AUTOMATA THEORY AND DEVELOPMENT, II
Morphogenesis of Simple Artificial Organisms

Richard Laing
Department of Computer and
Communication Sciences
The University of Michigan
Ann Arbor, Michigan, U.S.A.

and

Michael Arbib
Department of Computer and
Information Science
The University of Massachusetts
Amherst, Massachusetts, U.S.A.

Abstract

Although it is now clear that reproduction and other complex devel-
opmental processes can be modelled by mathematical automaton systems
(such as von Neumann's tessellation automata) such systems must be re-
focussed and augmented if they are to express clearly and directly more
of the reality of developmental processes. An automaton modelling system
is presented in this paper which permits the easy expression of biological
developmental hypotheses and the deduction of the consequences of those
hypotheses.

The modelling system has three principal parts: an *algorithmic
re-write system*, an *incarnation system*, and a *simplification system*.
In this paper the algorithmic rewrite system is described at length and
numerous examples of the consequences of particular developmental hypotheses
are given. Hypotheses are embodied in the modelling system by specifying
particular *initial biological components* (for example, cells, or tissues,
or organs) and particular *re-write rules* which say how a biological com-
ponent may be altered in the course of development, and a *control conven-
tion* which specifies the conditions under which the re-write rules may
be applied.

The incarnation system (which specifies the spatial properties of the
biological components) and the simplification system are discussed
briefly and in so far as they are necessary for the explication of the
example developmental hypotheses. (The incarnation system will be des-
cribed and discussed further in Part III of this paper.)

Automata Theory and Development: II

1. Introduction

The success of the digital computer has destroyed the belief that
complex chains of arithmetic calculation require the intervention of
some mysterious *mentalistic* force. In Part I of this paper (Arbib 1967)
we showed how the work of von Neumann (1966), Codd (1965), Thatcher (1964)
and Arbib (1966) could be placed in a biological framework, and thus
clinched the argument of von Neumann that reproduction was explicable as
a computational process without the intervention of some mysterious
vitalistic force. But just as the success of the computer at arithmetic
calculation has forced students of Artificial Intelligence to seek models
more akin to the reality of psychological processes, so must we - confident
that *a* process of development such as reproduction can be captured
in automata theory - now seek to evolve new theories which can capture
more of the reality of embryological processes.

If automata theory is to be a useful tool for the theoretical embry-
ologist, it is crucial that it be seen as an evolving discipline. It is
not enough to ask "How much of embryology can be captured in the language
of finite automata, Turing machines, and tessellations?" but we must
further ask "What new concepts must be injected into automata theory in
response to the need to capture essential properties of embryological
processes?" In Part I we emphasized answers to the first question; in the
present paper we shall also attempt to do justice to the latter.

Since Part I of this paper appeared, there has been a considerable
amount of further research bearing on the problems of formal approaches
to the study of biological development. This work includes (especially)

the papers by Lindenmayer (1971) and Hermann (1969, 1970) in this journal, in which string re-writing systems were employed in the modelling of developmental processes, and the paper by Laing (1970).

Related formal work includes the graph re-writing systems proposed by Pfaltz and Rosenfeld (1969) and Montanari (1970), and the tessellation automata and tessellation re-writing systems discussed by Smith (1971a, 1971b) and by Milgram and Rosenfeld (1970). Earlier work on nets of automata (Rosenstiehl, 1966) is also relevant here.

In this present paper we attempt to unify these approaches and techniques as we work toward providing a general framework for the modeling of biological (and other) developmental processes.

2. Goals for Models of Development

We begin by discussing what properties of model systems for development are desirable.

In order to be useful, a model system must be capable of *embodying* interesting, explanatory and predictive hypotheses about the biological systems it purports to represent. It is helpful if the model system permits the *precise* expression of theories of development, so that techniques of mathematical deduction can be employed. It is useful if such a model system also permits the *convenient generation* of the consequences of those theories. This means that the model must also make possible the comparison of its postulated consequences with the putative parallel empirical consequences of the (biological) systems under study.

In addition to the above very important properties, certain additional properties are also desirable; for example, the model system ought somehow to be so simple and direct   that investigator intuition

is not unduly inhibited by gratuitous complications and so that opportunity
for mathematical generalizing (i.e., statement and proving of theorems)
is thereby enhanced. In this paper we shall strive to meet the above
criteria, including the last mentioned one of simplicity and directness.

Unfortunately, in the course of designing a model system it is very
easy to obscure the relationship between the model and the empirical
system represented. Before commencing detailed presentation of our system
for modelling development, we here discuss briefly some problems of the
relationship between a model system and the empirical systems represented.

3. Relationships between Models and Empirical Systems

In modelling or simulating any process (in our case, the process of
biological development) we usually need not (indeed, generally *can not*)
provide a representational system capable of expressing *every* known
characteristic of the real process. We need of course model only those
characteristics of the real system which are relevant to the problem we
are studying. Often however we can not be certain, beforehand, of every
relevant characteristic of the system and we often therefore strive to
construct model systems which permit the representation of all those
characteristics which within reason, seem to bear on the problem.

In the course of its construction the model may also often come to
possess features not intended to be interpreted; that is, it will come to
possess features which though present in the model system, have no assign-
able counterpart in the empirical system under study. For example a
(mathematicized) model may come to include an apparatus of special symbols
or notations or punctuation introduced in order to assist the experimenter
in his understanding of the model and in the orderly and correct operation

of the model; this "bookkeeping" apparatus may have no simple or direct

relationship to the system under study. This non-interpretable apparatus

is even more striking and evident in the case of *simulation* (generating

one-by-one the temporal stages of a dynamic model); here (especially with

computer simulation) often the (programming) apparatus for running and

controlling the model will be considerably larger than the interpretable

part of the model itself. The interpretable part of the model we will call

the *model proper*.

The model proper usually has more flexibility than the real systems

to be represented, for the model must permit the expression and exploration

of several alternative possible explanations of the empirical evidence.

In expressing development, for example, the model should be flexible

enough to permit the generation of "abnormalities" (which of course are

just as "real" as normal development). Just as deformities give clues

to the understanding of biological development, so might we expect to

gain insight from the "monstrous" (or even biologically impossible)

arrays which a modelling system can produce, in addition to the balanced

or symmetric forms which are the likely product of "normal" genetic

potential under normal environmental conditions.

