Does complexity always increase during major evolutionary transitions?

Mercedes Bleda-Maza de Lizana[1], Hannelore Brandt[2], Ille C. Gebeshuber[3], Michael MacPherson[4], Tetsuya Matsuguchi[5] and Szabolcs Számadó[6]

[1] Manchester School of Management & Tyndall Centre for Climate Change Research, Manchester, UK

[2] Department of Mathematics, University of Vienna, Austria

[3] Institut für Allgemeine Physik, Vienna University of Technology, Vienna, Austria & Austrian Center of Competence for Tribology AC^2T research GmbH, Wiener Neustadt, Austria

[4] Theoretical Biology Laboratory, Stanford University, CA, USA

[5] Department of Biology, MIT, Cambridge, MA, USA

[6] Department of Plant Taxonomy and Ecology, Eötvös Loránd University, Budapest

Introduction

There is a long raging debate about the evolution of biological complexity. Despite the common sense observation that evolution produces more and more complex organisms, there is no accepted argument in favor of increasing complexity during biological evolution. In fact, many notable evolutionary biologists have explicitly warned against drawing such conclusion. For example, Maynard Smith and Szathmáry (1995) conclude that the increase of fitness cannot be equated with the increase of complexity, and not even the increase of fitness can be taken for granted during evolution. Others argue that the recognizable patterns can be explained by a ``random walk" which started from the simplest forms, thus complexity is expected to increase not because of some underlying trends but because of the boundary conditions of the process (Gould, 1996). On the other hand, Adami (2000) and his colleagues were able to show, with the help of evolving artificial life, that physical complexity must always increase in fixed environments for organisms whose fitness depends only on the sequence of their genome. The obvious shortcomings of this approach are that the environment is rarely fixed and that the fitnesses of organisms might depend on many factors other than their respective genomic sequences. Yet observing the increase of structural complexity of living creatures can make another line of argument in favor of increasing complexity. The major steps of this process sometimes called the "major transitions of evolution" (Maynard Smith and Szathmáry, 1995). Structural complexity increases during these transitions by definition, but what happens with physical complexity? Clearly the physical complexity of those parts of the genome that code for the structural changes should increase, but what happens with the other regions? Can we expect an increase of complexity or not? To study the question we developed a computer simulation of a generalized transition event. The properties of the transitions were taken from the book of Maynard Smith and Szathámry (1995).

Physical Complexity

Whereas the opinions on a general trend in evolution towards increasing complexity vary, increasing complexity during major evolutionary transitions may be obvious, as the result of a major transition might be exactly what we would call a more complex biological system. However, in order to decide on the question whether or not complexity always increases during major transitions in evolution, we

first have to make clear what complexity should denote precisely when referring to biological systems.

Although there is some common understanding of the meaning of complexity referring to biological organisms, to determine exactly what this notion stands for is not evident (Adami 2002). While several measures of complexity abound, not all are applicable for biology. The complexity of dynamical systems refers mostly to the features of processes, but for biological organisms complexity is rather related to form, function, or the underlying sequences. The number of different cell types, or the number of functions an organism can perform captures to some extent its complexity, but a general measure of structural or functional complexity is hard to define, as organisms differ in so many aspects.

However, it is possible to compare nucleic acid sequences, and even though the genotype-to-phenotype mapping is of great intricacy, the sequences should at least to some extent reflect the complexity of the corresponding organism.

Physical complexity, as stated in Adami (2002) is a measure of sequence complexity. This approach turns out to be very intuitive. It seems to correspond exactly to what biologists think is increasing when systems get self-organized. Furthermore, Adami argues that physical complexity must increase in molecular evolution under certain circumstances, due to natural selection.

An important aspect of information theory is that information is physical and must be about something. An arrangement of symbols becomes information only in some context to other physical objects. In biological systems, those parts of the genome which actually correspond to some function, in fact also store information about the environment and co-evolving species. The physical complexity of a sequence is meant to reflect the amount of information stored about the environment. Physical complexity is a relative measure, as it depends on the environment. It is possible to show that physical complexity must increase due to natural selection, if the environment does not change and the fitness of organisms only depends on their own sequence information. On the other hand, a drastically changing or fluctuating environment can lead to a decrease of physical complexity.

