J. theor. Biol. (1980) **87**, 421–456

On Order and Complexity. I. General Considerations

FRANK PAPENTIN

Institut für Biologie III, Universität Freiburg i.Br., Schänzlestr. 1, *D*-7800 *Freiburg i.Br., FRG*

(*Received* 29 *August* 1979, *and in revised form* 22 *December* 1979)

The length of the shortest possible description, called the minimal description of a pattern in a given language L is used as a measure of its complexity. This complexity is differentiated into an organized and an unorganized aspect. Organized complexity—the minimal description of the rules or laws underlying a pattern—is taken as a measure of its secondary order; unorganized complexity—the minimal description of the random aspects of a pattern—as a measure of its primary *dis*order. The average primary disorder of an ensemble of patterns is closely related to the entropy of that ensemble (if it is definable), whereas secondary order is beyond the entropy concept.

The procedures for the calculation of the complexity of a pattern are outlined and their relationship to formal language theory is given. The simplest of all patterns—binary sequences—are studied as a test case. It is found that unorganized complexity corresponds better to our intuitive notion of "disorder" than entropy does. The concepts of primary and secondary order are discussed with regard to the old controversy as to whether or not order increases or decreases in evolution. This controversy may be entirely settled if primary order is distinguished from secondary order: The first always tends to decrease in evolution and the second to increase. Moreover, the decrease of the first is the very driving force for the increase of the second; secondary order may, therefore, be said to emerge out of primary order.

1. Introduction

Order is one of the most basic concepts in science. All sciences can be said to be devoted to the study of one kind of orderliness or another. Yet the concept of order is poorly defined in science and elsewhere.

Perhaps most rigorously—but also most limitedly—order is defined in mathematics in terms such as the total and partial orderings of a set, the order of a differential equation or a group, etc. In physics, order, or rather disorder, is generally associated with entropy. Entropy, in its statistical version, is closely related to the concept of "information". In fact, the average information content of a set of signals has been called "entropy" by Shannon & Weaver (1949). Thus entropy and information—or concepts

0022–5193/80/230421+36 $02.00/0

derived from them, such as "negentropy" and "redundancy"—have been widely used in chemistry and biology as measures of order and organization (Schrödinger, 1951; von Foerster, 1960; Lwoff, 1962; Setlow & Pollard, 1967; Atlan, 1968; Blum, 1968; Gatlin, 1972; etc.). However, it has also been pointed out that the scale of order/disorder defined by entropy is inadequate to measure the kind of order—organization, structure, pattern—relevant for biology (Riedl, 1975; Saunders & Ho, 1976; Bresch, 1977).

Another way of measuring disorder or randomness is by means of "complexity", which has been defined as the length of the shortest possible description of a pattern or, more precisely, the shortest possible computer program to generate a pattern (Kolmogoroff, 1964; Chaitin, 1966). Complexity is a special way of estimating the information content of a pattern, and is as such related to entropy. The intuitively evident connection between complexity and entropy led van Emden (1970) in turn to define complexity in terms of the latter. Other notions of complexity have been discussed by Bremermann (1974). Recently, Bresch (1979) has developed a measure for complexity which he calls "pattern value", and Krüger (1979) for two aspects of order: homogeneity and symmetry.

The purpose of this series of papers is to bring "order" into the confusing issue of order by developing the notion of complexity pioneered by Kolmogoroff (1964) and Chaitin (1966). Practical procedures of calculating complexity will be layed out and applied to binary sequences (in this paper) and to more concrete patterns, such as snowflakes (Papentin, 1980) chemical, biochemical, and biological systems (in subsequent papers).

2. Derivation

The natural starting point for our derivation is *language*. Language not only is the universal means by which we analyse and describe all forms of order, but it also provides us with a means to *measure* order.

It is a general observation that less orderly, more complex, complicated, or random situations need more words for their description. For example, a number of books spread around in a "disorderly" manner in a room need quite a few words to be described: "Books A and B on the table, book C on the floor, book D in the drawer, etc. . . .," whereas an orderly arrangement of the books need much fewer words: "All books in alphabetical order on the shelf." We may, therefore, define tentatively *the length of the shortest possible description* of a system (describing it with a certain preset degree of accuracy) as being a *measure of its disorder.*

Of course, order by this definition would depend on language—or rather, the concepts implicit in language. This is, however, no basic flaw: it is simply in the nature of order for it to depend on the *state of our cognitive apparatus*. For example, the following sequence of numerals:

3 1 4 1 5 9 2 6 5 3 5 8 9 7 9 3

might look entirely random, i.e. "disorderly" for somebody not trained in mathematics. A mathematician, however, will immediately recognize it as the first 16 digits of π—the quotient of the circumference and the diameter of *any* circle—and thus something very systematic, non-random, i.e. "orderly".

Even more strikingly, the subjective nature of order—or rather, the dependence of order on the state of the cognitive apparatus of the observer—is seen in "alphabetic order":

a b c d e f g h i j k l m n o p q r s t u v w x y z.

For somebody not familiar with our letter system this arrangement of letters might look as arbitrary or "disorderly" as any other. For somebody grown up in our culture it does not look disorderly at all; in fact, by convention it has become a means of ordering all sorts of other objects (books, names in the telephone book, etc.).

Is the concept or order therefore something unscientific, useless?—No, it is just *relative* to a certain set of concepts: in other words, to language. Once the language has been fixed, order in our definition is exactly determined and is the same for all users of the language (see Appendix B, for proof).

Of course, there are more and less useful languages. For example, a language possessing different symbols for every pattern would not be very useful, because then each pattern could be described by a description of the same length—namely one—and thus achieve the same value of order.

Perhaps the most satisfactory kind of language would be one using only two kinds of concept, "zero" and "one" or "on" and "off", such as the language used by digital computers. In this case all other, more complicated concepts would have to be expressed in terms of these two most elementary concepts, and the length of the description of a pattern in such a language would be directly interpretable as the "information content" of the pattern. This is basically the approach taken by Kolmogoroff (1964) and Chaitin (1966) who defined the length of the shortest possible computer program, written in binary form, capable of generating a pattern, as a measure of its complexity, randomness or "disorder". However, in practice it is tedious to express everything in binary form or even to write a computer program capable of generating a given pattern. Moreover, this approach is not quite objective either: it is "subjective"—relative—to a given type of computer. Thus, in order to get a really practical measure for order we may as well use a "higher order" language. (The reader may excuse the multiple use of the word "order"—such are the intricacies of language!) We will see (section 3 and 4) that we obtain quite meaningful results if we stay close to the languages we already use, such as the symbols commonly used in mathematics.

We may now define more accurately the following terms.

(A) PATTERN

A certain number of objects connected together by a certain number of relationships we call with Bresch (1979) a "pattern". Thus a pattern for us generally means a connected graph, the objects being the "vertices" and the relationships being the "edges" of the graphs. However, we will also speak of a "pattern" when the relationships are not explicitly given, such as in sequences and arrays of symbols, building blocks put together, etc.

This definition of a pattern is sufficiently general and specific to be useful. Almost all natural or artificial systems can be modelled with a sufficient degree of accuracy by connected graphs and their simplified derivatives.

(B) COMPLEXITY

The length of the shortest possible or *minimal* description of a pattern, measured by the number of elementary ("structureless") symbols contained in it—with certain additions to be made in section 3.D—we call, in the spirit of Kolmogoroff (1964) and Chaitin (1966), the *complexity* of the pattern:

$$C := \mathrm{Min}(n_i),$$

where n_i are the length of all possible descriptions of the pattern in a given language L.

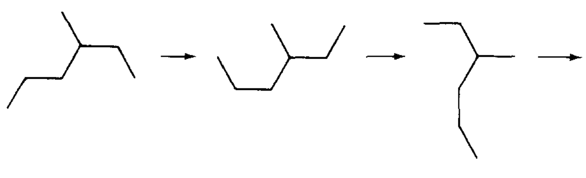
Complexity may be differentiated into *organized* and *unorganized* complexity: By organized complexity we understand the length of the minimal description of the rules or laws underlying the formation of a pattern, and by unorganized complexity the length of the minimal description of the random aspects of the pattern, such as the parameters entering the rules and/or the description of pattern aspects which cannot be described by rules. In many cases total complexity will simply be the sum of unorganized and organized complexities (see section 2.D, for numerical examples):

$$C = C_{\mathrm{org}} + C_{\mathrm{unorg}}$$

Of course, any differentiation of the complexity of a pattern into an organized and an unorganized part can only be made on the basis of further information, either direct information about the origin of the pattern or indirect information deduced from a large ensemble of patterns having the same origin, i.e. having been produced by the same mechanism, force or law. The differentiation of the complexity of an *individual* pattern with no other information than the pattern itself is nothing but a hypothesis regarding its origin: the rules, laws or forces underlying its formation. This is because an apparently very regular pattern might not have been generated by a rule at all, but by chance, since every—even the most regular—pattern has a certain probability of being generated by chance. Conversely, an apparently very irregular pattern might not have been generated by chance, but by a highly complicated rule which we are unable to comprehend. For example: the base sequence of an individual DNA molecule (supposing that we were able to decipher it) might look entirely irregular or random, however, if it has been extracted from an organism we know that it is not random at all: indeed, it codes for most of the organism!

