogeneous population and its evolutionary dynamics. We introduce the notion of grammatical coherence and quantify the condition for a universal grammar to induce coherent communication in a population. Thus, evolutionary considerations in addition to the traditional learning-theoretic ones are brought into play in the discourse on the maximum possible complexity of universal grammar.

Consider a universal grammar, $U$, which generates a search space that consists of $n$ candidate grammars, $G_1, \ldots, G_n$. Each grammar, $G_i$, is a rule system that defines a set of valid sentences. Denote by $a_{ij}$ the probability that a speaker who uses grammar $G_i$ formulates a sentence that is compatible with grammar $G_j$. The matrix $A = [a_{ij}]$ describes the pairwise relation among the $n$ grammars. We have $0 \leq a_{ij} \leq 1$ and $a_{ii} = 1$.

We assume there is a reward for mutual understanding. The payoff for an individual using $G_i$ communicating with an individual using $G_j$ is given by $F(G_i, G_j) = (1/2)(a_{ij} + a_{ji})$. This is the average probability that $G_i$ generates a sentence that is parsed by $G_j$ and vice versa. Note that $F(G_i, G_i) = 1$. Hence, in our first model, all $n$ grammars are equally powerful and allow the same level of communication ($29$).

We denote by $x_i$ the frequency of individuals who use grammar $G_i$. The average payoff of each of these individuals is given by $f_i = \Sigma_j x_j F(G_i, G_j)$. We assume that payoff translates into reproductive success: Individuals with a higher payoff produce more off-

spring. Children learn the language of their parents, and this learning process can be subject to mistakes. Denote by $Q_{ij}$ the probability that a child learning from a parent with grammar $G_i$ will end up speaking grammar $G_j$. With these assumptions, the population dynamics are given by
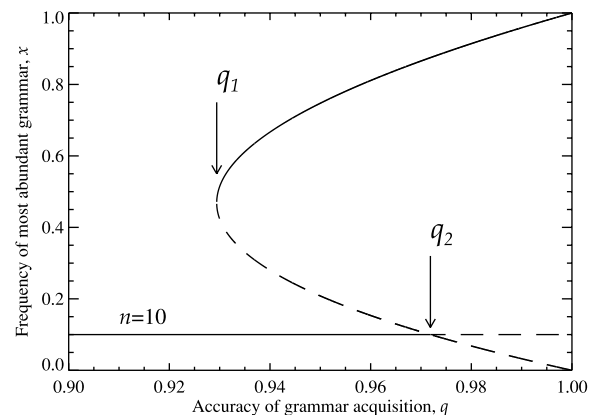
$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi x_i \qquad i = 1, \ldots, n \quad (1)$$

Here $\phi = \Sigma_i x_i f_i$ is the average fitness or grammatical coherence of the population; it is the probability that a sentence said by one person is understood by another person. The total population size is constant; we have $\Sigma_i x_i = 1$ ($30$).

In general, Eq. 1 can have multiple stable or unstable equilibrium solutions. For learning without mistakes, $n$ asymmetric equilibrium solutions of the form $x_i = 1$ and $x_j = 0$ (for all $j \neq i$) exist and are stable. Such solutions correspond to situations where all individuals of a population have adopted the same grammar. In contrast, for high error rates, the only stable equilibrium solution is one in which all grammars occur at roughly similar frequencies. We want to analyze the following question: How accurate does the learning process have to be for most individuals of the population to use the same grammar? In other words, when does a universal grammar induce coherent grammatical communication?

Consider the special case where all grammars have the same distance from each other, hence $a_{ij} = a$ for all $i \neq j$, where $a$ is a number between 0 and 1. In accordance, we have $Q_{ii} = q$ and $Q_{ij} = (1 - q)/(n - 1)$, where $q$ is the probability of learning the correct grammar or the accuracy of grammar acquisition. The symmetric solution, $x_i = 1/n$ for all $i$, always exists. Asymmetric equilibrium solutions with dominant grammar $G_i$ are given by $x_i = X$ and $x_j = (1 - X)/(n - 1)$,

