BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 26, 1964

ABSTRACT BIOLOGICAL SYSTEMS AS SEQUENTIAL MACHINES II: STRONG CONNECTEDNESS AND REVERSIBILITY

ROBERT ROSEN
Committee on Mathematical Biology, The University of Chicago

It was previously shown that the abstract biological systems called $(\mathscr{M}, \mathscr{R})$-systems could be regarded formally as sequential machines, and that when this was done, the reversibility of environmentally induced structural changes in these systems was closely related to the strong connectedness of the corresponding machines. In the present work it is shown that the sequential machines arising in this way are characterized by the property that the size of the input alphabet is very small compared with the size of the set of states of the machine. It is further shown that machines with this property almost always fail to be strongly connected. Therefore, it follows that one of the following alternatives holds: either most environmentally induced structural alterations are not environmentally reversible, or else many mappings in the category from which the $(\mathscr{M}, \mathscr{R})$-systems are formed must not be physically realizable.

1. *Introduction.* In a previous note (Rosen, 1964), we showed that an arbitrary $(\mathscr{M}, \mathscr{R})$-system could be represented as a sequential machine. We pointed out at that time that certain peculiarities arose in this representation, which have no counterpart in the general theory of sequential machines. In the present work, we explore these peculiarities somewhat further, and point out certain connections between our representation and some important problems relating to the realizability of abstract $(\mathscr{M}, \mathscr{R})$-systems, and to the reversibility of environmentally induced alterations in the structure of such systems.

We recall that it was shown in *loc. cit.* that a general sequential machine was specified by the set S of its states and a subset $\mathfrak{A}$ of the set $H(S, S)$ of all mappings of S into itself. This subset $\mathfrak{A}$ is connected with the "input alphabet" $\bar{\mathfrak{A}}$ of the machine in a 1–1 fashion via the next-state map δ; explicitly, we have the correspondence $\phi_a \to \bar{\alpha}$ determined by the relation $\phi_a(s) = \delta(\bar{\alpha}, s)$, where $\phi_a \in \mathfrak{A}$, $\bar{\alpha} \in \bar{\mathfrak{A}}$, $s \in S$. It was further shown that, if $\{f, \Phi_f\}$ is an $(\mathscr{M}, \mathscr{R})$-system of simplest form, where $f: A \to B$ and $\Phi_f: B \to H(A, B)$, then this system could be regarded as the sequential machine whose set of states is $S = H(A, B)$, and whose next-state map is defined by $\delta(a, f) = \Phi_f(f(a)), a \in A, f \in H(A, B)$. Combining the two results just stated, we find that the input alphabet to the sequential machine determined by the $(\mathscr{M}, \mathscr{R})$-system $\{f, \Phi_f\}$, which in this case is just A, may be regarded as a subset $\mathfrak{A}$ of the set $H[H(A, B), H(A, B)]$.

A heuristic description of our argument may now be given. Let us denote the cardinality of an arbitrary set T by $\nu(T)$. Our first observation is that, if $\nu(A) = \sigma$ and $\nu(B) = \tau$, where σ and τ need not for the present be finite, then the cardinality of the full set $H(A, B)$ of set-theoretic mappings of A into B is τ^σ, and the cardinality of $H[H(A, B), H(A, B)]$ is $(\tau^\sigma)^{\tau^\sigma}$. But clearly, if $\nu(B) > 1$, we will always have $\sigma \ll \nu(S)$; in fact, for $\sigma = 4$ and $\tau = 2$, $\nu(S)$ will be of the order of 10^{19}. That is: the input alphabet A to any sequential machine representing the $(\mathscr{M}, \mathscr{R})$-system must always be negligibly small compared with the total number of *mathematically possible* states of that machine (the use of the term "mathematically possible" will be clarified in a moment).

Next, we remark that a sequential machine whose input alphabet is very much smaller than its set of states is not likely to be strongly connected. The precise relationships existing between $\nu(A)$, $\nu(S)$, and the notion of strong connectedness do not seem to have been deeply explored in the literature; some partial results in this direction are given below.