If we were confronted with a large ensemble of DNA molecules whose source is unknown, we would regard them as organized, i.e. produced by some specific mechanism, if their base sequences were identical, however, as unorganized, if their base sequences were different. The distinction between organized and unorganized complexity is, therefore, rather easy in the case of large ensembles of patterns.

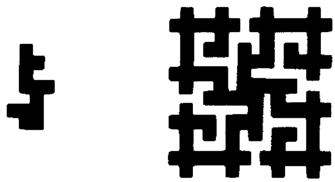
We may also consider a physical system viewed at different instants in time as an ensemble of patterns. For example, different "snapshots" of a hydrocarbon changing under the influence of thermal agitation may form the following ensemble of patterns:



The organized complexity of these patterns, of course, consists in the shortest possible description of the graph representing the binding relationships of the hydrocarbon, and the unorganized aspect in the shortest possible description of the specific three-dimensional configuration it assumes at a particular point in time. In general we may say that the organized complexity of a physical system has to do with its *invariants* (laws, symmetries, structural invariants), and the unorganized complexity with the *variant* expression of these invariants in space and time.

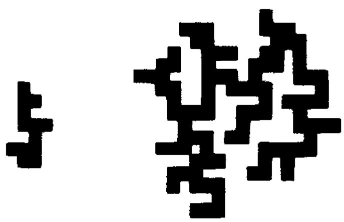
(C) ORDER

As a first attempt, as suggested in the introduction, we may simply take total complexity as a measure of disorder. This can, however, lead to odd results. For example, in comparing the following two patterns:



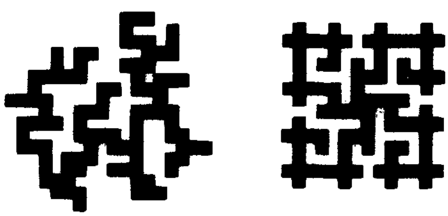
we would have to regard the right pattern as more disorderly, since it is clearly more complex (its minimal description is longer). This contradicts our intuition which tells us that the right pattern is more orderly, since it is much more regular.

Comparing the following two patterns:



our intuition might tell us—in accordance with the complexity criterium—that the right pattern is more disorderly. However, it may be argued that the two patterns are not comparable because they are different in size, and that it is thus meaningless to call one more disorderly than the other.

We will therefore restrict ourselves to the consideration of patterns of the same size, i.e. patterns which are composed of the same number of objects and relationships, such as the following two patterns:



No argument will arise over the assertion that the left pattern is indeed more disorderly than the right. Under the restriction made above we will now define the *disorder* $\mathscr{D}$ of a pattern as its *unorganized complexity*:

$$\mathscr{D} := C_{\text{unorg}}.$$

The order o of pattern may then be defined as the difference between its unorganized complexity and the unorganized complexity $C_{\max}$ of a maximally disorderly pattern of the same size:

$$o := C_{\max} - C_{\text{unorg}}$$

$C_{\max}$ is realized if all objects and relationships of a pattern are different and arranged—as a first approximation—in a totally random, chaotic way. (We will see in section 4.B that both by intuitive judgement and by the complexity criterium maximum disorder does *not* coincide with maximum randomness or chaos.)

It follows that:

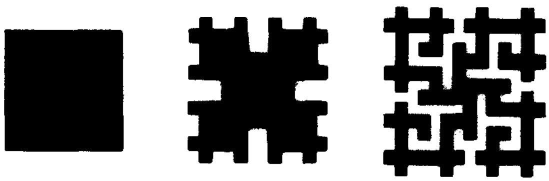
$$o_{\max} = C_{\max}$$

and

$$o + \mathscr{D} = C_{\max} = \text{const}$$

for all patterns of the same size.

Is the kind of order defined so far really the only one conceivable? If we compare the following three patterns:



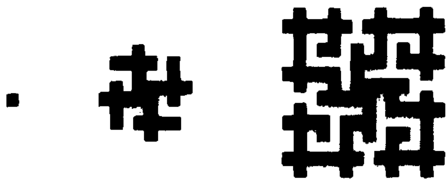
it makes sense to say that their "order" increases from left to right, even though their order according to the previous definition apparently remains the same (because their unorganized complexities remain the same, namely zero). This second, intuitively evident kind of order has to do with *organization*: the complexity of rule standing behind these patterns. It is, therefore, meaningful to define this *secondary order* by *organized complexity*:

$$o_{\text{sec}} := C_{\text{org}}.$$

For distinction we call the kind of order defined previously *primary order*:

$$o_{\text{prim}} := C_{\max} - \mathscr{D}_{\text{prim}}.$$

In contrast to primary order, secondary order may be applied without difficulties to patterns of different size. For example, the following three patterns



may well be said to increase in secondary order—degree of organization—from left to right. It again holds that:

$$o_{\text{sec,max}} = C_{\max}.$$

The maximum value of secondary order is reached if a pattern, when viewed in isolation, appears completely random, but when viewed in the context of a large ensemble of patterns—because all members of the ensemble are identical—turns out to be completely non-random (such as DNA as mentioned above).

The total order of a pattern may not exceed

$$2C_{\max} \geq o_{\text{prim}} + o_{\text{sec}}.$$

(D) SOME SIMPLE NUMERICAL EXAMPLES

(1) We may imagine a transmitter emitting just one type of letter sequence, which may consist of 10 different letters, for example:

$$p \quad a \quad x \quad n \quad l \quad f \quad b \quad c \quad g \quad m.$$

This sequence may be taken as its own (minimal) description, having a length of 10; thus a complexity of:

$$C = 10.$$

This is at the same time the maximum complexity possible for a sequence of this size:

$$C_{\max} = 10.$$

Because all sequences emitted by the transmitter are identical, their entire complexity is organized, thus:

$$C_{\text{org}} = \mathcal{O}_{\text{sec}} = 10, \quad \text{and} \quad C_{\text{unorg}} = \mathcal{D}_{\text{prim}} = 0,$$

and therefore:

$$\mathcal{O}_{\text{prim}} = C_{\max} - \mathcal{D}_{\text{prim}} = 10.$$

It follows that:

$$\mathcal{O}_{\text{total}} = \mathcal{O}_{\text{prim}} + \mathcal{O}_{\text{sec}} = 20 = 2C_{\max}$$

(2) For comparison, we may imagine another transmitter emitting just one type of letter sequence, consisting of the repetition only one kind of letter, for example:

$$a \; a \; a \; a \; a \; a \; a \; a \; a \; a$$

We may describe this situation by:

$$a^{10}$$

i.e. we are treating the gaps in between the as—in accord with mathematical convention—as multiplication signs (the description language is treated in more detail later). Simply counting the number of symbols used, we gain a complexity of

$$C = 3.$$

Again all of this complexity is organized, hence:

$$C_{\text{org}} = \mathcal{O}_{\text{sec}} = 3, \quad \text{and} \quad C_{\text{unorg}} = \mathcal{D}_{\text{prim}} = 0,$$

$$\mathcal{O}_{\text{prim}} = C_{\max} - \mathcal{D}_{\text{prim}} = 10, \quad \text{and}$$

$$\mathcal{O}_{\text{total}} = \mathcal{O}_{\text{prim}} + \mathcal{O}_{\text{sec}} = 10 + 3 = 13.$$

(3) Finally, we may imagine a transmitter emitting all possible letter sequences of length 10 in random order. Then the *average* complexity of these sequences is almost equal to the maximum possible complexity:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i \approx C_{\max} = 10$$

where $n = 26^{10}$ is the number of all possible sequences (presupposing the English alphabet), and C_i the complexity of the ith sequence.

In other words, the majority of the sequences of this ensemble may not be described by any sequence of symbols shorter than themselves. The organized complexity of every sequence in this case is:

$$C_{\text{org}} = \mathcal{O}_{\text{sec}} = 0,$$

hence the average unorganized complexity or primary disorder of these sequences

$$\bar{C}_{\text{unorg}} = \mathscr{D}_{\text{prim}} \approx 10,$$

and therefore:

$$\bar{o}_{\text{prim}} = C_{\text{max}} - \bar{\mathscr{D}}_{\text{prim}} \approx 0,$$

$$\bar{o}_{\text{total}} = \bar{o}_{\text{prim}} + \bar{o}_{\text{sec}} \approx 0.$$

The values for *individual* members of the ensemble may be quite different. For example the sequences:

(i) $a\ a\ a\ a\ a\ a\ a\ a\ a\ a,$
(ii) $a\ b\ a\ b\ a\ b\ a\ b\ a\ b,$
(iii) $a\ a\ a\ b\ b\ b\ c\ c\ c\ c,$

may be described by following symbol sequences:

(i) a^{10}
(ii) $(ab)^5$
(iii) $a^3b^3c^4$.