Thus systems for modelling biological development though they will

on the one hand be composed of abstractions and simplifications of some

features of real systems (both normal and "abnormal") on the other hand

may permit the expression of "pseudo-empirical" behaviors possibly not

exhibited in any known real system. Rarely do we achieve model systems

which permit in their models proper merely "all and only" the represen-

tation of some class of real systems under study. (Indeed, a model

developmental system which fitted its real systems so exactly that it

accounted for all and only the members of the real class of organisms, would be more a particular *theory* of development of the members of the class, than a general modelling system.)

Modelling systems tend, by design, to fall short by abstraction and simplification of features, and to over-shoot by the considerable speculative freedom permitted in the ways in which the features may be interrelated.

Since some particular interpretations of the developmental stages expressed in model may reflect closely the empirical systems under study and some not it is often useful to distinguish various degrees of fidelity with which the model proper might reflect real systems.

In employing artificial organism systems (as we will do here) in the modelling of development, we might ask only that the artificial organism system be able to produce the same ultimate forms as that possessed by the real organism system. On the other hand we can impose the very much more stringent requirement that an artificial organism system step-by-step mirror *every* significant developmental feature and *every* significant stage of development of the real organisms under study.

We might call the less stringent modelling (in which only the final outcome may correspond to the real system) a *weak* simulation, and the more stringent modelling (which feature by feature and stage by stage corresponds to the real system) a *strong* simulation. It should be clear that there are also many other kinds of simulation corresponding to many other kinds and degrees of fidelity that might be defined.

Of the two sorts of simulation focussed on above, a strong simulation seems clearly the more desirable goal; however the weaker forms of simulation can nevertheless still be extremely useful. Weak simulations are in effect alternative solutions to the same developmental problem. They

therefore implicitly pose the question: why does the real organism take
its particular course of development? This problem is especially inter-
esting when it appears that the real organism does not follow the most
efficient or elegant developmental route. (Such questions may of course
force us to examine the evolutionary history of the real organism.)

To summarize, a real system under study will generally be more complex
than its model; the model being first an abstraction of those characteristics
of interest, and secondly frequently given additional simplification for
both theoretical and practical reasons. Modelling systems however will
usually permit the exhibition of features beyond those possessed by the
real systems to be represented. The constraints imposed in the modelling
system so that the modelling system describes only a certain class of
real systems, constitute a *theory* of the real systems expressed in the
model system. In employing a model system we must distinguish sharply between
features of the model system intended to be interpretable in terms of the
systems represented and those which are merely ancillary "artifactual"
or computational features of the modelling or simulation process. We
may distinguish varying degrees of fidelity to the real systems; some
models represent the real systems strongly, some weakly; both notions
are useful.

[The mathematical foundations of modelling and simulation of systems
are discussed in Kalman, Falb, and Arbib (1969) and Zeigler and Weinberg
(1970).]

Finally let us note in passing that our long range goal is not
merely to model the development of form in individual organs or organisms
per se, but rather the understanding of the development of *functional*
systems which involve the whole organism. We have to think of the nervous

system, for example, not in terms of anatomically defined lumps of tissue, but rather in terms of an interacting overlapping collection of systems for carrying out biologically important functions. Thus, our task becomes even more complicated when we realize that it is not enough to look at one small part of the organism and explain how it grows, but we have to explain the sort of synchrony which allows functioning systems of various kinds to be available at birth and at later stages of maturation. At the moment, we hope to be able to look at one organ in a system and try to explain what sort of cellular interaction can give rise to its shaping We may hope that, later on, when we understand this, we will have the intellectual apparatus in place to combine together our models of several systems to understand what sort of overall synchronizing mechanisms allow the orderly coordination of their development.

We will next describe and discuss our proposed automaton system for modelling biological development.

4. A System for Modelling Development

Our proposed model system has three main parts;

(1) an *algorithmic* apparatus which operates on initially given developmental structures (composed of labelled points (standing for physical developmental components and their states) and lines connecting the points (standing for contiguity relationships among the components).

(2) an "*incarnation*" or "picturizing" apparatus which takes interpretable stages of the relational structures produced by the algorithmic apparatus and yields a "picture" of the physical-spatial properties of that model developmental stage. Thus, on the one hand (a) the incarnation

apparatus assists in the comparison between model consequences and the empirical systems purported to be represented, and on the other (b) permits the introduction into the relational structure (of the algorithmic apparatus) those revisions of elementary component contiguities which are a consequence of the changing spatial properties of the system under study.

(3) a *simplification* apparatus which provides techniques for the repeated introduction of new elementary components (e.g., aggregations of the previous elementary physical components) and an associated new algorithmic apparatus, which is to act upon the newly defined components.

The Algorithmic Apparatus. The algorithmic apparatus of our proposed modelling system for biological development is essentially a precise mechanism for generating representations of successive stages of growth of an organism. The algorithmic apparatus can be divided into two parts: (1) *re-write rules*, (2) a *control* which tells when, and where the individual re-write rules of (1) are to be employed.

The re-write rules operate on labelled relational structures consisting of individual elementary components (usually standing for organelles, cells, tissues, organs, individuals, etc. although they may also stand for components of the external environment). The re-write rules may be applied to any initially given relational structure of elementary components. An initial situation of especial interest for developmental studies is that of a single elementary component, standing for a single fertilized cell.

The components in the structure are re-written so as to change the state of components, to create or destroy components, or to create or destroy relationships among components. Two examples of re-write rules

would be: (1) ⓐ → ⓑ , (which says that a developmental component
in state a may be re-written so as to be in state b) or (2) ⓐ → ⓐ—ⓒ
(which says that a component in state a may be re-written as two connected
components, one in state a and one in state c).