Finally, we point out that not every mapping in $H(A, B)$ need be *realizable*, in the sense in which we have used that term (cf. Rosen, 1962, 1963). The totality of realizable maps in $H(A, B)$, which we may denote by $H_R(A, B)$, and which is bounded above by τ^σ, may actually have a cardinality very much less than τ^σ. Thus in any physical realization of the system $\{f, \Phi_f\}$, the total number of states which actually occur in the associated sequential machine may be very much less than the number $(\tau^\sigma)^{\tau^\sigma}$ of mathematically possible states, *depending on the size of* $H_R(A, B)$ *compared with* $H(A, B)$. Since we have shown (Rosen, 1964) that a close relationship exists between the notion of strong connectedness of a sequential machine associated with an $(\mathscr{M}, \mathscr{R})$-system Λ, and the reversibility of environmentally induced structural alterations in Λ, we are left with the following alternatives: *Either most mappings in an abstract set* $H(A, B)$ *are physically realizable, in which case most environmentally induced alterations in structure are irreversible in principle, or else many mappings are not realizable, and most environmental alterations can be reversed.* This rather paradoxical conclusion, which asserts roughly that the more things it is physically possible to do, the fewer things can be undone, will be further discussed below.

2. *Strongly Connected Sequential Machines.* The present section is devoted to a brief exploration of the problem of determining how many different sequential machines on the same set of states are strongly connected. More specifically, we wish to relate the number of such machines with the cardinality of the input alphabet, so that we may calculate how probable it is, roughly speaking, that an arbitrary sequential machine with a given set of states, and input alphabet of given cardinality, is strongly connected. As with all enumeration problems of this type, a complete solution is hindered by formidable combinatorial difficulties. However, a number of partial results in this direction are more or less readily obtainable, and will be described below. The discussion in this section is restricted to machines on a *finite* set of states S. To simplify our results we shall consistently write $N = \nu(S)$ and $K = N^N = \nu[H(S, S)]$.

According to our previous work (Rosen, 1964), a sufficient condition that a sequential machine (S, A) be strongly connected is that A contain a permutation consisting of a single cycle. If α is such a permutation, then the probability that a set $A \subset H(S, S)$, selected at random from the totality of all subsets of $H(S, S)$, shall contain α is just $\frac{1}{2}$ (since obviously there are 2^{K-1} subsets which do not contain α, and hence 2^{K-1} subsets which do). Thus we can already conclude that, in a certain sense, at least half of all possible sequential machines defined on S will be strongly connected.

This result can easily be sharpened. Since there are in general N permutations on S consisting of a single cycle, the probability that a set $A \subset H(S, S)$ shall intersect the set Π of all such permutations can readily be calculated. There are obviously 2^{K-N} subsets of $H(S, S)$ which do not intersect Π, and hence $2^K - 2^{K-N}$ subsets which do intersect Π. Thus, the probability that an arbitrary subset of $H(S, S)$ selected at random shall intersect Π is just

$$\frac{2^K - 2^{K-N}}{2^K} = 1 - \frac{1}{2^N}. \tag{1}$$

Stated otherwise, this result shows that the number of strongly connected sequential machines defined on S is at least $(1 - 1/2^N)$ times the total number of sequential machines defined on S, or $(1 - 1/2^N) \cdot 2^K$. We may note that, as N and K increase, the total number of strongly connected machines approaches the total number of machines; hence we may say that, if the set S is sufficiently large, almost every machine defined on S is strongly connected.

This result, although interesting and perhaps somewhat unexpected *a priori*, is only indirectly related to our main purpose. A moment's reflection will reveal that most of the sets A which contribute to the total number of strongly connected machines in the above argument are the "large" subsets of $H(S, S)$. For the "small" subsets A (i.e., those for which $\nu(A) \ll K$) the above line of argument gives no direct information concerning the number of sequential machines on S, with input alphabet of cardinality $\ll K$, which are strongly connected. This question, moreover, is our primary interest.

Let us therefore restrict attention to "small" subsets. If α represents a permutation without cycles, and A is a singleton (i.e., contains only one element), then the probability that A contains α is just $1/K$.