These yield unorganized complexities or primary disorders of 3, 5, and 6, respectively—all significantly below the average.

Whereas secondary order, by definition, is the same for all members of an ensemble (in the last case it equals zero) and thus may not serve to compare its different members, primary order is generally different for the members of an ensemble and thus may serve to compare them. However, if different ensembles are to be compared, both the secondary order and the *average* primary order or disorder of their members are useful. In fact, the average primary disorder of the members of an ensemble is very close to a concept used in information theory and thermodynamics—*entropy*—as we will see in Appendix A.

3. Procedure

In the last section we suggested that *complexity* is the most important concept with respect to order. As *unorganized complexity* it gives a measure of *primary disorder* and as *organized complexity* a measure of *secondary order*.

Then, how do we calculate complexity?—We may proceed in four steps:

(1) "Modelling" a given pattern, generally by a finite, connected graph i.e. a finite number of objects connected by a finite number of relationships;

(2) "coding" the model by a linear array of symbols, yielding what we may call the "rough" description;
(3) "compressing" the rough description according to certain, specific rules—amounting to a description of this description by another, higher order language—yielding the "fine" description; and
(4) "evaluating" that fine description in order to obtain a measure of complexity of the underlying pattern.

(A) MODELLING

The first step in modelling is to decide how accurately a pattern is to be represented; this decision will, of course, have a profound effect on the absolute value of complexity we will obtain. However, the *absolute* value of complexity is of little interest. What is of interest is the *relative* sequence of increasing or decreasing complexity a given set of patterns is arranged in. For this purpose it is sufficient that all patterns to be compared are modelled with the same degree of accuracy. The decision of how accurately a given set of patterns is modelled may be guided by the properties of the patterns, by practical considerations (a more detailed model, of course, means more work!), or be to a large extent arbitrary. In many cases, however, this decision will be suggested by the patterns themselves. For example, we may model society down to the level of its organizations, its individual members, the cells of its members, or even down to the molecules constituting these cells: For this and other *hierarchically organized* systems there exist clear cut layers, which offer natural steps to the model-building process.

(B) CODING

The first step in coding a model consists of devising a code for the elements occurring in the model. Taking a simple linear graph as an example:

○—▲=△~◫~◭~■····□—●~■····△~△=⦶

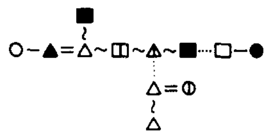
we may devise a code as follows:

a: ●	*d*: ■	*g*: ▲	*z*: —
b: ⦶	*e*: ◫	*h*: ◭	*y*: =
c: ○	*f*: □	*i*: △	*x*: ····
			w: ~

and describe (code) the above graph as follows:

$$c\ z\ g\ y\ i\ w\ e\ w\ h\ w\ d\ x\ f\ z\ a\ w\ d\ x\ i\ w\ i\ y\ b.$$

Branched graphs, for example:



may be coded by introducing a different type of symbol "R_i" indicating that branch i leaves the reference branch at a particular object.

We will agree to write these symbols directly before or after the symbol of the object at which the respective branch leaves the reference branch. The branch symbols are then specified after the specification of the "main branch", in the order of their indices, separated from it and from one another by semicolons. Thus, the above graph is coded as follows:

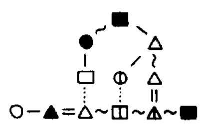
$$\underbrace{c\ z\ g\ y\ i\ R_1\ w\ e\ w\ h\ R_2\ w\ d\ x\ f\ z\ a}_{\text{main branch}};\ \underbrace{w\ d}_{R_1};\ \underbrace{x\ i\ R_3\ w\ i}_{R_2};\ \underbrace{y\ b}_{R_3}$$

Of course, it is arbitrary which branch is declared the "main branch". The above graph may also be coded in the following way:

$$\underbrace{c\ z\ g\ y\ i\ R_1\ w\ e\ w\ h\ R_2\ x\ i\ R_3\ w\ i}_{\text{main branch}};\ \underbrace{w\ d}_{R_2};\ \underbrace{w\ d\ x\ f\ z\ a}_{R_2};\ \underbrace{y\ b}_{R_3}$$

As a general rule, if there are more than one possibilities of coding a graph—i.e. if it possesses more than one rough description—we choose the rough description which yields the shortest fine description (see next section). The notation introduced above may also be used to code cycles simply by indicating the branch forming the cycle twice: first next to the object where it leaves the reference branch, and second next to the object where it joins the same or another reference branch. For example:

Graph:



Rough description 1:

$$\underbrace{a\ z\ g\ y\ i\ R_1\ w\ e\ R_2\ w\ h\ R_1\ w\ d;}_{\text{main branch}}\ \underbrace{x\ f\ z\ a\ w\ d\ x\ i\ R_2\ w\ i\ y;}_{R_1}\ \underbrace{x\ b\ z}_{R_2}$$

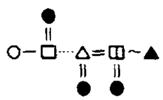
Rough description 2:

$$\underbrace{a\ z\ g\ y\ i\ R_1\ x\ f\ z\ a\ w\ d\ x\ i\ R_2\ w\ i\ y\ h\ R_1\ w\ d;}_{\text{main branch}}\ \underbrace{w\ e\ R_2\ w;}_{R_1}\ \underbrace{z\ b\ x}_{R_2}$$

etc.

Symbols for "true" branches may be used as often as this particular branch type occurs in a model. Symbols for branches leading to cycles may only be used twice, even if the same branch type occurs in several cycles. Using the same symbols in this case would lead to ambiguities in the connectedness of the model.
Examples:

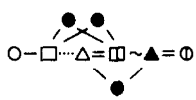
"True" branches:



Description:

$$c\ z\ f\ R\ x\ i\ R\ y\ e\ R\ w\ g;\ y\ a$$

Branches leading to cycles:



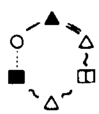
Ambiguous description (forbidden):

$$c\ z\ f\ R\ R\ x\ i\ R\ R\ y\ e\ R\ w\ g\ R\ y\ b;\ z\ a\ z$$

Proper description:

$$c\ z\ f\ R_1\ R_2\ x\ i\ R_1\ R_3\ y\ e\ R_2\ w\ g\ R_3\ y\ b;\ z\ a\ z;\ z\ a\ z;\ z\ a\ z$$

The cycle notation may also be used to specify simple rings, for example:



Rough description:

$$c R_1\ z\ g\ y\ i\ w\ e\ w\ h\ w\ d\ R_1,\ x$$

However, in this case we also use the notation "#", meaning that the respective branch is joined end to end. We agree to write this symbol always in front of the branch in question:

$$\#\ c\ z\ g\ y\ i\ w\ e\ w\ h\ w\ d\ x$$

With these symbols any connected graph, however complicated it may be, can be coded in an unambiguous way.

In case a certain set of graphs has a specially simple form, for example if all objects or relationships are identical, the procedure of coding them may be simplified, by neglecting the respective objects or relationships.

(C) COMPRESSING

The rough description of a pattern is a one-dimensional array of symbols and in the following will be treated as such, independent of its "meaning". It may be compressed according to two sets of criteria: "subsequences" and "regularities". The compressed version of the rough description of a pattern is also called (one of) its "fine descriptions". The purpose of compressing is, of course, to find the shortest possible fine description which will serve to measure the complexity of the pattern.

(*i*) *Subsequences*

Any subsection of a rough description may be declared a subsequence and replaced by an appropriate symbol. This symbol, however, has to be specified within the fine description. By analogy with the branches discussed in the previous section we will specify it after the main sequence, separated from it in this case by a comma. In case more than one subsequences are declared, they are indexed and specified again in the order of their indices, separated from one another and the main sequences by commas. As subsequence symbol we choose—in correspondence to formal language theory (see Appendix B)—the letter S. For example:

Rough description:

$$n\ r\ \underbrace{a\ b\ c\ d}\ q\ x\ v\ \underbrace{a\ b\ c\ d}\ h\ \underbrace{a\ b\ c\ d}\ y\ r\ v\ h, \quad \text{length} = 22.$$

Fine description:

$$\underbrace{n\ r\ S\ \ q\ x\ v\ S\ \ h\ S\ \ y\ r\ v\ h}_{\text{main sequence}},\ \underbrace{a\ b\ c\ d}_{S}, \quad \text{length} = 18.$$

It is, of course, only meaningful to declare subsequences if the resulting fine description is shorter than the original rough description. This is secured if a sequence of length 2 is repeated at least four times, a subsequence of length 3 at least three times, and a subsequence of length 3 at least two times.