For a calculation of the physical complexity, we proceed according to Adami (2002). The entropy of an ensemble of sequences X, in which sequences s_i occur with probabilities p_i is calculated as $H(X) = -\sum_i p_i \log p_i$, where the sum goes over all different genotypes i in X. By setting the base of the logarithm to the size of the monomer alphabet, the maximal entropy (in the absence of selection) is given by $H_{max}(X) = L$, and in fact corresponds to the maximal information that can potentially be stored in a sequence of length L.

The amount of information a population X stores about the environment E is now given by:

$$I(X:E) = H_{\max} - (X \mid E) = L + \sum_i p_i \log p_i$$

The entropy of an ensemble of sequences is estimated by summing up the entropy at every site along the sequence. The per-site entropy is given by:

$$H(j) = -\sum_{i=G,C,A,T} p_i(j) \log p_i(j)$$

for site j, where $p_i(j)$ denotes the probability to find nucleotides i at position j. The entropy is now

approximated by the sum of per-site entropies;

$$H(X) \approx \sum_{j=1}^{L} H(j),$$

so that an approximation for the physical complexity of a population of sequences with length L is obtained by

$$C(X) = L - H(X)$$

A Simple Model

We consider a simple model of a general transition in organization. It consists of a square toroidal lattice on which a number of *individuals* are placed and an *environment* consisting of k generic tasks. At the outset, each individual is assigned a *genome* of $k+2$ values drawn from the uniform distribution on [0, 1]. The first k genome entries represent the aptitudes of that individual in the k tasks; the last two entries are the *cooperativity* and *staying threshold*, to be discussed shortly.

At first the individuals move about the lattice in unbiased two-dimensional random walks, and without interacting with one another. Each seeks to maximize its fitness subject to the constraint that a fixed minimum level of fitness τ must be attained in each task, by dividing one *allocation* unit at each timestep. Thus, when g_{ij} and a_{ij} are the genome value and allocation fraction of the ith individual in the jth task and e_j is the value of the environment in task j, the fitness of individual i is:

$$F_i = \sum_{j=1}^{k} g_{ij} a_{ij} e_j \;,\; g_{ij} a_{ij} e_j \geq \tau \;,\; \sum_{j=1}^{k} a_{ij} = 1$$

The optimal strategy for an individual is to meet the threshold exactly in all tasks, and devote the remaining allocation fraction to that task j for which the product $g_{ij}e_j$ is greatest. The consequence to an individual which cannot meet the threshold in all tasks is simply that its fitness will not exceed $k\tau$.

To represent selection, at each timestep each individual has a fixed probability of replacement. The replacement is a near-copy of another individual in the population, where the jth value in the new genome is given as $g'_{ij} = g_{ij} + M$, where M is a uniform random variable on the interval [-μ, μ] and μ is the *mutation rate*. The individual serving as the basis for the replacement is chosen at random from the population with a probability proportional to its present fitness.

The organizational transition occurs when the individuals are permitted to form *collectives*. If two or more individuals find themselves on the same lattice point, they unite to form a collective. The collective computes fitness similarly to individuals: the differences are that the threshold is increased from τ to $n\tau$ where n is the number of individuals in the collective and that the allocation is not one but the sum of the cooperativities of each of the individuals, multiplied by a *benefit factor*. The collective can use the resources of any of the individuals in any of the tasks. After the collective's fitness is computed, it is divided equally among the individuals regardless of their contribution. Then each individual computes its fitness using the fraction of its allocation which it did not devote to the collective, *i.e.* one minus its cooperativity.

Finally, individuals decide at each timestep whether to remain in the collective by comparing $\exp(-(\log 2)\, f_C / f_I)$ to its staying threshold, where f_C is the fitness gotten from the collective and f_I that gotten alone, and staying put if the latter exceeds the former.

We measure the physical complexity as described above, with a small adjustment to our model. Physical complexity is computed from a set of fixed-length strings whose symbols come from a finite alphabet. In this model the genomic aptitudes are numbers between zero and one, so we convert them to bit strings before computing the complexity. To do this we chose a field width *c* and divided the interval [0, 1] into 2^c equally-sized intervals labeled with consecutive binary strings of length *c*. We associated to each genomic aptitude the string of the interval in which it lay. Thus, with field width 3 the genomic aptitude would get the string 111. These strings for each genomic aptitude were concatenated to form a single bit string for each individual.