In employing rules such as the above, in the representation of
developmental processes, we might for example be thinking in terms of
an interpretation of the symbols of the rules as denoting the states of
multi-cell *individuals*, (in which case the first rule might be taken as
representing a stage in the process of maturation and the second rule
as representing reproduction (such as budding-off); we might instead
interpret the symbols as the states of *organs* (in which case the first
rule might be taken as representing a stage in the development of that
organ and the second rule as representing the appearance of a new major
growth process on the organ). In this paper we shall very often be employ-
ing an interpretation of the symbols of our rules as denoting the states
of individual *biological cells*. Under this interpretation our first rule
above might be taken as representing a stage in cellular differentiation,
and the second rule as representing binary fission of a cell in which one
daughter cell is immature while the other is immediately prepared to
again divide.

Our two example rules produce *strings* of elementary components, but
multi-dimensional regular patterns of components (tessellations such as
those of von Neumann (1966) or Arbib (1967)) or even arbitrary graphical
structures are also produceable by such rules.

[Laing (1970) suggests the use of arbitrary graphical re-writing but
largely restricts himself to string examples; Lindenmayer (1971) restricts
himself to string production but in some cases includes in the string
ancillary symbols indicating how the strings may be re-written as other

structures, such as trees.]

A *set* of such re-write rules in the model defines all the potential developmental routes of the organism system. This full potential is narrowed if we require that the rules be used only in certain specified ways. This is the *control* part of our algorithmic system. The ways we allow the rules to be used can be interpreted in several ways: as a function of intrinsic control mechanisms of the organism, as a function of the impact of the extrinsic environment, or as some combination of these two.

If a system of re-write rules along with conventions of use of the rules retains some optionality (if the system is polygenic rather than monogenic) then the system will define not solely the course of development of a single individual but of a class. Thus, for organisms, this optionality could be interpreted as defining all possible phenotypic consequences of a single genotype.)

The properties of a system for modelling development will thus be affected by both the particular set of re-write rules permitted, and the *uses* of the rules permitted (in particular, specification of the time and place of application).

In order to attempt to generalize about what rules along with what rule-use conventions might be adequate to account for developmental processes, we might try to define whole classes of kinds of rules and conventions of rule uses. For example, we might consider rules which permit the re-writing of developmental components as a consequence of the states of those components alone (and not, for example, also taking into account the states of adjacent developmental components). Such kinds of rules (which we here call *autonomous-component rules*) might be contrasted with rules which *do* take into account the states of adjacent units. These latter we here

call *sensitive-component rules.* Other conditions may be imposed to
create other rule types.

We may also vary the way the defined rules may be employed. One
common convention is to allow *at most one* component at a time to be re-writ-
ten, employing freely any applicable rule for this purpose. (This is the
"traditional" convention in formal grammar theory.) A polar convention
requires that *every* component must be re-written "simultaneously" at
each step. (This is the "traditional" convention in most of cellular
automaton theory. Von Neumann (1966) employs this convention, as does
Lindenmayer (1971). Some flexibility can be gained by permitting compon-
ent re-write rules of the form x → x, so that the rule use convention
can be satisfied, while at the same time no change is produced at this
point.)

Other conventions might call for the ordering of the rules into a
list (and possibly also specifying an order in the developing entity)
and then imposing temporal priorities of rule application (and possibly
also priority of *place* of application). Still other conventions employ
computer-like programs to define what rules should next be employed.

(Different combinations of permissible re-write rule types and use
conventions can result in quite different models with quite different
powers; transferring results from one model system to another is therefore
not always straightforward. For example, employing autonomous-component
rules under the convention of simultaneous application of rules permits
the controlled production of organism symmetries which are quite impossible
with autonomous-component rules employed in the "one-rule-at-a-time"
convention.)

Although fundamentally all that is required of rule-use conventions,
is a clear, implementable statement of how the rules are to be used,

we will want the rule-use convention (as well as the rules themselves)
to reflect basic biological processes, known or postulated.

For example, the assumption of no structuring of the rule usage at
all will often define multiple outcomes, namely, the *full range* of develop-
mental potential inherent in the rules and initial conditions. On the
other hand, the explicit specifying (by a complete listing, say) of unique
time and place of rule application will define a *single course of develop-
ment,* while the assumption of simultaneous rule application at every
component at every stage would reflect the assumption of a uniformity
of (intrinsic and extrinsic) environmental impact on the course of development.

Other control conventions can define differential growth rates for
separate parts of the developing organism, can require obligatory rule
applications over-riding all other applications, can define whole ordered
rule sequences to be applied, can define conditions (e.g., of components
or of environment) which will change whe course of development by directing
the selection of one rule (or sequence of rules) over another.

It should also be noted that some components (either present initially
or produced by developmental rules) might be interpreted as serving as
"ports" for the entrance of external environmental influence. It would
then be "natural" to think of such components being "re-written" (differ-
entiating, or reacting) under the influence of uniform, or patterned, or
experimentally controlled environmental input, the changes in such "in-
put" components being transmitted (by repeated application of sensitive-com-
ponent rules) throughout the organism.

Indeed one way in which the effect of external environmental influence
can be incorporated formally into the proposed modelling system is to
permit the introduction (initially,or later - as a consequence of rule action)

of distinct spatial *environmental components* relationally associated with organism components (for example, associated with the components on the periphery of the organism). Just as with organism components, such environmental components would be subject to modification by appropriate re-write rules employed under some control convention.

(For a review of some of the results on alternative control conventions for re-write systems, see the papers by Rosenkrantz (1969), Fris (1968), Ginsburg and Spanier (1968), and especially Salomaa (1970).)

We have pointed out that a modelling system may go through stages which are not stages of development of the model proper; that is, they are not stages which are to be interpreted as standing for stages of development of the real system under study.

We can of course strive to set up algorithmic re-write systems having the property that we begin with an interpretable structure, and each re-write rule is so fashioned that its application always results in a new interpretable structure. This ideal situation may not always be possible, and we must therefore be prepared to introduce into our modelling system a way of distinguishing the interpretable from the non-interpretable stages of model development.

In the modelling of grammars, one way in which this is handled is by "terminal symbols". These terminal symbols may be produced by the re-write rules, but are themselves not subject to any further re-writing. When by their presence throughout the developing entity (in formal grammars, the entity would be a sentence form) no further re-writing is possible, an important stage in linguistic generation is thus signalled.