It may be that a subsequence is not repeated as such but only after a certain operation, such as an inversion, has been performed on it. For example:

$$n\ r\ \underbrace{a\ b\ c\ d}\ \underbrace{d\ c\ b\ a}\ q\ v\ x\ x\ \underbrace{d\ c\ b\ a}.$$

We will describe this situation by introducing another symbol "←" denoting "inversion". The above rough description may then be compressed as follows:

$$\underbrace{n\ r \leftarrow S\ S\ q\ r\ x\ x\ S,}_{\text{main sequence}}\ \underbrace{d\ c\ b\ a}_{S}.$$

Another operation which may be performed on a subsequence is "complementation". In the case where a rough description contains only two kinds of symbols complementation consists simply in the exchange of the two symbols. For example:

$$\underbrace{a\ b\ b\ b\ a\ a\ b}\ a\ a\ b\ a\ \underbrace{b\ a\ a\ a\ b\ b\ a}\ a\ b\ a.$$

This may be described by introducing another symbol "↑" denoting "complementation". The above rough description may thus be compressed as follows:

$$\underbrace{S\ \ a\ a\ b\ a \uparrow S\ \ a\ b\ a,}_{\text{main sequence}}\ \underbrace{a\ b\ b\ b\ a\ a\ b}_{S}.$$

We may extend the meaning of complementation to any *permutation* of the symbols occurring in a "reference sequence" leaving the structure of the reference sequence unchanged. (A concrete example of this situation is found in the complementary base sequences of DNA.) We will agree to specify the permutation in between the symbol denoting complementary and the respective subsequence symbol, and specify the permuted symbols in the order of their occurrence. As a reference sequence we will use the first relevant subsequence of the rough description.

For example:

$$r\ q\ \underline{c\ a\ a\ b\ b\ b\ c\ b\ c}\ p\ r\ s\ t\ \underline{b\ c\ c\ a\ a\ a\ b\ a\ b}\ a\ b.$$

Fine description:

$$\underbrace{r\ q\ S\quad p\ r\ s\ t \uparrow b\ c\ a\ S\quad a\ b,}_{\text{main sequence}}\ \underbrace{c\ a\ a\ b\ b\ b\ c\ b\ c,}_{S\,=\,\text{reference sequence}}$$

where the expression "$\uparrow\ b\ c\ a\ S$" means that

c is substituted for by b,

a is substituted for by c,

b is substituted for by a

in the reference sequence

(ii) Regularities

A regularity is anything which may be described by a mathematical rule, such as the n-fold repetition of the same symbol. For example:

$$a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a\ a$$

As introduced in section 2.E this situation may be described as:

$$a^{16}.$$

If a certain sequence is repeated n times, it either has to be included in brackets or to be declared a subsequence (which leads to a fine description of the same length). For example:

$$a\ b\ a\ b\ a\ b\ a\ b\ a\ b\ a\ b\ a\ b\ a\ b$$

Fine descriptions:

$$(a\ b)^8 \quad \text{or} \quad S^8,\ a\ b$$

For regular repetitions of a letter in an otherwise random sequence, for example:

$$p\ \underline{a}\ q\ \underline{a}\ b\ \underline{a}\ c\ \underline{a}\ f\ \underline{a}\ h\ \underline{a}\ n\ \underline{a}\ o\ \underline{a},$$

we adopt the following notation:

$$2i = a \mid p\ q\ b\ c\ f\ h\ n\ o.$$

understanding that the variable $i \in \mathbb{N}$ runs over the entire range relevant for the letter sequence in question, and that position $2i$ is occupied by "a" [in general: $f(i) = y$ means that position $f(i)$ is occupied by y]. All letters not specified by the rule are given in the order of their occurrence immediately after the symbol "|". If the regularity does not cover the entire sequence, we mark its beginning and/or end by the symbol "$\sim$".

A few more examples will clarify this notation:

(1) Rough description:

$$\underline{a}\ y\ \underline{a}\ f\ g\ \underline{a}\ k\ o\ r\ \underline{a}\ n\ v\ h\ c\ \underline{a}\ s\ r\ y\ m\ p\ \underline{a}$$

Fine description:

$$\Sigma i = a \mid y\ f\ g\ k\ o\ r\ n\ v\ h\ c\ s\ r\ y\ m\ p$$

(2) Rough description:

$$p\ n\ \underline{a}\ h\ x\ \underline{a}\ n\ y\ b\ r\ \underline{a}\ v\ o\ p\ a\ k\ l\ \underline{a}\ m\ x\ y$$

Fine description:

$$p\ n \sim i^2 = a \mid h\ x\ n\ y\ b\ r\ v\ o\ p\ a\ k\ l \sim m\ x\ y,$$

In the last case, however, it does not pay to use this notation, since the "fine description" is longer than the original rough description.

(3) The notation may also be used without difficulties twice:

Rough description:

$$n\ \underline{a}\ x\ \underline{a}\ b\ \underline{a}\ r\ \underline{a}\ s\ \underline{a}\ b\ \underline{a}\ p\ \underline{a}\ h\ \underline{a}\ b\ \underline{a}\ x\ \underline{a}\ v\ \underline{a}\ b\ \underline{a}\ z\ \underline{a}\ n\ \underline{a}\ b\ \underline{a}$$

Fine description 1:

$$2i = a \mid n\ x\ \underline{b}\ r\ s\ \underline{b}\ p\ n\ \underline{b}\ x\ r\ \underline{b}\ z\ n\ \underline{b}$$

Fine description 2:

$$2i = a \mid 3i = b \mid n\ x\ r\ s\ p\ n\ x\ r\ z\ n$$

(D) EVALUATING

Evaluating the length of a fine description does not seem to pose any problems: it simply consists in counting the number of symbols occurring in it.

However, in certain cases this simple approach may lead to odd results: for example, when an ensemble of patterns consists of linear sequences of just one type of element. Then the fine descriptions all have the form "a^n", where "a" denotes the type of element and "n" the number of its repetitions. The complexities of the patterns then would be a discontinuous function of n; all sequences of length 1–9 would gain the complexity 2, all sequences of length 10–99 the complexity 3, etc. (using the usual decadic numerals). In order to avoid this oddity we may introduce the following refinement: we do not count the "exponents" and instead take them into account logarithmically. Thus, our final definition of the "length" of a fine description looks as follows:

$$n = n_a + \Sigma \log m_i,$$

where n_a is the number of symbols except the "exponents", and m_i the "exponents" occurring in it. The type of logarithm used is arbitrary. In this and following papers we use decadic logarithms, because in this way a simple counting of the total number of symbols occurring in a fine description can be used as an (upper) approximation of the complexity of a pattern.

One final remark: Complexity was defined (section 2.B) as the length of the shortest of *all* possible fine descriptions of a pattern (understanding "length" now in the refined sense defined above). In practice, however, it is seldom possible to examine all possible fine descriptions of a pattern; the complexity of a pattern (even though it exists, see Appendix B) may therefore only be given as an upper approximation, i.e. having found a "very short" fine description one generally cannot be sure whether this really is the shortest one possible. How close one gets to the real complexity depends on the amount of time and work one invests.

4. Application

In this section we will apply the formalism introduced in the previous sections to the simplest of all conceivable patterns: binary sequences. For better optical evaluation we display them as sequences of "x" and ".". Binary sequences can be regarded as their own models and rough descriptions so that it only remains to find their shortest possible fine descriptions.