$j \neq i$. Here $X$ is a number between 0 and 1. It is possible to show that such asymmetric solutions exist and are stable provided that $q$ exceeds a threshold value given by $q_1 = 2\sqrt{a}/(1 + \sqrt{a})$. The symmetric solution loses its stability when $q$ exceeds the threshold $q_2 = 1 - (1 - a)/(na)$. These results hold for $n \gg 1/a$. (If instead $1/a > n \gg 1$, then $q_1 = 2/\sqrt{n}$ and $q_2 = 1/2$.) Therefore, if $q < q_1$, only the symmetric solution is stable. If $q_1 < q < q_2$, then both the symmetric and the asymmetric solutions are stable; which one will be adopted depends on the initial conditions. Finally, for $q > q_2$, only the asymmetric solutions are stable. Hence, $q > q_1$ is a necessary condition for the population to converge to a coherent grammar, whereas $q > q_2$ is a sufficient condition (Fig. 1).

These conditions specify a "coherence threshold" for universal grammar. In general, $q$ will be a declining function of $n$. Therefore, $q(n) > q_1$ is an implicit condition for the maximum size of the search space generated by universal grammar. The coherence threshold is a necessary condition for evolution of complex language: Only a universal grammar that satisfies the coherence threshold can lead to the emergence of grammatical communication.

Let us now calculate the coherence threshold for two specific learning procedures that determine how to evaluate the input sentences. We will consider a "memoryless learner" and a "batch learner." Roughly speaking, a memoryless learner is the simplest and a batch learner the most sophisticated mechanism within a range of reasonable possibilities. Therefore, the actual learning mechanism used by children will have a performance between these two bounds.

The memoryless learner 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 sen-



**Fig. 1.** Bifurcation diagram showing the frequency of the most abundant grammar in a population versus the probability, $q$, that children correctly acquire the grammar of their parents. All $n$ grammars have the same distance from each other. We have $a_{ij} = a$ for $i \neq j$ and $a_{ii} = 1$. The $n$ asymmetric solutions are given by $x_i = X$ and $x_j = (1 - X)/(n - 1)$ for a specific $i$ and all $j \neq i$. For $n \gg 1$, we have $X = (q/2)(1 \pm \sqrt{D})$, where $D = 1 - 4[(1 - q)/q^2][a/(1 - a)]$. The equilibrium frequency of the predominating grammar does not depend on $n$. In the limit where $q$ approaches 1, the frequency of the most common grammar is simply given by $q$ and grammatical coherence is given by $q^2$. The asymmetric solutions exist for $q > q_1$. Each solution has a stable and an unstable branch. The symmetric solution, $x_i = 1/n$, loses stability for $q > q_2$. Parameter values are $a = 1/2$ and $n = 10$. Dashed lines correspond to unstable solutions.

tences consistent with $G_k$. As long as these sentences are also consistent with $G_i$, the learner maintains his hypothesis. If a sentence occurs that is not 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. For the special case where $a_{ij} = a$ for all $i \neq j$ and $n \gg 1$, the probability of learning the teacher's grammar is $q = 1 - [1 - (1 - a)/n]^b$. The threshold values, $q_1$ and $q_2$, can be restated in terms of the minimum number of sampling sentences per individual required for the population to converge to a coherent grammar. From $q > q_1$, we obtain

$$b > C_1 n \qquad (2)$$

where $C_1 = [1/(1 - a)] \log[(1 + \sqrt{a})/(1 - \sqrt{a})]$. Hence, for a memoryless learner, the

number of sample sentences has to exceed a constant times the number of candidate grammars.

The memoryless learner makes the minimum demand on the cognitive ability of the individual. The other extreme is a "batch learner" who memorizes $b$ sentences and then chooses the grammar that is most consistent with all memorized sentences. For the batch learner, we can show that the probability of learning the correct grammar, in a generic case, is given by $q = [1 - (1 - a^b)^n]/[na^b]$. Together with $q > q_1$, this leads to

$$b > C_2 \log n \qquad (3)$$

where $C_2 = 1/\log(1/a)$. Hence, a batch learner requires that the number of sample sentences exceeds a constant times the logarithm of the number of candidate grammars.

Because any realistic learning procedure has a performance somewhere between memoryless learners and batch learners, Eqs. 2 and 3

provide boundaries for the maximum size of the search space that is compatible with grammatical coherence within a population (*31*).