In modelling biological development we can also employ such methods of distinguishing important or interpretable stages from unimportant or uninterpretable stages of development in the model. That is, we can

designate certain sets of states of components as important, interpretable, or (perhaps, temporarily) "final", and so distinguish stages of interest from those which are merely by-products and way-stations of the modelling technique. It should also be mentioned that the control component of the modelling algorithm might embody conditions for specifying the important or interpretable stages of the modelling process.

In later pages, we shall employ various kinds of rules and rule conventions, and discuss their relevance in modelling biological development.

As we have pointed out above, specifying the precise rules and conventions of use of rules defines an algorithm (which may be monogenic or polygenic--in the first case there always being at most a single next step, where in the second case there may be alternative next steps and ultimate outcomes, thus defining a *set* of organism courses of developments).
This notion of an *algorithm* or *program* or *routine,* or *recipe* lies at the heart of our proposed modelling systems. The basic justification for this is that by this means many complex information-processings can be specified effectively in a way which is much more compact than any actual 'run' of the process. That is, an algorithm can provide a concise, economical, and revealing specification of a process. To see this, consider the following algorithm (in the form of a computer program) consisting of four instructions:

1. Set n equal to zero.

2. Print out n.

3. Replace n by n+1.

4. Return to the second instruction.

If you observe a computer executing this program, it will emit a stream of numbers which is endless - at least till you have exhausted the capacity of the computer.

Let us tie this point to our concern with embryology by seeing that the precise connections of an animal's nervous system can be specified far more economically than by explicitly listing for each pair of cells the manner and distribution of synaptic connections, if any, between them. Arguments that a comparison of the number of DNA bases with the number of connections in the brain shows that the brain must be a random network are as naive as comparing the four instructions of the above program with the number of positive integers and concluding that the sequence of positive integers, since it has more than four elements, must be a random sequence!

From programming computers then, we know the flexibility of programs having loops within them which are hierarchically structured to provide for a great deal of economy in the way we specify processes. As a biological example of a plausible "use" of such "nested subroutines", interneurons of the frog retina's second layer and the ganglion cells have schematized the dendrites of the ganglion cells as falling into two or three segregated layers. A plausible wiring scheme would then prescribe that certain types of axons from the interneurons terminated in one layer and so are highly likely to connect one level of the dendrites of the ganglion cells while other types of axons bearing different transforms of the visual input would terminate in the other layer thus hitting other parts of the ganglion cell dendrites. By this means, one can very simply specify how to get a retina that would function perfectly for the frog trying to snap flies in its world, without having to specify point-by-point interconnections. Hence, a sort of "nested subroutine" approach could probably explain a great deal of the specificity of the nervous system without requiring an immense investment in genetic material. In short, hierarchical structures of nested subroutines in genetic programs provide the right framework to

replace a one-gene one-character view and allow an economy of description without any imputation of randomness. Such economy of genetic description tion augurs well for economy of functional description when we come to describe organizational principles for neurophysiological processes. (For the organization of frog retino-ganglion connections, see DeLong and Sidman (1962)).

The Incarnation Apparatus. The algorithmic apparatus produces representations of connected aggregates of developmental components (such as, e.g., cells). Although the *states* of the components may carry all the required information as to the size and shape of the components, and the components' spatial relationship to neighboring components, the physical resemblance to the empirical systems being studied may be only very weakly apparent (or even indiscernible).

A function which allows us to compare the status of predictions in the model, with putatively parallel empirical consequences is of crucial importance in making the model work and provide explanation.
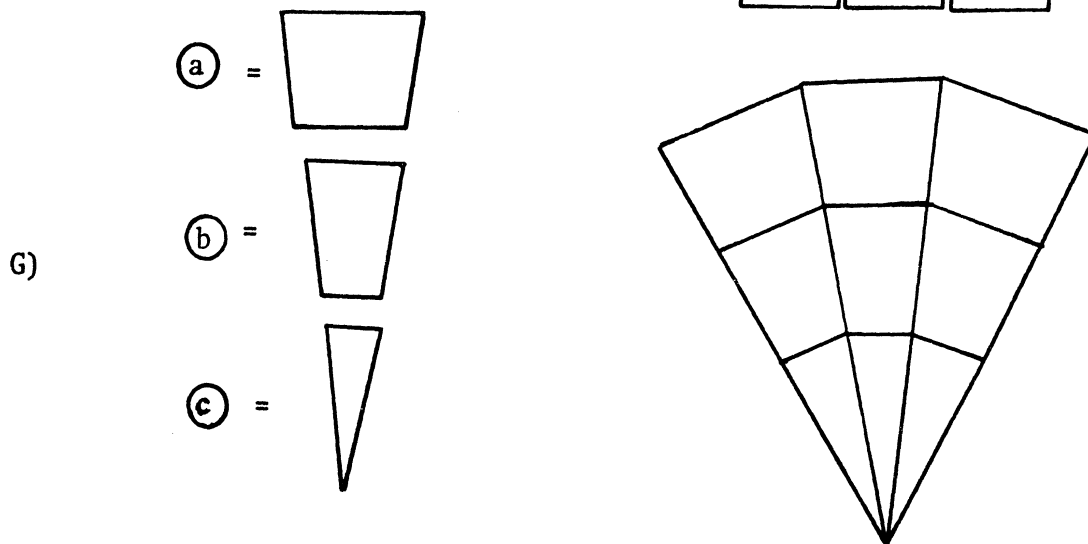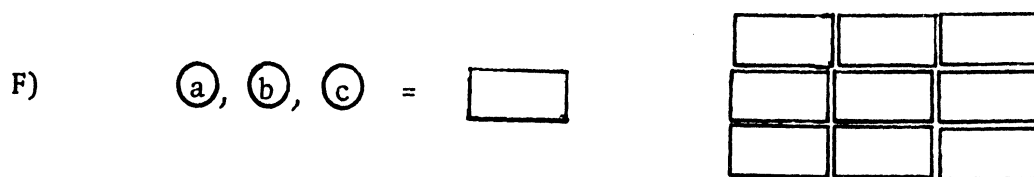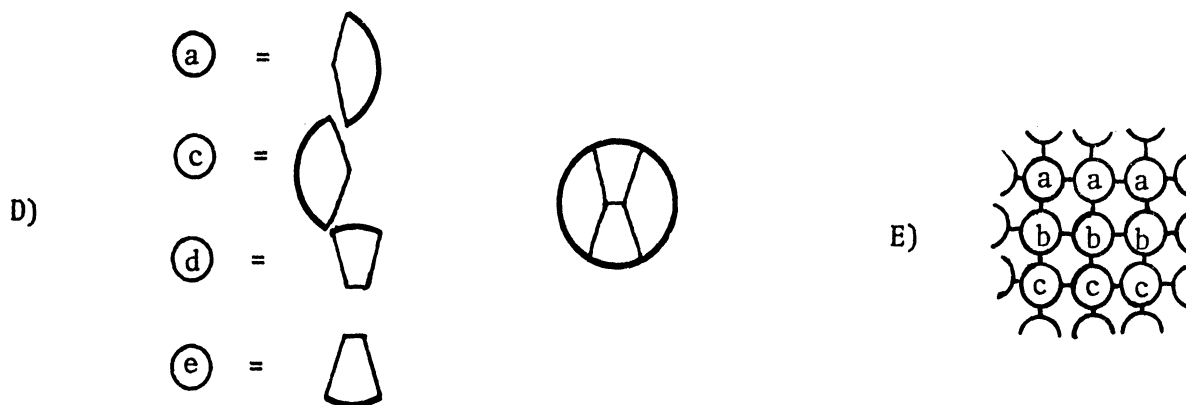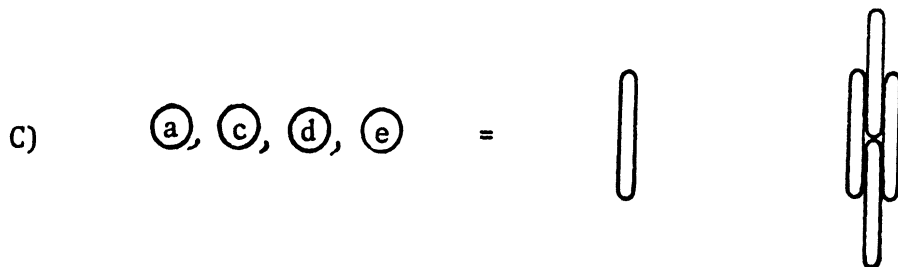
In some cases the required association is quite simple and direct, and the notion of a separate function relating the model realm and the empirical realm does not even explicitly arise.