Results and Discussions

We implemented the model on a computer and ran it with *50* individuals on a *10×10* grid with parameter settings as described below. The results of a typical run appear in (Figure 1-3). The mean genomic aptitudes tend towards one as individuals are replaced with fitter individuals. No trend is apparent in the mean cooperativity and staying threshold. Collectives come into and go out of existence, while the mean fitness rises to plateaus whose heights vary from run to run.

The model always results in an increase in both structural and physical complexity. The former is true by definition since a collective is structurally more complex than the individuals of which it is composed; whenever collectives are permitted, the structural complexity increases. The physical complexity increases as a result of the fitness scheme and replacement rule, independently of whether collectives can form or not. An organism with genomic aptitudes all equal to one has greater fitness in this model than an organism without, which results in populations with genomic aptitudes all close to one after enough generations have passed. In such a population the physical complexity, effectively a measure of homogeneity, must increase.

We computed several statistics about each collective during a run, including its mean size, mean values of the fitness and genomic values of its constituents, mean number of freeriders (members with cooperativity less than 1/2), and the mean of the variance of its genomic aptitudes. The last statistic measures the extent of division of labor within a collective; it is large when members specialize in particular tasks and small when they do not. We also computed the mean lifetime over all the collectives during a simulation, reasoning that conditions favoring collectives would result in longer lifetimes.

We found that collective stability was strongly related to both the birth-death rate and the benefit factor. In Table 1, each value is the mean collective lifetime averaged over ten simulations under some combination of birth-death rate and benefit factor. Increasing the birth-death rate undermines collective stability. When replacement occurs within an existing stable collective, the new individual might be uncooperative or require a high return from the collective where the old was not. It could also be that

the new individual is better for the collective than the old, but since uncooperative and/or demanding individuals tend not to be in collectives in the first place, high birth-death tends to break up collectives.

Increasing the benefit factor tends to increase collective stability. This is because the fitness obtained from the collective increases, increasing the likelihood that the collective fitness to individual fitness ratio will be above each member's threshold. This effect is extremely important to collective stability; when the benefit factor drops to 1, a mean lifetime *1.4* indicates that most collectives disband a single timestep after forming.

We found a number of strong correlations between other collective statistics, averaged over ten simulations with a particular combination of birth-death rate and benefit factor. As one expects, there were strong positive correlations between the size, genomic aptitudes, and fitness within a collective. In Table 2, we show the other strong correlations between collective statistics. The pairs are not surprising: fitness correlates positively with cooperativity, mean lifetime, and staying threshold, in decreasing order of correlation strength, and negatively with mean number of freeriders. The mean lifetime correlates strongly with collective size.

Changing the birth-death rate doesn't affect the magnitude of any of the correlations very much. It appears that increasing the benefit factor strengthens the correlation between fitness & mean lifetime and mean lifetime & size, and decreases that between fitness and staying threshold. The correlations are probably insensitive to the birth-death rate because it only affects collectives externally. Increasing the benefit factor, however, does affect the dynamics within a collective, evidently making it more stable. The correlation trends make sense; longer-lived, stable collectives have the chance to grow larger and accrue fitness, and will be able to support more members with lower staying thresholds.

Some correlations we had expected to be strong were curiously weak under this model. Division of labor as measured by the mean variance of genomic aptitudes correlated with nothing strongly. Mean lifetime correlated strongly with fitness and size but not cooperativity or the number of freeriders. It could well be that division of labor doesn't occur because the replacement rule tends to generate like individuals, each of whom is about as good as any other in a given task. As for the mean lifetime, one way to account for its independence from cooperativity is if collectives are incapable of selecting against noncooperative individuals. Then collectives would tend to have as many freeriders as contributors and thus no correlation between their lifetimes and their composition. In this model, the only way for collectives to rid themselves of freeriders is if the freeriders also happen to have lower fitness than contributors. From the table, it appears that they do, so one would expect a negative correlation to develop between mean lifetime and number of freeriders if the simulations were to be run longer than 10000 steps.

Conclusions

The physical complexity of those regions that do not code for the structural changes increases independently from structural complexity. There is no correlation between fitness of collectives and division of labor (possible causes will be discussed later). On the other hand, there is a correlation

between the size of colonies and fitness. Thus, it seems that there is no additional increase of physical complexity other than those regions that code for the structural changes.