Now consider the situation where the candidate grammars, $G_1, \ldots, G_n$, have different distances from each other. Figure 2 shows the equilibrium solutions for a case with $n = 20$ candidate grammars, where the numbers $a_{ij}$ are randomly chosen from a uniform distribution on (0,1). For a small number of sample sentences, $b$, all grammars occur roughly at the same frequency and the grammatical coherence of the population is low. As $b$ increases, equilibrium solutions become stable, where the majority of the population uses a particular grammar. The critical transition occurs at a $b$ value that is approximately given by Eq. 2 with $a = 1/2$. It can be shown that there are exactly $n$ stable, one-grammar solutions if $b$ is large enough and $a_{ij} < 1$ for all $i \neq j$.

The candidate grammars could differ in their overall performance. Some grammars could describe a larger number of concepts or be less ambiguous than others. Hence, candidate grammars can have different fitness values (*32*). In such a scenario, the one-grammar solutions assume different fitness values even for large $b$. Therefore, we can imagine an evolutionary process where the population is searching for fitter candidate grammars. Suppose a population uses a particular grammar, $G_1$. Someone invents a modification that alters the grammar to $G_2$. A fluctuation could shift the whole population to adopt $G_2$. Such transitions are more likely to occur in a small population. They are favored if the two grammars are fairly similar and $G_2$ has a higher fitness than $G_1$. Hence, the model provides a framework for studying the cultural, evolutionary adaptation of grammar within the same universal grammar (*33*).

To further illuminate the selective pressures that act on the design of universal grammar, we study the competition between different universal grammars (*34*). We state two specific results.

First, consider universal grammars with the same search space and the same learning procedure, the only difference being the number of input sentences, $b$. This quantity is proportional to the length of the learning period. We find that natural selection leads to intermediate values of $b$ (Fig. 3). For small $b$, the accuracy of learning the correct grammar is too low. For large $b$, the learning process takes too long and is too costly. This observation can explain why there is a limited language acquisition period in humans.

Second, consider universal grammars that differ in the size of their search space, $n$, but have the same learning mechanism and the same value of $b$. In general, there is selection pressure to reduce $n$. Only if $n$ is below the coherence threshold can the universal grammar