In other cases, a distinct apparatus for displaying the spatial consequences of the algorithmic apparatus is extremely useful. Indeed, there will often be cases where the "skeleton" graphical structure produced by the algorithmic apparatus can give rise to markedly different three-dimensional spatial forms depending on the labelling and interpretation of the states of components. An example is given in Figure 1.

In many cases the incarnation function need be applied but once;

Figure 1: The algorithmic re-write apparatus of A), if given a single
           initial a-state component will permit the production of the
           labelled relational structure B). C) and D) (by assigning
           different spatial properties to the components) show two
           possible alternative spatial consequences of the relational
           structure of B).
               In E), the developing structure, shown embedded in
           a tessellation space, could have many alternative spatial
           interpretations, among them F) and G).

A)

(a) → (a)—(b)

(b) →)—(b)—(c)

(b) → [diagram with d on top, e on bottom]

B)

[diagram: d top, a left, c right, e bottom]

C)

(a), (c), (d), (e) =

D)

(a) =

(c) =

(d) =

(e) =

E)

[grid with a a a / b b b / c c c]

F)

(a), (b), (c) =

G)

(a) =

(b) =

(c) =
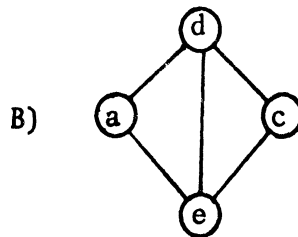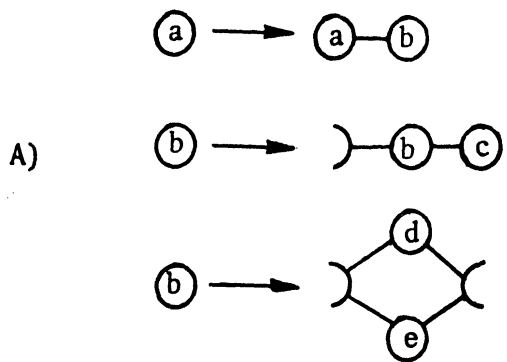
that is, at the completion of the process of generating the ultimate consequences of theory in the model. In the developmental modelling system we propose here however, it may be necessary to apply an incarnation function repeatedly in the course of calculating intermediate interpretable stages of the model.

Although we usually need not apply the incarnation function after every application of the re-write rules, we must, (especially in model systems in which the rules are sensitive both to the states of components and the states of contiguous neighbors) be prepared to apply the incarnation function frequently enough always to detect any spatial changes which could cause different (contiguity sensitive) re-write rules to be brought into operation.

For example, re-write rules applied repeatedly might step-by-step change the states of components (e.g., cells in a strip of tissue) so that use of the incarnation function would reveal that the tissue had rolled up into a cylinder, bringing many previously distant cells into contiguity. The incarnation function would have to be applied frequently enough to detect the newly arisen contiguity relations, so that the re-write component of the model could take them into account.

Thus, in the system for modelling development we propose, an incarnation function which not only gives us an understandable picture of the model consequences and its relation to the empirical data but also serves to introduce modifications into the algorithmic apparatus as required.

[This aspect of our model will be further discussed in Part III of this paper.]

The Simplification Apparatus. As long as the organisms whose developmental processes we are observing and whose outcome we are calculating have

only a few structural components (for example, if the organisms consist of only a few scores of cells) it is possible that our algorithmic apparatus would yield dynamic stages of growth which are clearly "graspable" by the mind. At low levels of combinatorial complexity, the effect of repeated application of various sorts of local rules on global form can often be intuited. We may even possess or acquire enough analytic insight into the processes to enable us to express and prove generalizations (theorems) about rules and their processes and their consequences. By such means we often can extend our intellectual grasp beyond what can be seen immediately and directly. Usually however our analytic insight begins to fade swiftly in the face of increasing combinatorial complexity.