(A) SOME SAMPLE SEQUENCES

(*i*) *A random sequence*

Generated by tossing a coin:

$$\underbrace{.\ x\ .\ x\ x\ .\ x}_{S}\ .\ \underbrace{x\ x}_{S}\ x\ x\ .\ .\ .\ .\ .\ .\ \underbrace{x\ .\ x\ .\ .\ x}_{S}\ .\ .\ x\ x\ x\ .\ .\ .$$

Fine description:

$$S^2xS^2x^3\ .\ {}^5S^2\ .\ S\ .\ Sx^2\ .\ {}^3,\ .\ x.$$

Complexity:

$$C \leq 14 + 4 \log 2 + 2 \log 3 + \log 6 \approx 16{\cdot}94.$$

The average of ten such random sequences was found to be $\bar{C} \leq 16{\cdot}84$. This is not—as implied by the term "average"—the maximum value possible for a binary sequence of this length. It is possible to construct sequences with a complexity greater than 19, for example the following

(*ii*) *non-random above average sequence*:

$$x\ .\ .\ .\ x\ .\ .\ x\ x\ .\ x\ x\ .\ .\ x\ x\ x\ .\ x\ x\ .\ x\ .\ x\ .\ .\ x\ x\ .\ x\ .\ x$$

$$S_2 \qquad \uparrow S_1 \qquad S_1 \qquad\qquad S_1 \qquad\qquad \uparrow S_1 \quad S_2\ S_2$$

Fine description:

$$S_2\,.\uparrow\ S_1^2\ .\ x^2 S_1 x \uparrow S_1 S_2^2 x,\ S_2 x S_2,\ x\ .$$

Complexity:

$$C \leq 19 + 3, \log 2 \approx 19{\cdot}90.$$

This sequence is "non-random" in the sense that it is not typical for sequences generated by tossing a coin. It takes quite an amount of effort (attention, intelligence) to construct such as sequence. The highest complexity and thus primary disorder (if the complexity is of unorganized nature) may therefore *not* be associated with the highest degree of randomness or chaos. We will come back to this interesting point in the next section and consider first some non-random (in the above sense) sequences of below average complexity:

(*iii*) *A doublet sequence*

Generated by tossing a coin, but "head" was interpreted as "xx" and "tail" as ". .":

$$.\ .\ x\ x\ .\ .\ x\ x\ x\ x\ .\ .\ x\ x\ .\ .\ x\ x\ x\ x\ x\ x\ x\ x\ .\ .\ .\ .\ .\ .\ .\ .$$

$$S \qquad S \qquad\quad S \qquad S$$

Fine description:

$$S^2 x^2 S^2 x^6\ .^8,\ .^2 x^2$$

Complexity:

$$C \leq 8 + 5 \log 2 + 2 \log 4 + \log 8 \approx 11{\cdot}19.$$

This value is clearly below the average for random singlet sequences.

(*iv*) *A symmetrical sequence*

First half generated by tossing a coin, second half by inversion of the first half:

$$x\ .\ .\ x\ x\ x\ .\ .\ .\ .\ .\ .\ x\ x\ .\ x\ |\ x\ .\ x\ x\ .\ .\ .\ .\ .\ .\ x\ x\ x\ .\ .\ x$$

Fine description:

$$S \leftarrow S,\ x\ .\ {}^{2}x^{3}\ .\ {}^{6}x^{2}\ .\ x$$

Complexity:

$$C \leq 11+2\ \log\ 2+\log\ 3+\log\ 6 \approx 12{\cdot}85.$$

Again this value is clearly below the average for random, asymmetrical sequences. The same is true for

(*v*) *A complementary sequence*

First half generated by tossing a coin, second half by complementation of the first half:

$$.\ x\ x\ x\ x\ .\ .\ x\ x\ .\ x\ x\ .\ x\ x\ x\ |\ x\ .\ .\ .\ .\ x\ x\ .\ .\ x\ .\ .\ x\ .\ .\ .$$

Fine description:

$$S_1 \uparrow S_1,\ S_2x^2\ .\ S_2^3x,\ .\ x$$

Complexity:

$$C \leq 12+2\log 2+\log 3 \approx 13{\cdot}08$$

These and other sequences can be most easily compared in a table in descending complexity (see Table 1).

The complexities of these sequences have to be regarded as unorganized if they are considered to be samples of the ensemble of all possible binary sequence of length 32 (there is nothing organized about these sequences except that they are binary sequences of length 32). From this standpoint the above values have to be regarded as measures of primary disorder. Since the maximum complexity of a sequence of length 32 is $C_{\max} = 32$ (realized if all of its symbols are different) the primary order of the above sequence is

$$\mathcal{O}_i = 32 - \mathcal{D}_i$$

where $\mathcal{D}_i$ are the values given in Table 1. Thus, even the most disorderly of the above sequences (no. 1) has a comparatively high relative primary order:

$$\mathcal{O}_{\text{prim,rel}} = \frac{C_{\max} - \mathcal{O}_{\text{prim}}}{C_{\max}} - \frac{32-19{\cdot}90}{32} = 0{\cdot}378,$$

this is simply due to the fact that it is a binary sequence. The most orderly of

TABLE 1

Characteristic	Sequence	Fine description	$C \leq$
1. Non-random, above average	x · · · x · · x x · x x · · x x x · x x · x · x · · x x · x · x	$S_2 \cdot \uparrow S_1^2 \cdot x^2 S_1 x \uparrow S_1 S_2^2 x,\ S_2 x S_2,\ x \cdot$	19·90
2. Random	· x · x x · x · x x x x · · · · · · x · x · · x · · x x x · · ·	$S^2 x^2 \cdot {}^6\uparrow Sx \cdot {}^2x^3 \cdot {}^3, \cdot x \cdot x^2$	16·94
3. "Semi" random†	x · x · x · x x · x x · · x x · · x x x · · x x x · · x x x x x	$S_1^3 x S_1 x \cdot {}^2x^2 S_2^3 x^2,\ x \cdot, \cdot {}^2x^3$	16·64
4. "Semi" random†	x x x · x x x x x x x · x · x · · · x · x · x x · · · · · · · ·	$x^3 S x^6 S^2 \cdot {}^2S^3 x \cdot {}^8, \cdot x$	14·24
5. "Semi" random†	· x · · x · x · · · · · · · · · x x x x x x x x x x · x · · x x	$SxS \cdot {}^7x^{10} Sx^2, \cdot x \cdot {}^2$	13·45
6. Complementary random	· · · x · · x · · x x · · · · x x x x · x x · x x · · x x x x ·	$S_1 \uparrow S_1,\ S_2 x^2 \cdot S_2^3 x, \cdot x$	13·08
7. "Semi" random†	· x · x x · x x · · · · · · · · x x x x x x x x x · · · · · x x	$\cdot x S^2 \cdot {}^8x^9 \cdot {}^4S, \cdot x^2$	13·06
8. Repetitive	x · · x · x x x · · · · x x · x x · · x · x x x · · · · x x · x	$S^2, x \cdot {}^2x \cdot x^3 \cdot {}^4x^2 \cdot x$	12·98
9. Symmetrical random	x · · x x x · · · · · · x x · x x · x x · · · · · · x x x · · x	$S \leftarrow S, x \cdot {}^2x^3 \cdot {}^6x^2 \cdot x$	12·85
10. Random doublet	· · x x · · x x x x · · x x · · x x x x x x x x · · · · · · · ·	$S^2 x^2 S^2 x^6 \cdot {}^8, \cdot {}^2x^2$	11·19
11. "Lawfull"	x x · x x · · x x · · · x x · · · · x x · · · · · x x · · · · ·	$\Sigma i = x^2 \| \cdot {}^{20}$	7·60
12. Repetitive	x x x x · · · · x x x x · · · · x x x x · · · · x x x x · · · ·	$S^4, x^4 \cdot {}^4$	5·81
13. Repetitive	x · x · x · x · x · x · x · x · x · x · x · x · x · x · x · x ·	$S^{16}, x \cdot$	5·20
14. Homogeneous	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	x^{32}	2·51

† The term "semi" random denotes that the sequences were generated by some mixture of randomness and law which, however, does not need to be specified here.

the above sequences (no. 14) has a relative primary order of:

$$o_{\text{prim,rel}} = \frac{32-2\cdot 51}{32} = 0\cdot 922,$$

which is the highest value possible in this case.

(B) COMPLEXITY AND ORDER VS. RANDOMNESS

The previous section has shown—perhaps surprisingly—that unorganized complexity or primary disorder does not assume a maximum for typical random sequences. In this section we shall investigate this phenomenon in more detail. We will study 11 ensembles of binary sequence, each 32 bits long, which have been generated by different mixtures of "randomness" and "law", i.e. the first bit of the sequences was determined by pure chance ($p = 0\cdot 5$ for 1 and 0) for all ensembles alike, whereas the probability of shifting to the alternative bit from one position to the next was changed from one ensemble to the next. Starting from the value one it was reduced to zero in steps of 0·1 from the first to the eleventh ensemble. In other words, ensemble no. 1 (probability of shifting = 1) and no 11 (probability of shifting = 0) were both highly constrained by law: the bits alternated regularly in all sequences of ensemble 1, and remained the same in all sequences of ensemble 11. The amount of chance vs. law increased from the first to the sixth ensemble (probability of shifting = 0·5, representing pure chance) and then decreased again. Ten randomly selected members of each ensemble are shown in Table 2. The entropies (a measure of randomness) of the 11 ensembles were determined as follows:

$$H = -p \log_2 p - (1-p) \log_2 (1-p),$$

where p is the probability of shifting.