**Fig. 2.** Grammatical coherence of a population versus the number of sample sentences, $b$, per individual for stable equilibrium solutions of Eq. 1. There are $n = 20$ grammars with randomly chosen pairwise distances; the $a_{ij}$ values (for $i \neq j$) are taken from a uniform distribution on (0,1), and $a_{ii} = 1$. Children learn the grammar of their parents according to a memoryless learning algorithm. The grammatical coherence (or average fitness) of the population is given by $\phi = \Sigma_i x_i f_i$, where $f_i = (1/2)\Sigma_j x_j(a_{ij} + a_{ii})$. It is a measure of mutual understanding in the population. For small $b$, all grammars occur at roughly similar frequency; the coherence is low. For larger values of $b$, stable equilibria appear with the majority of the population adopting the same grammar. The first critical transition occurs roughly at $b = 3.5n$, which is Eq. 2 with $a = 1/2$. Some grammars lead to stable equilibrium solutions only for large numbers of sample sentences. In the limit $b \to \infty$, there are $n$ stable equilibria corresponding to all people using one of the $n$ grammars.
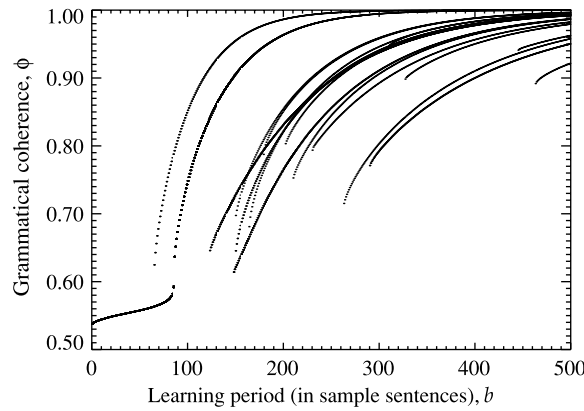
**Fig. 3.** Natural selection chooses a limited period of grammar acquisition. The time it takes to learn grammar is proportional to the number of sample sentences, $b$, that are being evaluated. The evolutionarily stable value of $b$ maximizes the product $r(b)q(b)$ (solid line), which represents the rate of producing offspring that have acquired the correct grammar (same grammar as the parent). Here $r(b)$ is the rate of producing offspring that have acquired grammatical communication, and $q(b)$ is the probability of learning the correct grammar. The selected value of $b$ is marked as ESS (evolutionarily stable strategy). The value $b_c$ represents the coherence threshold. For this figure, we chose $r(b) = 1/(1 + 0.01b)$ and $q(b) = 1 - [1 - (1 - a)/n]^b$ (as defined for the memoryless learner). The evolutionary stability analysis uses the two–universal grammar equation of (*34*) and is exact for large values of $n$. The evolutionarily stable strategy does not maximize the fitness of the population, which is given by the product $r(b)\phi$ (dashed line). Parameter values are $n = 10$, $a_{ii} = 1$, and $a_{ij} = 0.1$ for $i \neq j$.
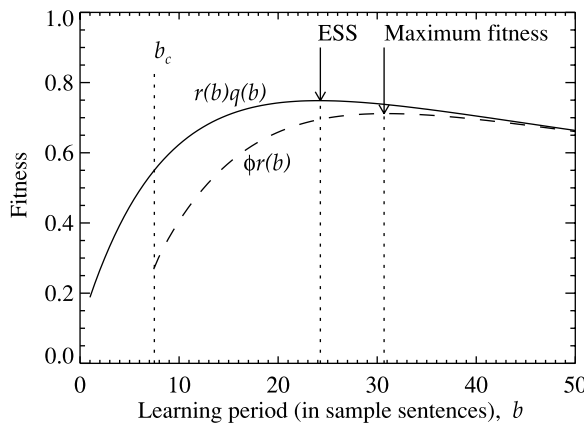
induce grammatical communication. Moreover, as $n$ declines, the accuracy of grammar acquisition increases (*35*). There can, however, also be selection for larger $n$: Suppose universal grammar $U_1$ is larger than $U_2$ (that is, $n_1 > n_2$). If all individuals use a grammar, $G_1$, that is both in $U_1$ and $U_2$, then $U_2$ is selected. Now imagine that someone invents a new advantageous grammatical concept that leads to a modified grammar $G_2$ that is in $U_1$, but not in $U_2$. In this case, the larger universal grammar is favored. Hence, there is selection both for reducing the size of the search space and for remaining open minded to be able to learn new concepts. For maximum flexibility, we expect search spaces to be as large as possible but still below the coherence threshold.

We now explore the conditions under which natural selection favors the emergence of a rule-based, recursive grammatical system with infinite expressibility. In contrast to such rule-based grammars, one might consider list-based grammars that consist only of a finite number of sentences. Such list-based grammars can be seen as very primitive evolutionary precursors (or alternatives) to rule-based grammars. Individuals would acquire their mental grammar not by searching for underlying rules, but by simply memorizing sentence types and their meaning (similar to memorizing the arbitrary meaning of words). List-based grammars do not allow for creativity on the level of syntax. Nevertheless, whether or not natural selection favors the more complicated rule-based grammars depends on circumstances that we need to explore.

Current human grammars can generate infinitely many sentence types, but for the purpose of transmitting information, only a finite number of them can be relevant. Natural selection cannot directly reward the theoretical ability to construct infinitely long sentences. Let us therefore consider a group of individuals that use $N$ different sentence types (or syntactic structures). $N$ specifies the number of sentence types that are relevant from the perspective of biological fitness.

Now imagine individuals that learn their mental grammar by memorizing lists of sentence types. We can ask how many sample sentences, $b$, a child must hear for the whole population to maintain $N$ sentence types. If all sentence types occur equally often, we simply obtain $b > N$ (*36*).

We can compare the performance of individuals using list-based versus rule-based grammars. Using the result for batch learners, which have comparable memory requirements to the list learners, we find that the number of relevant sentence types, $N$, has to exceed a constant times the logarithm of the number of candidate grammars, $n$. We have

$$N > \frac{\log n}{\log(1/a)} \qquad (4)$$

If this condition holds, then rule-based grammars are more efficient than list-based grammars and will have a fitness advantage (*37*).