If A contains two elements, we can calculate the probability that A contains α as follows: there are $K(K-1)/2$ subsets of $H(S,S)$ which contain two elements, and clearly only $(K-1)$ of them can contain α. Thus the probability in question is

$$(K-1)\bigg/\frac{K(K-1)}{2} = \frac{2}{K}. \tag{2}$$

Likewise, if $\nu(A) = 3$, the probability that A contains α may be calculated: there are $K(K-1)(K-2)/2\cdot 3$ different subsets $A \subset H(S,S)$ such that $\nu(A) = 3$; moreover, the number of these subsets which contain α is readily seen to be the same as the number of different two-element subsets of $S - \{\alpha\}$; i.e., is given by $(K-1)(K-2)/2$. Thus the probability in question is given by

$$\frac{(K-1)(K-2)/2}{K(K-1)(K-2)/2\cdot 3} = \frac{3}{K}. \tag{3}$$

In general, if $\nu(A) = r$, then the probability that $\alpha \in A$ is given by r/K.

We may note explicitly that all these probabilities tend to zero as $K \to \infty$. Thus, if the above computations were perfectly general, it would follow that the number of strongly connected machines with alphabet A such that $\nu(A) \ll K$ is almost empty, in a sense, when K (i.e., S) is very large. The above results fail to be perfectly general for two reasons: (A) we have thus far considered only a single permutation α, whereas there are in general N such permutations α_1, $\alpha_2, \ldots, \alpha_N$; (B) in any case, the condition $\alpha_i \in A$, where α_i is one of the N permutations just specified, is only a sufficient and not a necessary condition for the strong connectedness of our machine. We can dispose of the second objection by noting that, for A small compared to S (which is the only case we are interested in) the condition expressed in (B) is "almost necessary" as well as sufficient; i.e., most of the strongly connected sequential machines which do have small alphabets owe their strong connectedness to the presence of such a permutation. Therefore, taking (B) into account does not affect our computations appreciably. We further assert that (A) also does not affect the limiting value of the probabilities derived above. We can show this by rederiving the probabilities in question, taking (A) into account. For example, if $\nu(A) = 1$, the probability that A contains at least one of the N permutations $\alpha_1, \alpha_2, \ldots, \alpha_N$ is just N/K. Remembering that $K = N^N$, we see that this ratio still tends to zero as K (and N) become larger. The computation for $\nu(A) = 2, 3$, etc. is complicated by the fact that sets containing more than one of the α_i must be counted only once in determining the number of sets of desired cardinality containing *at least one* of the α_i. For example, for $\nu(A) = 2$, the reader may verify that the expression corresponding to (2) above is

$$\frac{(K-1) + (K-2) + \cdots + (K-N)}{K(K-1)/2} = 1 - f(K, N),$$

where $f(K, N)$ is a readily computed expression converging to unity as K and N increase without bound; hence the entire expression converges to zero, as asserted.

The above arguments have shown that, in effect, the smaller is the ratio $\nu(A)/\nu(S)$, the more unlikely is the sequential machine determined by S and A to be strongly connected. The same result can be derived in a somewhat different manner by using methods of an asymptotic nature. These methods, which were developed in some detail by P. Erdös and A. Rényi (1960) in their work on the evolution of random graphs, seem to be applicable to a wide class of combinatorial problems of biological interest. Accordingly, it may be useful to briefly sketch how these ideas may be applied to the present problem.

Erdös and Rényi (*loc. cit.*) consider graphs formed from a set of N vertices, and a number n of edges chosen at random without repetition from the totality of all possible edges. The problem is to determine how the structure of the graph changes as n increases relative to N; in particular, it is desired to determine at what point in the "evolution" of these graphs a particular kind of structure (i.e., connectedness) becomes manifested by almost all graphs at that "evolutionary stage." This is accomplished by putting $n = n(N)$, and studying the asymptotic behavior of the probability that a particular structure is manifested in an arbitrary graph satisfying $n = n(N)$ as $N \to \infty$.

The results of the above-mentioned authors which are of particular interest for our present purposes are the following: (A) If $n(N) = cN$, where c is a real number greater than one-half, then the size of the largest component of a graph with N vertices and cN edges is given approximately by $G(c)N$, where

$$G(c) = 1 - x(c)/2c \quad \text{and} \quad x(c) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (2ce^{-2c})^k, \tag{4}$$

with probability tending to unity as $N \to \infty$; and (B) if the above conditions are satisfied, then almost every point of such a graph belongs either to a small component, which is a tree, or to the single "giant" component of size $G(c)N$.

