LIFE, INFORMATION THEORY, AND TOPOLOGY

N. Rashevsky

COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

The information content of an organism determines to a large extent its ability to perform
the basic vital functions: selection of food, breaking up of the food molecules into appropriate
parts, selection of those parts, and their assimilation. The information content needed is very
large and requires a sufficiently large complexity of the organism. The information content of
an organism is largely determined by the information content of the constituent organic mole-
cules. The information content of the latter is in its turn determined by the number of physically
distinguishable atoms or radicals of which the molecule is composed. The different arrange-
ments of atoms in a molecule are represented by the structural formula, which is basically a
graph. It is shown that the topology of this graph also determines to a large extent the infor-
mation content. Different points of a graph may be physically indistinguishable; in general,
however, they are different in regard to their topological properties. A study of the relations
between the topological properties of graphs and their information content is suggested, and
several theorems are demonstrated. A relation between topology and living processes is thus
found also on the molecular level.

Information-theoretical problems in biology have been recently raised
by a number of authors (Quastler, 1953; Morowitz, 1955). The following
information-theoretical aspect of very basic biological phenomena, though
not explicitly stated, seems to be implied in the interesting considerations
of J. von Neumann (1951) on self-reproducing automata.

Perhaps the most basic characteristic of living organisms is their ability
to construct other similar or even more complex organisms through as-
similation of certain component parts of their non-living environment.
The organism, whether multicellular or unicellular, or even a hypothetical
“living molecule,” does not have available in the environment its com-
ponent parts in a “pure form.” Those component parts are contained in
the environment as parts of other non-living entities. The organisms must
break up those non-living entities into smaller units in a proper way, and
then again select from this debris the component units needed for building
other similar living organisms. After such a selection the organism ar-
ranges the component units in a proper manner. This is the essence of
digestion, ingestion, and assimilation— the basic properties of the living.

Viewed as an automaton, the organism must have, in von Neumann’s
terminology (loc. cit., pp. 29–31), instruction for the proper break-up of
environmental units, the proper selection of the necessary components, and their proper arrangement. From the point of view of information theory, the organism must have a sufficiently large information content. It must contain all the information necessary for the proper break-up of the environmental unit (digestion), the selection of proper units resulting from the break-up, and the building of a duplicate organism from those units (absorption and assimilation).

Leaving aside the different, sometimes divergent, estimates of the information contents of cells and organisms (Quastler, loc. cit., pp. 251, 263), we can see offhand that the information content must be very large. In order to have a sufficiently large information content, the organism must be sufficiently complex. Therefore the above considerations put a lower limit on the complexity of even the simplest organisms.

This complexity may be to some extent measured by the number of different, distinguishable units of which the organism is composed. If we have any large number of indistinguishable units, the information content of each unit and of the whole aggregate is zero. If we have \( n \) distinguishable units with equal probability of choice, then the information content is equal to \( \log_2 n \).

If we consider as elementary units the different kinds of atoms of which living organisms are made, we find that this number is very limited—of the order of 10. Therefore second-order units must be made from combinations of those elementary units. Such second-order units are the diverse kinds of molecules of which an organism is composed. The situation is quite analogous to the formation of a very large number of distinct words from a limited number of letters in the alphabet of a written language.

From the point of view of this analogy the above problem of minimum complexity becomes formally analogous to the following one:

Given a written language, how large, in terms of the total number of words, must a book printed in that language be in order to contain complete information necessary to manufacture this book?

The answer depends on the nature of the environment, on what is given for the manufacturing. If a complete linotype machine, bookbinding machine, paper, ink, etc., are available, the book need not be too large. If, however, only rough materials are available out of which one must make the paper ink, machines, etc., then the size of the book will be beyond the practical limits of a single volume, and we shall have the problem of reproducing not a book but a whole library. In the case of the living the situation is nearer to that of the second alternative; hence the tremendous complexity of organisms.