We have formulated a mathematical theory for the population dynamics of grammar acquisition and calculated the conditions for universal grammar to induce coherent grammatical communication within a population. The key result is a "coherence threshold" that relates the maximum size of the search space to the performance of the learning procedure. Only a universal grammar that satisfies the coherence threshold can promote the evolution of grammatical communication. Our theory also describes the cultural evolution of different candidate grammars within the same universal grammar. We have studied competition between universal grammars and outlined that natural selection leads to a limited language learning period and search spaces of intermediate size. Finally, rule-based, generative grammars can only evolve if the number $N$ of sentences types, which are relevant with respect to biological fitness, exceeds the logarithm of the number of candidate grammars, $\log n$. Otherwise it would be more efficient to memorize sentence types associated with arbitrary meaning. In this case, language would have remained a rather dull communication system without any creative ability on the level of syntax. If, however, rule-based grammars are selected, then the potential for "making infinite use of finite means" (*38*) comes as a by-product.

The general connection between learning theory and evolutionary dynamics should be applicable to a wide variety of learning situations that arise in biology or artificial intelligence. For language acquisition, the theory makes testable predictions about the relation between the size of the search space and the performance of the learning procedure. For language evolution, this report is a step toward a quantitative understanding of how universal grammar can arise by natural selection.