These results can be applied to our present problems in the following way. Let (S, A) be a sequential machine. We shall say that two elements s_i, s_j of S are connected by an edge if there exists a map $f \in A$ such that $f(s_i) = s_j$ or $f(s_j) = s_i$ (we exclude the case $i = j$, in accordance with the usage of Erdös and

Rényi. This definition turns the set S into a graph, and it will be noted that the sequential machine (S, A) can be strongly connected only if the graph we have just defined consists of a single component. Suppose as before that $\nu(S) = N$, and let us restrict attention for the moment to sets A which are small relative to N. If A consists of only a single map, then the number n of edges in the graph we have just defined is precisely N. If A consists of a pair of mappings, then the number n of edges in that graph lies between N and $2N$; if A consists of three mappings, then $N \leqslant n \leqslant 3N$, etc. In general, the number of edges in the graph associated with a sequential machine (S, A) is bounded above by $n(N) = \nu(A)N$ (and indeed, if N is large compared to $\nu(A)$, it will be readily seen that the number of edges will in general actually be equal to $\nu(A)N$).

As we have noted previously, if the graph associated with a sequential machine has more than one component, then that sequential machine cannot be strongly connected. To see the meaning of the results (A and B) of Erdös and Rényi in this situation, we must calculate $G(\nu(A))$ for the various (small) cardinalities of A. We have approximately

$$G(1) = .8; \quad G(2) = .98; \quad G(3) = .999.$$

This means that for $N = 100$, say, $\nu(A) = 1$ implies that there are, on the average, only 80 vertices of the associated graph in the largest component; $\nu(A) = 2$ implies that there are approximately 98 vertices in the largest component, and $\nu(A) = 3, 4, 5 \ldots$ implies that all vertices lie in a single component. On the other hand, for $N = 1000$, $\nu(A) = 1, 2, 3$ implies respectively that 800, 980, and 999 vertices lie in the largest component, etc. Thus, we see as before that as the number of states of a sequential machine increases without bound, it becomes more and more improbable that a sequential machine with a small input alphabet will be strongly connected. We obtain, moreover, some numerical idea about the relative magnitudes involved. It must be observed, however, that the magnitudes so obtained are far from precise, because the strong connectedness of a sequential machine, and the connectedness of the graph which we have above associated with that machine, are not really closely related concepts. That is, many (if not indeed most) sequential machines whose graphs are connected will not themselves be strongly connected. As a result, the probability that an arbitrary sequential machine, with an input alphabet of specified cardinality, will be strongly connected is substantially smaller than the probability computed above, that the associated graph consists of a single component.

3. *Applications.* We have seen in the preceding section that, in a sense, it is infinitely unlikely for a sequential machine (S, A) satisfying the condition that $\nu(A) \ll \nu(S)$ to be strongly connected. According to the discussion in Section

1, moreover, any sequential machine arising from an $(\mathscr{M},\mathscr{R})$-system will *always* have this property, if $H(A, B)$ is taken to be the full set of set-theoretic maps of A into B. A computation of the orders of magnitude involved, even using the rough estimates of the preceding section, shows very quickly that, if A and B are finite, then *any sequential machine arising from the simplest possible* $(\mathscr{M},\mathscr{R})$-*system* $\{f, \Phi_f\}$ *is infinitely unlikely to be strongly connected.* From the manner in which the magnitudes behave, it is a fair conjecture that the same result will hold in the infinite case, although this would involve a much deeper analysis than was carried out in the preceding section. Moreover, it is seen immediately, via the representation of arbitrary $(\mathscr{M},\mathscr{R})$-systems as sequential machines, that the discussion given for the case of the simplest system carries over verbatim to the general case. That is, *any sequential machine which represents an* $(\mathscr{M},\mathscr{R})$-*system is infinitely unlikely to be strongly connected.*

We have seen in our previous work (Rosen, 1964) that the reversibility of environmentally induced alterations of the metabolic structure of an $(\mathscr{M},\mathscr{R})$-system, by means of purely environmental means, is possible only if the associated sequential machine is strongly connected. The result just derived makes it clear, however, that it is virtually certain for any such sequential machine to fail to be strongly connected. Thus we must conclude that, in any category which is as rich in mappings as the full category of sets, almost all environmentally induced metabolic alterations *will be irreversible by purely environmental means* (i.e., by applying a particular set of environmental inputs to the altered system). This should be contrasted with the results obtained in previous work (Rosen, 1961, 1963) which had indicated that reversibility required the existence of a large number of maps within the system (i.e., implied a *minimal* degree of "richness").