Analysis of organisms consisting of several thousand cells (or other basic components) usually places a severe burden on our intuition and it often becomes difficult even with computer simulation and electronic visual display facilities, to "see" what the effect of local developmental rules is. That is, it becomes difficult to grasp *significance*, as opposed to merely observing that changes are taking place. When this happens, it is a clear signal that if our modelling system is to continue to be explanatory it must be simplified.

We have been considering rules of development whose immediate range of communication and effect is a single component (usually, in our examples, a single cell) and its immediately contiguous neighbors. That is we have been considering rules whose "neighborhood" (region of communication with other units) is of fixed, finite (and rather small) scope. Computationally speaking there is nothing inherently limited in such rules. Indeed there are theorems to the effect that any effective computational process can be expressed in such systems. If the system ceases to communicate to

us any sense of what is taking place developmentally, it is not an inherent
failing of the system, but rather a sign of the limitations of the powers
of the human mind in the face of great combinatorial complexity. If we
wish, then, to understand complex developmental processes, our model must
undergo simplification.

In simplifying there are several techniques we may employ. Suppose
we have a small aggregate of some specific number of structural components.
For each component there is a rule of development applicable. If we wish
to express the next stage of development of this clump of components, we
apply our "micro" rules, one-component-at-a-time. It might be possible
however to define a new "macro" rule which takes the original clump, and
in a single rule application gives us the next state of the entire clump.
If the "macro" is obtained by merely compounding exactly what the sequen-
tial piecemeal application of the original rules would have produced any-
way, then such macro rules are clearly a legitimate simplification of the
rule system. If any criticism is levelled against such a macro rule (as
perhaps leading to false conclusions) reference can always be made to the
micro rules of which the macro rule was compounded.

In simplifying our system, we may wish to search our developing
organism model for other such "clumpable" aggregates of components. We
may even wish to convert the entire developing system into a number of
such clumped entities, not concerning ourselves further with the histories
of the original individual components, but only of the new individuals
consisting of clumps of the old components. In such a case, we might thereby
be purchasing rule simplification (and thus explanation) at the cost of
loss of information about the fate of individual original components.

Notice also that (depending on the way the original micro rules were
to be applied) we may also simplify the time-scale (and thus the number of

separate developmental stages we are able to express). That is, some macro

rules may by such as to cause the system to reflect *every stage* of devel-

opment that was reflected in the original micro system, while other macro

rules might operate so as to omit some stages of development. The micro

rule time scale and a macro rule time scale may differ. Clearly this sim-

plification (coarsening) of time scale will result in a system simplification.

Thus, just as we must be aware that in our spatial lumping (our lumping

of cell-components) we may lose micro size physical properties of interest,

we must be aware that in our temporal lumping we may be omitting develop-

mental stages of interest.

As an example, let us consider a small fixed aggregate of cells which

behaves in the following fashion: if one of the cells in the aggregate

receives a stimulus from the outside, then, by means of inter-cell commun-

ication within the aggregate, there is a later time (a fixed number k of

steps later) in which all the cells in the aggregate simultaneously "diff-

ferentiate" thereby assuming some particular state or form. Clearly we

can define a macro rule which operates upon the clump at time n and at

time n+k produces the differentiated aggregate. Such a macro will express

precisely the fate of the cell aggregate; such a rule will not however tell

us anything about the stage-by-stage development of the separate cells of

the aggregate. If it is this micro-time, micro-component information we

seek, then clearly the macro rule omits it; if however, we are trying to

understand the more global course of development, then such macro rules

may be very valuable tools.

In our examples of simplification we so far have limited ourselves

to cases where the new macro re-writing rule was applicable only to a

fixed finite number of cells. In such cases, several micro rules might

be reduced to a single macro rule and in turn the new macro rules combined to form yet higher level macros. By this means considerable system simplification can be achieved.

Let us now consider a more complex issue.

Suppose (as in our last example) that an aggregate of cells is such that if one of its members is stimulated in a certain fashion then after a certain amount of time all the cells of the aggregate assume the same differentiated state. Contrary to the earlier example however, the number of cells in the aggregate is not *fixed*. Thus any macro rule whose condition for application is the state of a fixed finite aggregate of cells (a rule of fixed finite scope) will be insufficient. Indeed because of (in general) the indefinite size of cell aggregates, either an indefinite number of finite scope rule applications will be required to handle all the cases which might arise, or (since *all* of our re-writing rules have been of the finite scope type) instead of achieving instant simplification of our rule system, we may be forced, *ad hoc* to augment or revise indefinitely (and trivially) the set of developmental rules, so as to accomodate each larger aggregate which may arise. Thus, our very attempt to simplify our system by substituting macro rules operating on cell clumps for micro rules operating on individual cells, may be foiled.

One solution to this is to employ a new form of re-writing rule: the *transformation rule*.

A transformation rule need not operate over a fixed finite region of the developing organism. Let us explicate further by means of an example. Let us suppose that in a developing organism there is a region of cells which is producing some hormonal substance. This substance produced at site A will bring about changes in distant region B of the organism. Intervening between

regions A and B there is a region C, through which the hormonal material can be transported, but (for purposes of this example) is not itself otherwise affected by the hormonal substance. Let us also assume that Regions A, B, and C have an indefinitely large number of cells. (That is, although at any moment these regions are composed of a finite number of cells it is not reasonable to assign any specific bound to the number of cells they may be composed of.)

Employing our earlier "immediate cell neighborhood rules", we can represent the cell-by-cell transmission of a hormonal substance arising in the cells of A, passing through the intervening cells of C, and causing changes in the cells of B. If the numbers of cells of A, B, and C are large, then step-by-step cell-by-cell changes (even though precisely modelling real organism changes) may be difficult to follow, and their significance obscured. If we successively increase the scope of the rules (while each time keeping that scope fixed and finite) we will only superficially improve the situation. If the organism with its A, B, and C regions can be indefinitely large, then we will be adding new rules (and new complexity) to our descriptive system. Such new rules will be no more enlightening than the old rules, and, since the organism can be indefinitely large, new rules of yet larger scope may have continually to be introduced.

What is required is a rule (a transformation) which will operate on the *entire regions* A, B, and C, and will, from the state of region A at time t assign the state of region B at time t+k. Such a transformation rule clearly cannot be limited to affecting a fixed finite number of organism cells. It must operate on the regions regardless of sizes and distances involved.