**References and Notes**
1. S. Pinker, *Words and Rules* (Basic Books, New York, 1999).
2. G. A. Miller, *The Science of Words* (Scientific American Library, New York, 1996).
3. N. Chomsky, in *The View from Building 20*, K. Hale, S. J. Keyser, Eds. (MIT Press, Cambridge, MA, 1993), pp. 1–52.
4. E. M. Gold, *Inform. Control* **10**, 447 (1967).
5. N. R. Hornstein, D. W. Lightfoot, *Explanation in Linguistics* (Longman, London, 1981).
6. R. Jackendoff, *The Architecture of the Language Faculty* (MIT Press, Cambridge, MA, 1997).
7. N. Chomsky, *Rules and Representations* (Columbia Univ. Press, New York, 1980).
8. K. Wexler, P. Culicover, *Formal Principles of Language Acquisition* (MIT Press, Cambridge, MA, 1980).
9. E. Gibson, K. Wexler, *Ling. Inquiry* **25**, 407 (1994).
10. D. Lightfoot, *The Development of Language: Acquisition, Changes and Evolution* (Blackwell/Maryland Lecture in Language and Cognition, Oxford, 1999).
11. R. Manzini, K. Wexler, *Ling. Inquiry* **18**, 413 (1987).
12. P. Niyogi, *The Informational Complexity of Learning* (Kluwer Academic, Boston, 1998).
13. D. Osherson, M. Stob, S. Weinstein, *Systems That Learn* (MIT Press, Cambridge, MA, 1986).
14. Universal grammar and the innateness of grammatical principles of human language are controversial issues. Here, we use a very general definition of universal grammar denoting "mechanism of language acquisition," which is certainly innate. The continuity hypothesis states that universal grammar is available to the child at all stages of development (*15*), whereas the maturationist hypothesis holds that universal grammar is changing during language acquisition (*16*).
15. N. Hyams, *Language Acquisition and the Theory of Parameters* (Reidel, Dordrecht, Netherlands, 1986).
16. A. Radford, *Syntactic Theory and the Acquisition of English Syntax* (Blackwell, Oxford, 1990).
17. S. Pinker, *The Language Instinct* (Morrow, New York, 1994).
18. M. Gopnik, in *Language, Logic, and Concepts*, R. Jackendoff, P. Bloom, K. Wynn, Eds. (MIT Press, Cambridge, MA, 1999), pp. 263–284.
19. J. Maynard Smith, E. Szathmary, *The Major Transitions in Evolution* (Freeman Spektrum, Oxford, 1995).
20. R. Brandon, N. Hornstein, *Biol. Philos.* **1**, 169 (1986).
21. S. Pinker, A. Bloom, *Behav. Brain Sci.* **13**, 707 (1990).
22. F. Newmayer, *Lang. Commun.* **11**, 3 (1991).
23. T. Hashimoto, T. Ikegami, *Biosystems* **38**, 1 (1996).
24. M. D. Hauser, *The Evolution of Communication* (Harvard Univ. Press, Cambridge, MA, 1996).
25. J. R. Hurford, M. Studdert-Kennedy, C. Knight, Eds., *Approaches to the Evolution of Language* (Cambridge Univ. Press, Cambridge, 1998).
26. M. A. Nowak, D. C. Krakauer, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8028 (1999).
27. V. Vapnik, *The Nature of Statistical Learning Theory* (Springer, New York, 1995).
28. L. G. Valiant, *Commun. ACM* **27**, 436 (1984).
29. A grammar mediates a mapping between form and meaning. The countably infinite number of possible linguistic expressions can be represented as strings over a finite syntactic alphabet $\Sigma_1$. The set of all possible strings of $\Sigma_1$ is denoted by $\Sigma_1^*$. Similarly, one can enumerate all possible meanings, $\Sigma_2^*$ as strings over a primitive semantic alphabet $\Sigma_2$. Therefore, $\Sigma_1^*$ is the set of all possible linguistic expressions and $\Sigma_2^*$ is the set of all possible meanings. A grammar $G_i$ generates a subset of $\Sigma_1^* \times \Sigma_2^*$, that is, a (potentially infinite) set of sentence-meaning pairs. Mathematically, $G_i$ is specified by a measure $\mu_i$ on $\Sigma_1^* \times \Sigma_2^*$. We can define $a_{ij} = \mu_i(G_i \cap G_j)$ to be the proportion of sentence meaning pairs that $G_i$ and $G_j$ have in common. Hence, $a_{ij}$ is the probability that a user of $G_i$ speaks an utterance that a user of $G_j$ can understand.
30. Equation 1 is similar to the standard quasi-species equation but has frequency-dependent fitness functions. For quasi-species theory, see M. Eigen and P. Schuster [*The Hypercycle. A Principle of Natural Self-Organisation* (Springer, Berlin, 1979)].
31. This result can also be discussed in the principles and parameters framework. Universal grammar is determined by genetically inherited principles, which limit the number of candidate grammars and specify the learning mechanism. The parameters have to be learned by evaluating input sentences. We can calculate the maximum number of parameters consistent with the coherence threshold. Suppose there are $k$ independent parameters that can be represented as binary switches. Therefore, $n = 2^k$. For the memoryless learner, we obtain $k < \log_2(b/C_1)$. For the batch learner, we obtain $k < b/(C_2 \log 2)$. The innate principles have to reduce the number of parameters, $k$, to fulfill these conditions. A different approach is optimality theory. There are $k$ constraints. Each grammar is given by a specific ordering of these constraints. Hence, $n = k!$ [A. Prince, P. Smolensky, in *Technical Report RuCCS TR-2* (MIT Press, Cambridge, MA, 1993), pp. 234–272; B. B. Tesar, P. Smolensky, *Lingua* **106**, 161 (1998)].
32. Let us consider a generalization that allows us to define grammars of varying intrinsic fitness. Each user of $G_i$ is characterized by an encoding matrix $P$ and a decoding matrix $Q$. Here, $P_{kl} = \mu(s_k, m_l)/\Sigma_l \mu(s_l, m_l) = \mu(s_k, m_l)$, which is simply the probability of

using the expression $s_k$ to convey the meaning $m_l$. Similarly, $Q_{kl} = \mu(s_k, m_l)/\Sigma_j\mu(s_k, m_j) = \mu(m_l, s_k)$ is the probability of interpreting the expression $s_k$ to mean $m_l$. The need to communicate meanings is related to events in the shared world of the linguistic community. Therefore, one can define a measure $\sigma$ on the set of possible meanings $(\Sigma_2^*)$ that speakers and hearers might wish to communicate with each other. Given this, we can define $a_{ij} = tr[P^{(i)}\Lambda(Q^{(j)})^T]$, where $\Lambda$ is a diagonal matrix such that $\Lambda_{ii} = \sigma(m_i)$. This is the probability that an event occurs and is successfully communicated from a user of $G_i$ to a user of $G_j$. $F(G_i,G_i)$ is the probability that users of $G_i$ will have a successful communication with each other. Communication might break down in one of two ways: (i) poverty: an event happens whose meaning cannot be encoded by $G_i$, and (ii) ambiguity: an event happens whose meaning has an ambiguous encoding in $G_i$ leading to a possibility of misunderstanding. Thus, $F(G_i,G_i)$ is a number between 0 and 1 and denotes the fitness of $G_i$. Maximum fitness, $F(G_i,G_i) = 1$, is achieved by grammars that can express every possible meaning (zero poverty) and have no ambiguities.