The unorganized complexities of the ten sample sequences of each ensemble were determined, averaged, and plotted—together with the entropies—as a function of p (Fig. 1). Clearly, the average unorganized complexities are biased towards higher values of p, having a maximum (within the 11 ensembles studied) at $p = 0\cdot 7$. This corresponds precisely to the subjective notion of "disorderliness" as it was determined by asking 20 subjects (scientists and science students) to choose the ensemble they regarded as most "disorderly" (without any specification with regard to primary or secondary order). The result is shown in the staircase curve in Fig. 1. Thus, maximum (primary) disorder—both by the complexity measure and subjective evaluation—does not coincide with maximum randomness

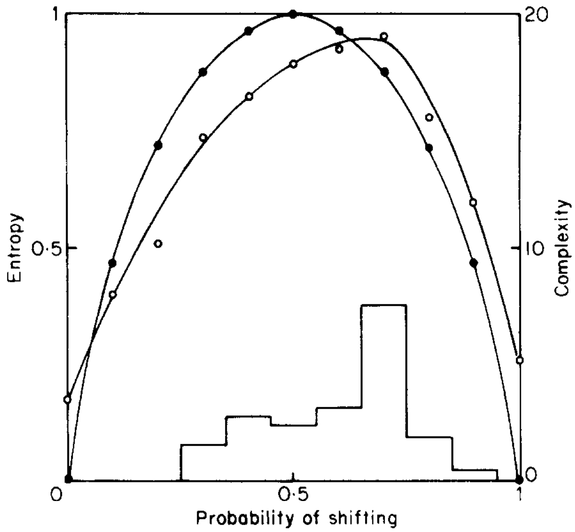


FIG. 1. Entropy (●) and complexity (○) as a function of the probability of shifting for the 11 ensembles studied (see text). The staircase curve represents the relative frequency at which the respective ensembles were regarded as "most disorderly" by a population of 20 scientists and science students.

(entropy). Interestingly enough Krüger (1979) gains a similar result with his measure of order as homogeneity and symmetry.

This result suggests that order and disorder seem to be judged by something resembling the complexity measure introduced in this paper—the difficulty of describing a pattern internally— rather than by a detector of randomness or its absence. Our psychological apparatus seems to be poorly fitted to detect randomness. It tends to see order or law where there is pure chance. Untrained subjects regard the comparatively long sequences of one's or zero's occurring in ensemble 6 as non-random or due to some kind of "law" and, therefore, considered ensemble 4 in which such sequence occur less frequently as more disorderly, even though it is, in fact, more lawful in the sense that the arrangement of its digits is less random. Our bias towards order also is revealed if we are asked to draw "random" sequences. What invariably appears are sequences with an unorganized complexity *greater* than that of the average random sequence. (The average for ten subjects—again scientists and science students—was 18·16, with 18·98 as the highest and 17·16 as the lowest value, whereas the average for random sequences is ≈16·84.)

TABLE 2

Eleven groups of ten binary sequences. The binary sequences of each group were generated by a certain mixture of "randomness" and "law" which differed from group to group (see text)

1	2	3	4
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

5	6	7	8
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]

9	22222111122222222112222211111111	10	12222212211111111111111111111111	11	11111111111111111111111111111111
	22111111122221122221111222111111		11111221122212222122222222222222		11111111111111111111111111111111
	22111111112222221111111222222111		11111111222222222222211111111111		22222222222222222222222222222222
	11111122122222222221222222211111		22222211122222222221112111111111		11111111111111111111111111111111
	11122111111111111111111111122222		22222222222222222222221111111111		11111111111111111111111111111111
	21112222211111112222222212222112		22222211122222222222121122222222		11111111111111111111111111111111
	22222221222222111112222222222222		11111111112111111111111111112222		22222222222222222222222222222222
	11222222222222221211111111111211		11111111111111111112221122111111		22222222222222222222222222222222
	22222222222211111222112211222211		22111111111111111111122222111222		22222222222222222222222222222222
	21111222111121122211111111111111		22221111111111111122222221111111		11111111111111111111111111111111

5. Summary and Conclusion

Starting with the observation that more complex or complicated patterns need more words for their description we introduced the length of the shortest possible descriptions—called the minimal descriptions—of these patterns in a language L as a measure of their complexity. Complexity may be differentiated into an organized aspect—the minimal description of the rules or laws underlying the formation of the pattern—and an unorganized aspect—that minimal description of the random aspects of the pattern. Unorganized complexity gave us a measure of what we called *primary disorder* and its complement—the deviation of a pattern from the (in the first approximation) totally random, chaotic state—a measure of its *primary order*. Organized complexity in turn gave us a measure for *secondary order*: the degree of organization inherent in a pattern. Whereas primary order corresponds to the inverse or complement of entropy, known as negentropy, redundancy, or even "information" (in a complete reversal of the meaning of the term originally introduced by Shannon), secondary order *may not* be associated with entropy or any of its derivations.

Primary and secondary order are in a way opposites of one another: the first decreases with (unorganized) complexity and the second increases with (organized) complexity. Both are, nevertheless, inherent in our cognitive apparatus as criteria for order, as the illustrations in section 2.c show. A distinction between these two kinds of order therefore appears necessary.

This distinction between primary and secondary order also allows us to dissolve one of the oldest controversies in evolutionary thinking: whether or not order increases or decreases during evolution. We suggest that this controversy is due to a confusion—non-distinction—of these two kinds of order.

A decrease in "order" means to the one camp—which we may call the "Boltzmannian" camp—an increase in entropy, and thus a decrease of primary order; an increase in "order" to the rivaling camp—which we may call the "Darwinian" camp (even though an increase in "order" in its sense of the word is also observed in areas other than biology)—means an increase in organization, and thus of secondary order. Not only is there no contradiction if we distinguish between primary and secondary order, but the increase in secondary order is seen to be directly dependent upon the decrease in primary order: increase in secondary order, such as atoms forming molecules, molecules forming cells, cells forming organisms, is, like any other process in nature, "driven" by an increase in entropy (second law of thermodynamics). We may even say that secondary order emerges out of—and at the expense of—primary order. Then, the most basic trends of evolution (of a sufficiently large part of the universe viewed for a sufficiently long time) may be summarized as follows:

$$\Delta o_{\text{prim}} \leq 0,$$

$$\Delta o_{\text{sec}} \geq 0,$$

whereas the direction of change of total order is questionable:

$$\Delta o = \Delta o_{\text{prim}} + \Delta o_{\text{sec}} \overset{?}{\lesseqgtr} 0.$$

However, it is certain that:

$$\Delta C = \Delta C_{\text{unorg}} + \Delta C_{\text{org}} \geq 0,$$

since

$$\Delta C_{\text{unorg}} \geq 0 \quad \textit{and} \quad C_{\text{org}} \geq 0.$$

In other words: it is complexity which always increases in evolution, be it unorganized or organized complexity. No such statement can be made about (total) order, because one kind of order (primary) tends to decrease, and the other (secondary) to increase.

In section 3 we laid out the steps for the calculation of complexity: modelling, coding, comprising, and evaluating. The difficulties involved—especially in the first three steps—are, indeed, formidable. Modelling involves the description of a concrete object by an abstract graph, and thus depends critically on what the model builder regards as essential and as non-essential.

The rough description (coding) of a model in turn depends critically on the symbols used for the rough description, and likewise the fine description (comprising) of a rough description on the symbol used for this process.

The selection of these symbols—the creation of an appropriate description language—is largely arbitrary. Moreover, even if one (or a few) such language(s) could be agreed upon as being optimal by the majority of scientists the task of finding the "shortest possible fine description" of a pattern in such a language—even though it exists—is probably beyond the capacity of even the fastest computers today, except for very simple patterns and description languages. Only upper approximations are feasible.

There is an irreducible subjectivity associated with the problem of complexity and the related problem of order. Other attempts to measure complexity and order—Chaitin's "length of program", Bresch's "pattern value", and Krüger's "homogeneity" and "symmetry",—are afflicted by exactly the same difficulties. If we want to make complexity and order more than just vague, intuitive notions at all, we have to face these difficulties.

Certainly, the procedures for calculating complexity, introduced in this paper also have to be regarded as mere suggestions, which seemed, however, natural to the author. Using the length of the minimal description as a measure for complexity has a strong appeal, because language is deeply ingrained in our way of thinking and evaluating and, therefore, really may represent the core of the problem of complexity. The empirical result on the relationship between randomness (entropy), complexity, and order reported in section 4.B supports this idea. The author hopes to show in further publications that his approach to complexity also yields meaningful results when it is applied to less abstract sets of patterns (see Papentin, 1980).

I am indebted to Dipl. math. M. Krüger and Dr. J. Cullum for carefully going through the manuscript and making many useful suggestions.