This brings into focus a critical problem of this simplification

process (and of many simulation and modelling simplification processes):

how do we determine and define aggregable entities such as A, B, and C

which are to become the new basic components manipulated by our revised

and augmented re-write and transformation rules? In particular how do we

find the "natural" multicell aggregates which become the basic units to be

manipulated by "higher" level systems of rules" A technique of automatic

hierarchy generation in developmental theory would be very useful, but

at present does not exist. One very suggestive prospective element in such

a theory is to consider as candidates for aggregation cells which have

elements of common history (a technique which has long been employed by

developmental biologists in distinguishing biological structures). One

such common element would be possession of the same ancestor cell.

This is illustrated in the derivation tree shown in Figure 2.

Of course, cells of common ancestry need not be the only guide to

aggregability.

[The ideas on transformations and aggregability criteria by derivation

history are of course adapted from aspects of the contemporary linguistic

theory of N. Chomsky (1963 , 1965). It should also be pointed out that in

part our finite scope rules parallel the grammarian's constitutent structure

or phrase-structure rules, while our incarnation function is related to

the linguist's lexical or phonological insertion, and our simplification

transformations parallel Chomsky's grammatical transformations. There

are differences however, between the biological-developmental and the

grammatical situations, which are worth bringing out. Among the very strong

reasons for introducing transformations into grammatical theory is that

there may be relationships between indefinitely separated parts of sentence

forms and these parts determine new sentence forms. The new forms usually

cannot conveniently be produced by constitutent structure rules alone. For
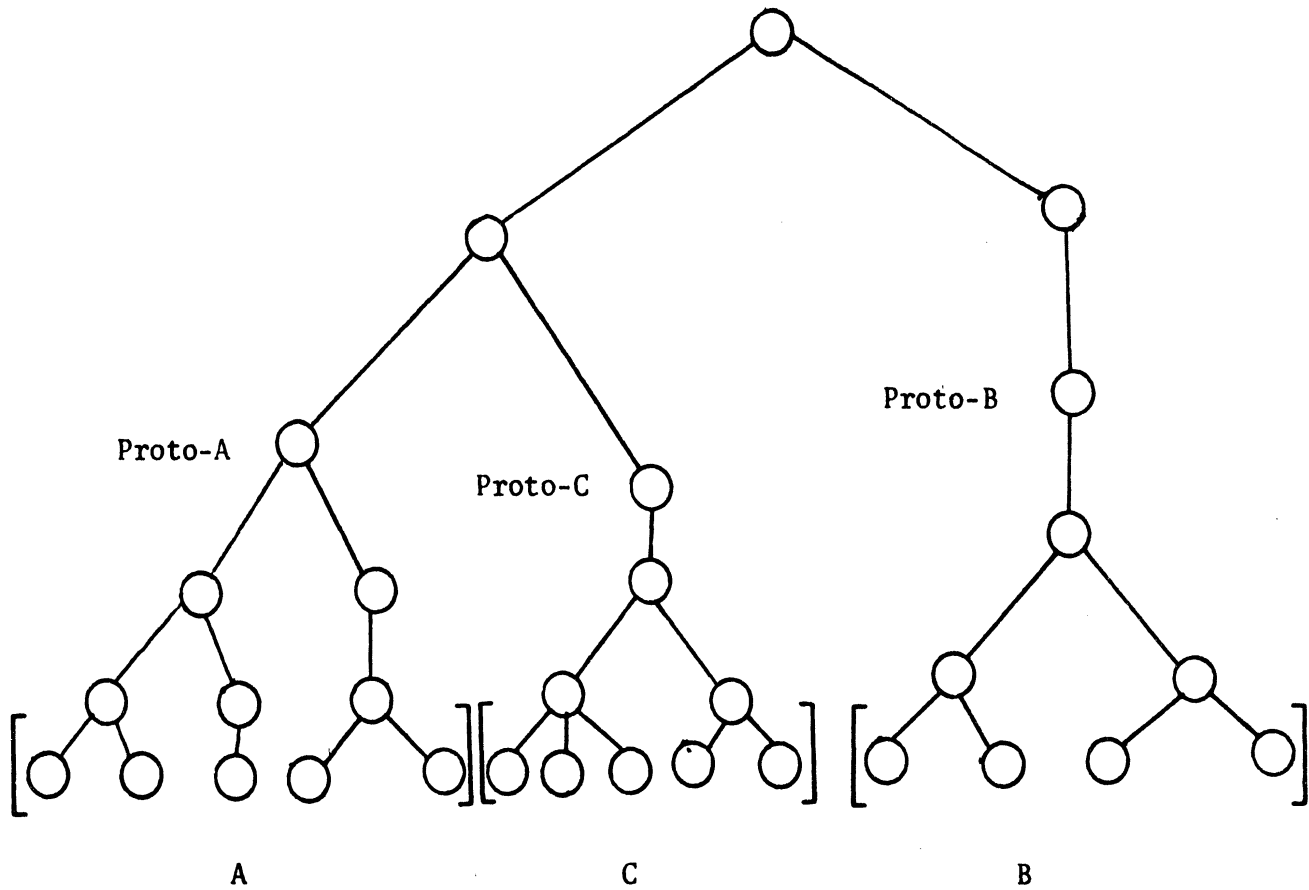
Figure 2: Illustration of possible determination of aggregable structures by
means of ancestry. In the general case, there might be an indefin-
ite number of rule applications between an "ancestor" cell and
the structure to be aggregated, so that a structure might be
indefinitely large, or indefinitely far from some other structure
affecting it.

what would seem to be required is some sort of communication-interaction between distantly separated sentence elements, and there is no reason to view the intervening sentence elements as a communicating medium transmitting the interactions. In the biological case the absolute necessity of transformation rules is not, on the face of it, so compelling, because here there is a communicating medium, albeit of indefinite extent, and phrase-structure rules could, piecemeal, transmit the desired effect. A clear reason for their adoption is, as we have indicated, their simplifying-explanatory power (which is also a compelling reason for their use in grammatical theory) as we try to define new combined biological units and to define the nature of the interaction of these units.]