33. To study the effect of finite (small) population sizes, the deterministic Eq. 1 is replaced by a stochastic process. In this case, we observe that the population adopts one of the candidate grammars (that admits a stable equilibrium) for some time and then jumps to another equilibrium. If the candidate grammars differ in their fitness, then the stochastic process performs an evolutionary optimization on the space of all grammars.

34. Denote by $x_i$ the fraction of individuals who use $G_i$ of universal grammar $U_1$; denote by $y_i$ the fraction of individuals who use $G_i$ of universal grammar $U_2$. $U_1$ and $U_2$ contain, respectively, $n_1$ and $n_2$ candidate grammars. Some of the candidate grammars can be part of both universal grammars. The universal grammars, $U_1$ and $U_2$, can also differ in the number of sample sentences, $b_1$ and $b_2$, that are being considered. Therefore, we have to take into account the rate of producing offspring with grammatical communication; this rate is given by the declining function $r(b)$. An alternative interpretation is that $r(b)$ describes the cost that is associated with learning. The dynamics are described by

$$\dot{x}_i = r(b_1)\sum_{j=1}^{n_1} x_j f_j^{(1)} Q_{ji}^{(1)} - \phi x_i \qquad i = 1, \ldots, n_1$$

$$\dot{y}_i = r(b_2)\sum_{j=1}^{n_2} y_j^{(2)} f_j^{(2)} Q_{ji}^{(2)} - \phi y_i \qquad i = 1, \ldots, n_2$$

We have $f_i^{(m)} = \Sigma_{j=1}^{n_1} x_j F(G_i^{(m)}, G_j^{(1)}) + \Sigma_{j=1}^{n_2} y_j F(G_i^{(m)}, G_j^{(2)})$, $m \in \{1, 2\}$, and $\phi = \Sigma_{i=1}^{n_1} f_i^{(1)} x_i r(b_1) + \Sigma_{i=1}^{n_2} f_i^{(2)} y_i r(b_2)$, where the superscripts 1 and 2 refer to $U_1$ and $U_2$, respectively.

35. In general, it is advantageous to reduce the size of the search space, because a smaller $n$ leads to a larger accuracy of grammar acquisition. The situation is more complex, however. Consider two universal grammars $U_1$ and $U_2$ with $n_1 > n_2$. Suppose $U_1$ is resident and $U_2$ is an invading mutant. If $n_1$ exceeds the coherence threshold, then $U_2$ will always outcompete $U_1$. If $n_1$ is below the coherence threshold, then $U_2$ can only invade if the specific grammar adopted by the population of $U_1$ speakers is also part of $U_2$; otherwise $U_1$ can resist invasion by $U_2$. The selective difference between $U_1$ and $U_2$ is small if both $n_1$ and $n_2$ values are either well above or well below the coherence threshold. Hence, selection is strongest close to the coherence threshold (if $n_1 \approx n_2$).

36. This problem has been solved before, in a different context. How many words, $N$, can be stably maintained in a population if each child hears $b$ words during its language acquisition period and has a probability, $\rho$, to memorize a new word after one encounter? The answer is $N < b\rho$ [M. A. Nowak, J. B. Plotkin, V. A. A. Jansen, *Nature* **404**, 495 (2000)].

37. The implicit assumption here is, of course, that the rule-based grammars can generate at least these $N$ sentence types. In a principles and parameters framework, condition 4 implies that the number of parameters $k$ has to be less than $N$.

38. W. von Humboldt, *Linguistic Variability and Intellectual Development* (Univ. of Pennsylvania Press, Philadelphia, 1972).