REFERENCES

ATLAN, H. (1968). *J. theor. Biol.* **21,** 45.
BLUM, H. F. (1968). Time's Arrow and Evolution. Princeton: Princeton University Press.
BOLTZMANN, L. (1877). *Wiener Berichte* **76,** 373.
BREMERMANN, H. (1967). In *Progress in Theoretical Biology I* (F. M. Snell ed.) New York: Academic Press.
BREMERMANN, H. (1974). In *Physics and Mathematics of the Nervous System* (M. Conrad, W. Güttinger and Dal Cin eds), Lecture Notes in Biomathematics 4, Berlin: Springer.
BRESCH, C. (1977). *Zwischenstufe Leben.* München: Piper & Co. Verlag.
BRESCH, C. (1979). *Zur Quantifizierung von Mustern* (preprint), University of Freiburg, Germany.
CHAITIN, G. (1966). *J. Assoc. Comput. Mach.* **13,** 547.
GATLIN, L. L. (1972). *Information Theory and the Living System.* New York: Columbia University Press.
KOLMOGOROFF, A. (1964). *Information Transmission* **1,** 3.
KRÜGER, M. (1979). Definition von Ordnungsmaßen auf Punktmengen, Dissertation (in progress), Univ. Bremen, Germany.
LWOFF, A. (1962). *Biological Order.* Cambridge: MIT Press.
MOROWITZ, H. J. (1955). *Bull. math. Biophys.* **17,** 81.
PAPENTIN, F. (1980). Die Naturwissenschaften **67,** 174.
VAN EMDEN, M. H. (1970). *An Analysis of Complexity.* Amsterdam: Mathematisch Centrum.
VON FOERSTER, H. (1960). In *Self-Organizing Systems* (M. C. Yovits & C. Scott, eds) New York: Academic Press.
RIEDL, R. (1975). *Die Ordnung des Lebendigen.* Hamburg: Paul Parey.
SAUNDERS, P. T. & HO M. W. (1976). *J. theor. Biol.* **63,** 375.
SCHRÖDINGER, E. (1951). *Was ist Leben?* München: Lehren.
SETLOW, R. B. & POLLARD, E. C. (1967). *Molecular Biophysics.* Palo Alto: Addison-Wesley.
SHANNON, C. E. & WEAVER, W. (1949). The *Mathematical Theory of Communication.* Urbana: University Illinois Press.

APPENDIX A

Relationship to Information Theory and Thermodynamics

We defined order and complexity as the length of the shortest possible description of a pattern. As such they must have something to do with information, because the information content of a non-redundant (as it is implied by the phrase "shortest possible") description will be an increasing function of its length. It would, therefore, be desirable to express complexity in terms of "bit", the unit of information, rather than in terms of the "length" of a description. However, this is in most cases exceedingly difficult, if not impossible, because the information content of the symbols constituting the description and the rules concerning the interpretation of this information are not generally known. If we want to express complexity in terms of bits at all, we will have to assume a more simplistic viewpoint on information—as is generally done in information theory (Shannon & Weaver, 1949)—and only consider the number, kinds, and relative frequencies of the symbols occurring in a description; in other words, we consider information only in so far as it exceeds a certain "background information" consisting of the meaning of the symbols and the rules for their interpretation.

The information content I of an "event" is defined by the amount of "choice" inherent in it, where choice is defined as:

$$I = \log_2 N, \tag{1}$$

where N is the number of possibilities available. The unit of information—one bit—is given by $N = 2$, i.e. by simple "yes–no" decision. The information content of a description consisting of n symbols falling into k classes, each represented n_j times, where $n = \sum_{j=1}^{k} n_j$, is then given by the choice inherent in selecting one of the possible permutations of these symbols, which is expressed by a well known formula from statistical mechanics, and information theory (Boltzmann, 1877; Shannon *et al.*, 1949);

$$I = \log_2\left(\frac{n!}{\prod_{j=1}^{k} n_j!}\right) = \log_2 \; n! - \sum_{j=1}^{k} \log_2 \; (n_j!) \tag{2}$$

For large n this expression may be approximated by:

$$I \approx -n \sum_{j=1}^{k} f_j \log_2 f_j = n\bar{I}, \tag{3}$$

where $f_j = n_j/n$ is the relative frequency of the jth kind of symbol, and $\bar{I}$ the average information content of the n symbols. For $n \to \infty$, $\bar{I}$ approaches Shannon's "entropy":

$$H = -\sum_{j=1}^{k} p_j \log_2 p_j \tag{4}$$

where p_j is the probability of occurrence of the jth kind of symbol. The information content of an individual symbol is then given by:

$$I_j = -\log_2 pj \tag{5}$$

The "information content" of a description may now be determined by determining the relative frequencies of the symbols occurring in it according to formula (2), or approximately according to formula (3).

Complexity and order in this way may be defined not as the length of the minimal description of a pattern but as the minimum amount of "information" inherent in it.

However, this is very cumbersome to do in practice. For all practical purposes it is sufficient to use the length-of-description approach. In this series of papers we will, therefore, only employ this approach.

It might be useful, however, to compare the length-of-description measure more closely with the physical entropy measure, since the latter is often used as a measure of order and even of complexity (van Emden, 1970). Information theory and thermodynamics are very closely related subjects. Without going into details, we may say that a more entropic system needs *on the average* more information for its description. As such entropy is, of course, closely related to the average unorganized complexity of an ensemble of patterns, as mentioned earlier. We would, therefore, expect the entropies and average unorganized complexities of a set of ensembles to be proportional, or at least to be increasing functions of one another. Thus, if entropy is taken as a measure of disorder we have to specify that it measures what we called *primary* disorder.

There is *nothing* in the entropy concept which measures secondary order. We can illustrate this once more by the example of a chemical compound: Entropy can give us a number for the randomness or chaos inherent in the ensemble of all possible conformations of a chemical compound at a given temperature and thus allow us to compare different chemicals with respect to this property, but it cannot give us a number which allows us to compare them with respect to their organization, i.e. their invariant binding relationships or structure. It is, therefore, *incorrect* to use the lack of entropy as a measure for organization. Certainly, any degree of organization entails a reduction of entropy with reference to the maximally entropic, random or chaotic state, but *the same amount of entropy reduction may correspond to* *very different degrees of organization.* If lack of entropy and organization were the same, a perfect crystal would possess the highest degree of organization, which is absurd, as pointed out by Saunders & Ho (1976).

Another weakness in the physical entropy concept is that it is bound to ensembles which—as a further limitation—have to be at equilibrium or close to it. An individual pattern, such as a sentence, a piece of art, or a machine, and a non-equilibrium structure, such as a Bernard cell, a candle flame or a biological organism, may not be discussed in terms of entropy at all. No such limitations exist for the length-of-description measure. Any set of structures, whether at equilibrium or not, physical or otherwise, may be compared with regard to their complexity and order as long as they are described in the same way (i.e. with the same degree of accuracy and by the same language).

APPENDIX B

Relationship to Formal Language Theory

A *formal language L* is defined as a subset of the set A_T^+ of all possible combinations of a finite set of symbols $A_T = \{a_1, a_2, \ldots, a_n\}$, called the *alphabet*—more precisely: the "terminal" alphabet—of the language:

$$L \subseteq A_T^+$$

Any element $w \in L$ is called a *word* of the language L.

One of the main themes of formal language theory is the classification of all possible languages according to their properties. One useful concept in doing so is that of a *generative grammar.* A generative grammar is generally displayed as a quadruplet.

$$G := \{A_T, A_N, S_o, R\},$$

where $A_T \neq \varnothing$ is the "terminal" alphabet, as specified above; $A_N \neq \varnothing$, with $A_N \cap A_T = \varnothing$, the "non-terminal" alphabet, i.e. a finite set of "metalinguistic" variables which themself stand for sequences of terminal symbols, combinations of terminal symbols with metalinguistic variables, or combinations of metalinguistic variables with themselves. $A = A_T \cup A_N$ is also called the joint alphabet. $S_0 \in A_N$ is a special metalinguistic variable, called the "start symbol"; and $R = \{(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)\}$ a finite set of ordered pairs, called "rules" which are also written in the form: $u_i \rightarrow v_i$ $(i = 1, \ldots, n)$, where $u_i, v_i \in A^+$, meaning that u_i may be substituted for by v_i.

The successive application of a certain number of rules, starting with S_o and leading to the word w_k, is called a *derivation* of w_k; in symbolic representation:

$$S_o \xrightarrow[R_1]{} w_1 \xrightarrow[R_2]{} w_2 \xrightarrow[R_3]{} \cdots \xrightarrow[R_k]{} w_k,$$

or for short:

$$S_o \xrightarrow[*]{} w_k,$$

where "$*$" denotes the successive application of certain rules $R_i \in R$.

A grammar where all rules have the form:

$$u \to v, \text{ with } u \in A_N \text{ and } v \in A^+,$$

is called a *context-free* grammar.