Of course, the results should be viewed in the light of the assumptions of the model. Some of these assumptions are deliberate simplifications, elaborating them in detail might provide ground for further research. One of these assumptions is the fixed nature of the environment; the results would almost certainly change if we allowed the environment to vary. Introducing fluctuations in the environment (which would alter the effectiveness of different genotypes) might trigger a greater diversity of collectives and might favor further division of labor. Second, we exogenously control the transition, that is, it is a built-in feature whether or not individuals are allowed to form collectives. Arguably, it would be much more interesting to produce an evolutionary model in which a transition occurs endogenously. However, it was beyond the scope of our current investigation. Third, as long as an organism has allocation left over it can continue to improve its fitness at the same rate. That is, there is no diminishing return when the organism decides to invest all of its remaining resources into the same trait. Introducing such diminishing returns might favor greater diversity and hence greater division of labor. Fourth, it seems that it is too easy to achieve the optimal genomic aptitude; as a result the whole population becomes homogenous with respect of the task genes. In turn this might be one reason why we cannot observe division of labor. If everyone has the best available genes for a given task then there is no reason to specialize. This homogeneity might result from the limited size of the population, from the lack of environmental perturbations, and from the low mutation rate.

Finally, it is worth noting that collectives won't form if the birth/death rate is too high, or if the public goods benefit factor is too low. The later result is expected but the former is somewhat surprising. It might be due to the fact that newborn individuals with low cooperativeness destroy the collectives.

Figures and Tables

Figure 1

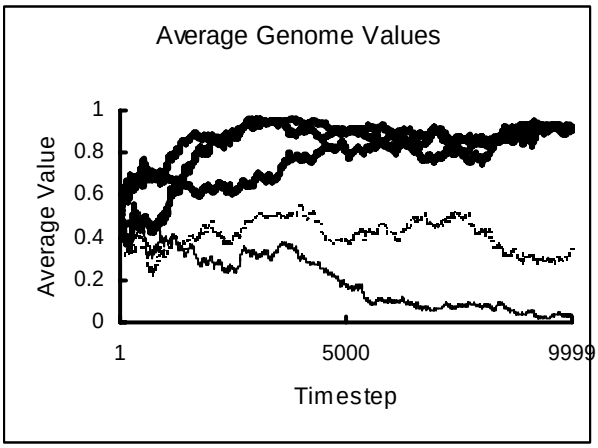


Figure 2

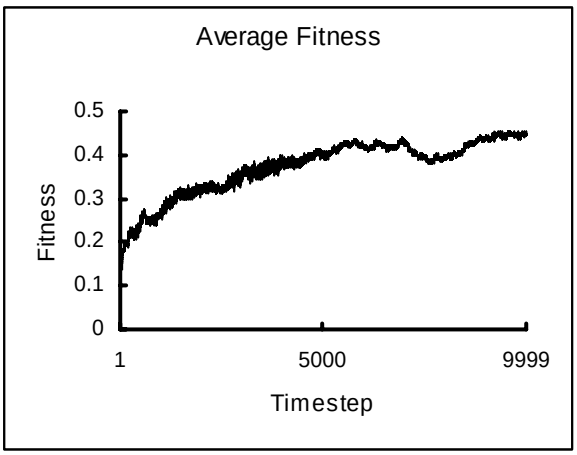
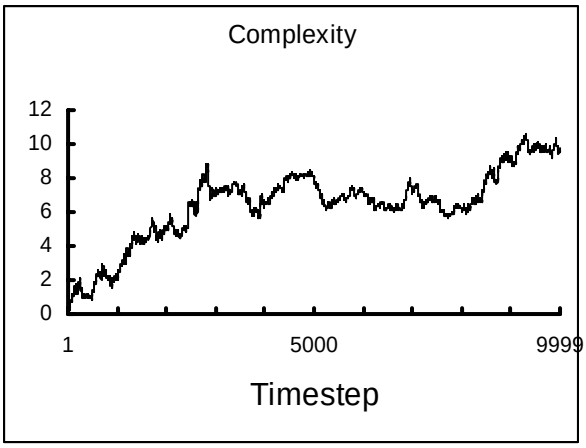


Figure 3



Figures 1-3. Results from a typical run of 10,000 timesteps. Figure 1 shows average aptitudes of individuals (dark solid lines), cooperativity (solid line), and staying threshold (dashed line).