39. 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 NSF is gratefully acknowledged.

21 September 2000; accepted 10 November 2000

# Broad-Spectrum Mildew Resistance in *Arabidopsis thaliana* Mediated by *RPW8*

Shunyuan Xiao, Simon Ellwood, Ozer Calis, Elaine Patrick, Tianxian Li, Mark Coleman, John G. Turner*

Plant disease resistance (*R*) genes control the recognition of specific pathogens and activate subsequent defense responses. We show that the *Arabidopsis thaliana* locus *RESISTANCE TO POWDERY MILDEW8* (*RPW8*) contains two naturally polymorphic, dominant *R* genes, *RPW8.1* and *RPW8.2*, which individually control resistance to a broad range of powdery mildew pathogens. Although the predicted RPW8.1 and RPW8.2 proteins are different from the previously characterized *R* proteins, they induce localized, salicylic acid–dependent defenses similar to those induced by *R* genes that control specific resistance. Apparently, broad-spectrum resistance mediated by *RPW8* uses the same mechanisms as specific resistance.

The majority of characterized *R* genes participate in gene-for-gene interactions, in which the *R* product appears to act as a receptor that recognizes a product of the corresponding avirulence (*Avr*) gene from the pathogen, inducing defense responses. The *R* gene–mediated defenses typically involve a rapid, localized necrosis, or hypersensitive response (HR), at the site of infection, and the localized formation of antimicrobial chemicals and proteins that restrict growth of the pathogen (*1, 2*). Many crops rely on *R* genes for resistance to specific pathogens, but resistance fails in the presence of strains of the pathogen that lack the corresponding *Avr* genes. Broad-spectrum disease resistance is therefore desirable, and has been achieved through the use of recessive mutations (*3*); a challenge is to develop broad-spectrum resistance with dominant *R* genes (*4*). More than 20 of the *R* genes that confer specific resistance have been characterized and they form five classes of protein with differing combinations of five conserved structural motifs. With the exception of Hm1, a toxin reductase, and Pto, a protein kinase, the characterized *R* proteins contain a leucine-rich repeat (LRR) motif believed to specify recognition (*1, 5, 6*). All of the characterized *A. thaliana R* genes encode proteins with motifs for a nucleotide-binding site (NBS) and an LRR, and similar *R* genes have been isolated from

other plant species (*1*). Probes for conserved sequences in NBS-LRR motifs have detected numerous homologs in the genomes of crop plants (*7*), and more than 100 in the genome of *A. thaliana* (*8*). The *A. thaliana* loci *RPW7* and *RPW8* from accession Ms-0 map to the same interval on chromosome 3, and confer resistance to the powdery mildew pathogens *Erysiphe cruciferarum* UEA1 and *E. cichoracearum* UCSC1, respectively (*9*). During the mapping of *RPW8*, we discovered that this locus segregated from the mapped NBS-LRR *R*-gene homologs (*8*), suggesting the presence of a different type of resistance gene, which we characterize here.

We genetically mapped *RPW8* to a fragment of genomic DNA from Ms-0 in cosmid B6 (Fig. 1A) (*10*), and confirmed that Col-0 plants containing the B6 transgene (T-B6) were resistant to *E. cichoracearum* UCSC1 (Fig. 2A) (*10*). The B6 DNA sequence (*11*) revealed three open reading frames (ORFs) (Fig. 1A) encoding a serine/threonine protein kinase 2 (*SPK-2*) (GenBank accession number S56718) and two uncharacterized genes, which we named *MSC1* and *MSC2*. Subclones of B6 in a plant transformation vector were introduced into Col-0 plants by *Agrobacterium*-mediated transformation (*11, 12*), and those that contained either *MSC1* or *MSC2*, or both of these ORFs, conferred resistance to *E. cichoracearum* UCSC1 (Fig. 1A). This indicated that *RPW8* comprises two independently acting genes, *MSC1* and *MSC2*, which we therefore renamed *RPW8.1* and *RPW8.2*, respectively. Ms-0 cDNAs for *RPW8.1* and *RPW8.2* (*13*) were introduced

School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

*To whom correspondence should be addressed. E-mail: j.g.turner@uea.ac.uk