These grammars are of special interest to us, because finding the shortest possible description of a pattern can be viewed as finding the shortest possible derivation of the pattern in an appropriate context-free "pattern language" whose only word is the rough description of the pattern in question.

For example, let

$$w := x\ y\ z\ p\ q\ r\ s\ b\ f\ p\ q\ r\ s\ x\ y\ z\ g\ h\ x\ x\ y\ z$$

be the rough description of a pattern, then its shortest possible fine description—according to the rules laid out in the previous sections—is:

$$S_1\ S_2\ b\ f\ S_2\ S_1\ g\ h\ S_1,\ x\ y\ z,\ p\ q\ r\ s$$

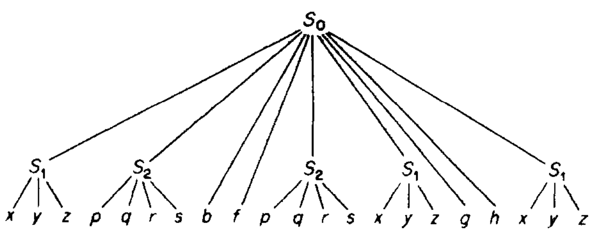
This fine description can be viewed as a set of context free rules:

$$S_0 \to S_1\ S_2\ b\ f\ S_2\ S_1\ g\ h\ S_1,$$

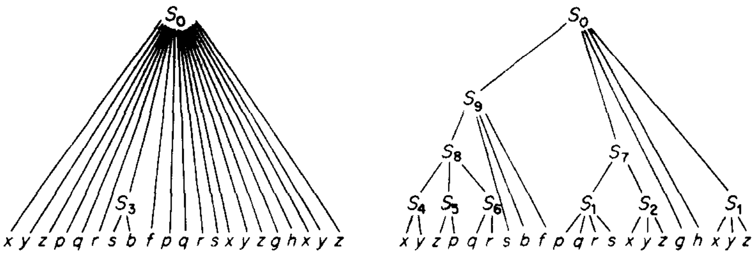
$$S_1 \to x\ y\ z,$$

$$S_2 \to p\ q\ r\ s.$$

These rules are capable of generating the rough description, i.e. starting from S_o we will obtain w by successive application of the rules. Such a "derivation" of w may be represented by a hierarchy (an inverted tree):



Of course, there are many more derivations of w conceivable. For example:



The problem of finding the shortest possible description of a pattern then amounts to finding the shortest possible derivation (i.e. the derivation involving the least total number of symbols) in an appropriately defined language.

Such a language—in the above example—should consist of the terminal alphabet containing at least all symbols occurring in w:

$$A_T = \{b, f, g, h, p, q, r, s, x, y, z\},$$

and a non-terminal alphabet containing all partitions (connected subsequences) of w:

$$A_N = Q(w),$$

where $Q(w)$ denotes the set of all partitions of w.

The system of rules R of the repective language should be declared by means of all possible "coverings" of a given subsequence $S_i \in A_N$, by subsequences $S_{ij} \in A_N$, $j = 1, \ldots, n$, such that

$$S_i = S_{i1} S_{i2} \quad \cdots \quad S_{in};$$

We write:

$$S_i \rightarrow S_{i1} S_{i2} \cdots S_{in}$$

denoting a rule of R.

In case S_{ij} consists of one terminal letter $a \in A_T$ only, we will agree to substitute S_{ij} for a.

In a language defined in this way, all conceivable derivations (corresponding to all conceivable decompositions of w into subsequences, see 3.3) may be tested and the shortest possible one selected. We define the "length" of a derivation in accord with the previous definition as the number of symbols to the right of the arrows of all rules used in the derivation (where each rule is counted only once even if it is applied more than once during the derivation), plus the number of commas in between the rules. The rules applied in the 3 examples of derivations given are:

(i)
$$S_0 \rightarrow S_1\ S_2\ b\ f\ S_2\ S_1\ g\ h\ S_1,$$
$$S_1 \rightarrow x\ y\ z,$$
$$S_2 \rightarrow p\ q\ r\ s,$$

yielding a total length of 18

(ii)
$$S_0 \rightarrow x\ y\ z\ p\ q\ r\ S_3\ f\ p\ q\ r\ s\ x\ y\ z\ g\ h\ x\ y\ z,$$
$$S_3 \rightarrow s\ b,$$

yielding a total length of 23.

(iii)
$$S_0 \rightarrow S_9\ S_7\ g\ h\ S_1,$$
$$S_1 \rightarrow x\ y\ z,$$
$$S_2 \rightarrow p\ q\ r\ s,$$
$$S_4 \rightarrow x\ y,$$
$$S_5 \rightarrow z\ p,$$
$$S_6 \rightarrow q\ r,$$
$$S_7 \rightarrow S_1\ S_2,$$
$$S_8 \rightarrow S_4\ S_5\ S_6,$$
$$S_9 \rightarrow S_8\ s\ b\ f,$$

yielding a total length of 35.

Thus the first derivation, yielding a length of 18, is the shortest possible—the complexity of the pattern in question.

By a simple extension of the notion of a generative grammar we may now completely embed our measure of complexity into formal language theory. We introduce a third finite alphabet of *functional signs*:

$$A_F = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ +,\ -,\ \Sigma,\ |,\ \uparrow,\ \leftarrow,\ \ldots\}$$

which may be used in formulating the rules of a grammar. The extended definition of a generative grammar is now as follows:

$$G := \{A_T,\ A_N,\ A_F,\ S_o,\ R\}.$$

All rules of a context-free grammar of this type have the form:

$$u \rightarrow v,$$

where $u \in A_N$ and $v \in A^+ = (A_T \cup A_N \cup A_F)^+$. The "length" of a derivation in such a language is determined—again applying the principle used previously—by counting the number of symbols occurring to the right of the arrows of the rules used in the derivation, except for numbers which are taken into account logarithmically, plus the number of commas between the rules.

We may now show that the complexity of a pattern is a well-defined, calculable quantity. We show firstly that for each rough description of a pattern there exists a shortest possible fine description whose length—which we will refer to as the "complexity" of the rough description—is calculable.

Proposition

If w is a word (the rough description of a pattern) of length n of a language L, based on the finite alphabets A_T, A_N, and A_F, with $A_N = Q(w)$; the complexity of w is effectively calculable.

Proof

We have to show that there exists an algorithm (most conveniently imagined as a computer program) which stops after a finite number of steps and has the complexity of w as its output.

In order to determine the complexity—the length of the shortest possible fine description—of a rough description w no fine descriptions longer than w have to be considered, since fine descriptions are by definition shorter than their respective rough description. As $A = A_T \cup A_N \cup A_F$ is finite, only finitely many symbol sequences have to be examined. One may proceed by first forming all singlets of the $a_i \in A' = A^+ \cup \{,\}$, then testing them to see whether they are capable of generating the rough description w—i.e. whether they are fine descriptions of w—or not; then by forming all doublets and testing them to see whether they are capable of generating w or not, then all triplets, etc.... The length of the first—or one of the first—multiplets capable of generating w is the complexity of w (since numbers are taken into account logarithmically, it might not be the first sequence encountered which is the "shortest" possible). In any case, maximally $Z = \sum_{i=1}^{n} k^i$ sequences—a finite number—have to be examined, where k is the cardinality of A'. The complexity of w may therefore be determined in finitely many steps, i.e. it is effectively calculable.

That the complexity of the underlying pattern is effectively calculable can be seen by an argument similar to the one used above: since a pattern (graph) consists by definition of finitely many objects and relationships and the number of symbols used to code such a pattern is finite as well, only finitely many rough descriptions may be constructed. As the shortest possible fine description of each of these rough descriptions may be determined in finitely many steps, the shortest of all these may be found in finitely many steps, i.e. the complexity of the pattern is effectively calculable.

The measure of complexity introduced here is not only useful in comparing patterns—the words of a language—but also in comparing *entire languages.* The complexity of context-free languages (only such were considered so far) may simply be defined as the *length of the shortest possible grammar* to generate these languages, where "length" again is defined as the number of symbols appearing to the right of all their rules, with the exception of numbers which are taken into account logarithmically, plus the number of commas separating the rules.

However, the notion of complexity may probably be extended easily to other types of languages as well. Every language is defined as a certain subset $L \subseteq A^+$. Rules are just a convenient way of describing this subset L concisely—if it can be done this way. Other languages may be described concisely by some other method. Some languages may be so irregular that they cannot be described in any other way than by listing their individual words (these are the most complex ones). In any case, the length of the minimal description of a language may be taken as a measure of its complexity. Since there is a one-to-one correspondence between languages and automata, the measure of complexity for the former may also be taken as a measure of complexity for the latter (cf. Bremermann, 1974).