The only way to increase the probability that a strongly connected machine associated with an $(\mathscr{M},\mathscr{R})$-system be strongly connected (apart from adjoining mappings to the system, as described in Rosen, 1964) is to increase the size of the alphabet A relative to the set of states $S = H[H(A, B), H(A, B)]$. The only way to do this is to discard mappings in S; the natural candidates for pruning are the maps which are not realizable, in the sense in which we have employed this term. We may now ask: how many maps will have to be discarded before it becomes appreciably probable that the associated sequential machine is strongly connected? We can give a coarse estimate by employing some of the computations we have already carried out. We have seen above, for instance, that in the simple case for which $\nu(A) = 4$, $\nu(B) = 2$ we had $\nu(S) \sim 10^{19}$. If we calculate the size of the largest component of the graph with 10^{19} vertices for which $c = 4$ according to (4), we find this number to be $G(4)N \cong (.999634) \times 10^{19}$, which is quite significantly less than 10^{19}; in fact there will be some 3.66×10^{15} vertices which do not lie in the largest component. From these computations, we see that it will be necessary to reduce the size of S by a factor of at least 10^6 in order to satisfy even a weak necessary condition for the strong connectedness of a sequential machine on S with four input symbols. Stated otherwise, if the above example were to be taken seriously in a model for abstract biology, it would for at least one-third of the mappings in S fail to be realizable (in all probability the actual number of nonrealizable mappings in S would be of the order of half the total number of possible mappings in this case).

It is easily seen how this type of computation may be carried out in general. It should also be noticed that results of this type have implications for the number of realizable maps in $H(A, B)$ as well. In the example just discussed, for instance, it is necessary to reduce the size of the set S from about 10^{19} to about 10^{10} in order to render it appreciably probable that the sequential machine associated with the simple $(\mathscr{M}, \mathscr{R})$-system $\{f, \Phi_f\}$ is strongly connected (which is in turn sufficient to allow environmentally induced structural alterations to be reversed by appropriate sequences of environmental inputs alone). From this it follows that the number of maps in $H(A, B)$ itself which are realizable will be approximately 10 (instead of the full 16 maps which comprise the totality of set-theoretic maps of A into B), if we make the assumption that the nonrealizability of mappings in $S = H[H(A, B), H(A, B)]$ arises from the nonrealizability of maps in $H(A, B)$ itself. Once again, the principles employed in this special example may immediately be extended to the most general case; the details are left to the reader.

It is hoped to provide a detailed discussion of the relation between the above results and our previous work on reversibility and realizability in a forthcoming paper.

This study was supported in part by the Air Force Office of Scientific Research Grant #AF-AFOSR-9-64 and in part by a United States Public Health Service Research Career Development Award (#GM-K3-18, 431) from the National Institute of General Medical Sciences.

LITERATURE

Erdös, P. and A. Rényi. 1960. "On the Evolution of Random Graphs." *Publications of the Mathematical Institute*, Hungarian Academy of Sciences, **5**, 17–60.

Rosen, R. 1961. "A Relational Theory of the Structural Changes Induced in Biological Systems by Alterations in Environment." *Bull. Math. Biophysics*, **23**, 165–171.

———. 1962. "A Note on Abstract Relational Biologies." *Ibid.*, **24**, 31–38.

———. 1963. "On the Reversibility of Environmentally Induced Alterations in Abstract Biological Systems." *Ibid.*, **25**, 41–50.

———. 1964. "Abstract Biological Systems as Sequential Machines." *Ibid.*, **26**, 103–111.

RECEIVED 9-27-63