Table 1

A	Benefit factor	1	3	5	10
	Mean lifetime	1.35	2.9	39.2	194.6
	*Birth-death rate was 0.05				

B	Birth-death rate	0.005	0.01	0.05	0.1
	Mean lifetime	117.7	39.2	13.4	8.5
	*Benefit factor was 5				

Table 2

A	Benefit Factor	1	3	5	10
	Fitness/Cooperativity	0.76	0.8	0.73	0.54
	Fitness/Lifetime	0.08	0.17	0.39	0.61
	Fitness/Freerider	-0.03	-0.08	-0.1	-0.28
	Fitness/Threshold	0.07	0.17	0.18	-0.35
	Lifetime/Collective Size	0.08	0.19	0.44	0.71
	*Birth-death rate was 0.05				

B	Birth-death rate	0.005	0.01	0.05	0.1
	Fitness/Cooperativity	0.64	0.73	0.73	0.7
	Fitness/Lifetime	0.47	0.39	0.41	0.39
	Fitness/Freerider	-0.14	-0.1	-0.27	-0.28
	Fitness/Threshold	0.09	0.18	0.06	0.09
	Lifetime/Collective Size	0.44	0.44	0.48	0.51
	*Benefit factor was 5				

Bibliography

Adami C. (1998) *Introduction to Artificial Life*, New York: Springer Verlag.

Adami C. (2002) "What is complexity?", *BioEssays* 24, 1085-1094.

Adami C., Ofria C. and Collier T.C. (2000) "Evolution of biological complexity", Proc. Nat. Acad. Sci. USA 97, 4463-4468.

Bonner J. T. (1988) *The Evolution of Complexity*, Princeton: Princeton University Press.

Carroll S.B. (2001) "Chance and necessity: the evolution of morphological complexity and diversity", *Nature* 409, 1102 – 1109.

Emmeche C. (1997) "Aspects of Complexity in Life and Science", *Philosophica* 59(1), 41-68.

Heylighen F. (1999) "The Growth of Structural and Functional Complexity during Evolution" in F. Heylighen, J. Bollen & A. Riegler (eds.) *The Evolution of Complexity*, Kluwer Academic, Dordrecht, 17-44.

Heylighen F. (2000) "Evolutionary transitions: how do levels of complexity emerge?", *Complexity* 6(1), 53-57.

May R.M. (1988) "How many species are there on earth?", *Science* 241, 1441-1449.

Maynard Smith J. and Szathmáry E. (1995) *The Major Transitions in Evolution*, W.H. Freeman, Oxford.

Maynard Smith J. and Szathmáry E. (1999) *The Origins of Life: From the Birth of Life to the Origin of Language*, Oxford University Press.

McShea D. W. (1996) "Metazoan complexity and evolution: is there a trend?", *Evolution* 50(2), 477-492.

McShea D.W. (2000) "Sense and depth", *Biol. Philosophy* 15, 751-758.

McShea D.W. (2000) "Functional complexity in organisms: parts as proxies", *Biol. Philosophy* 15, 641-668.

McShea D.W. (2001) "The minor transitions in hierarchical evolution and the question of a directional bias", *Journal of Evolutionary Biology* 14, 502-51.

Michod R.E. and Nedelcu A. M. (2003) "On the reorganization of fitness during evolutionary transitions in individuality", *Integrative and Comparative Biology* (in press).

Michod R. E. (2003) "Cooperation and conflict mediation in the evolution of multicellularity" in P.Hammerstein (ed.) *Genetic and Cultural Evolution of Cooperation*. MIT Press, Cambridge, MA.

Michod R. E. (2003) "Cooperation and conflict in the evolution of complexity" in H.Lipson, E.K.Antonsson, and J.R.Koza, (eds) *Computational Synthesis: From basic building blocks to high level functionality*. AAAI Press, Stanford, CA.

Sharpe O. (1999) Evolution at the Origins of Life: Modelling Simple Persistent Replicating Objects in Advances in Artificial Life, 5, 407-411, European Conference ECAL .

Stewart J. (2000) *Evolution's arrow: the direction of evolution and the future of humanity*. Rivett, ACT, Chapman Press.

Watson R.A. and Pollack J.B. (2003) "A computational model of symbiotic composition in evolutionary transitions", *Biosystems* 69,187-209.

Zuckerkandl, E. (1997) "Neutral and Nonneutral Mutations: The Creative Mix—Evolution of Complexity in Gene Interaction Systems", *Journal of Molecular Evolution* 44, S2-S8.