5. Discussion and Formal Results

We have outlined our modelling system for development above. The system has great flexibility and, as we have pointed out can serve as the representing vehicle for "impossible" as well as "real" organisms and organism courses of development.

We now wish to discuss which among the formally definable possible accountings for development expressible in our modelling system, might best explain real biological development. That is, which *theories* of development expressible are good, fruitful, revealing, informative, and which are not.

It is clear that some theories, expressible in this system, of how development takes place, are clearly inadequate. For example, one theory might declare that all development is properly accounted for by employing only two re-write rules: 1) a *binary fission* developmental rule, applicable to any cell, in which the original cell is split into two daughter

cells; these in turn are to be connected to each other and to all the cells

the original cell was connected to, and 2) a *connection deletion* rule

applicable to any connection between cells, so as to remove that connection.

With this pair of rules any graphical cell structure can be formed. For

example, to create the cell structure of a hippopotamus, we start with a

single initial cell and employ the binary fission rule to produce a

completely connected mass of n cells (where n is at least as large as the

number of cells in the hippopotamus we desire; some simple calculations

on hippo volume and minimum cell size, plus a generous safety factor

will yield n, the number of cells required). The growth rule applied at

least n-1 times will yield the desired cell mass. We next (employing

the connection deletion rule) selectively disconnect all non-hippo cell-con-

tiguity relationships and also setting aside any excess non-hippo material.

The hippopotamus cell structure will gradually unfold.

This hippopotamus recipe is on the face of it absurd, but let us

take the time to be very explicit about some of its failings. First of

all, in the above "theory" of development, the time and place of appli-

cation of each developmental rule is completely controlled from the outside.

In effect, a highly structured and intrusive "environment" (the clever

hand and mind of the hippo designer) controls each step of organism develop-

ment, while the organism proper plays no important part in the process.

But in any adequate theory of development we take it as essential that

development as an interaction between organism and environment be expressible,

and that in most interesting and revealing cases the rules of development

of the organism itself will exercise considerable control over the course

of development.

Secondly, in the above scheme, the developmental process (as outlined)

does not step-by-step mirror the significant stages of development of a

real organism. That is, it is not a "strong representation" of the developmental processes of an organism. (Of course, as we have pointed out earlier, it may be useful or informative at times to produce a "*weak* representation" of the development of an organism, but here even the weak representation embodied is really a vast class of alternative courses of development.)

We therefore seek theories which of themselves contain the features which explain how organisms might be formed.

Such theories of development can be expressed in our modelling system. (For simplicity in explicating our ideas we will frequently restrict our examples to *linear* structures even though in most cases the results are applicable to more complex graphical structures of Physical components.) That *any* /(fixed, finite, linear) array of cells can be produced by an autonomous-component (context-free) system is an immediate consequence of the following lemma:

LEMMA 1: *For every string* $a_0 a_1 \cdots a_{n-1}$ *there exists a context-free grammar (autonomous-component organism system) with at most* $2^{[\log_2 n]}$ *nonterminals which produces it.*

Proof: Let $M = [\log_2 N]$. We introduce one nonterminal $\overline{\theta}$ for each binary string of length j, $0 \le j \le M$. Let $\tilde{\theta}$ be the number whose binary expansion is θ. Then a suitable grammar has initial symbol $\overline{\Lambda}$ and productions

$$\overline{\theta} \to \overline{\theta 0}\ \overline{\theta 1} \quad \text{for} \quad \ell(\theta) \ < M\text{-}1$$

and

$$\theta \to a_{\tilde{\theta 0}}\ a_{\tilde{\theta 1}} \quad \text{for} \quad \ell(\theta) \ = M\text{-}1$$

it being understood that $a_k = \Lambda$ for $k > n$.

EXAMPLE: The string $a_0 a_1 a_2 a_3 a_4 a_5$ would be generated as follows:

$$\overline{\Lambda}$$

$$\overline{0} \quad \overline{1}$$

$$\overline{00} \quad \overline{01} \quad \overline{10} \quad \overline{11}$$

$$a_0 a_1 \quad a_2 a_3 \quad a_4 a_5$$

Thus, we can grow *any* pattern of cells if we allow ourselves an indefinitely large number of component states and corresponding rules.

In this scheme every cell, save those in the final stage, bears within it its complete history - the prefixes of its state tell it what cells it is descended from, and the complete string tells the cell its precise coordinate in the linear (or arbitrary) cellular array. Thus, when the cell stops dividing it can unambiguously "interpret" this "positional information" (cf., Wolpert, 1969) to differentiate into epidermis, or patella, or what have you.

Note also that an even more elementary system of autonomous component productions, employing only *restricted* autonomous productions, could also yield any particular array of cells. (In *restricted* productions, when a cell divides in two, one of the resulting pair of cells is no longer capable of further division.) Using restricted rules only, any particular specified linear array can be produced, moving from left to right (or alternatively from right to left). The following example set of example rules should make the process clear.

$$A_0 \rightarrow B_0 A_1$$

$$A_1 \rightarrow B_1 A_2$$

$$A_2 \rightarrow B_2 A_3$$

$$\vdots$$

$$A_n \rightarrow B_n$$

At each step a new B cell is produced and the "master" A cell goes into a new state which defines the production of the next cell to be produced. The master cell can thus have "packed up" inside it, in a very simple fashion, (a backwards and forwards deterministic loop-free path of states) complete instructions for the growth of any particular (in the example, linear) array of cells.

Although by the use of the above kinds of production rule systems we can produce any particular finite organism array whatsoever, such a scheme clearly is not completely satisfactory in accounting for the development of organisms. Where our "hippopotamus" growth was completely under the control of the environment, here the organism development is completely determined by the system itself. This in itself is welcome, but unfortunately it does not express the complex organism-environment interaction we know exists. Such systems of complete organism specificity of the process of development undoubtedly have a role in explanations of development; they may be useful in the description of some particular developmental sub-processes; they cannot however be the *general* mechanisms either actually employed in development nor can they, in this form, serve as adequate basis for a useful explanatory or predictive model of development. In these particular systems, the final organism cell organisation must be known beforehand and the production system would have to be designed to "spew out" the pre-recorded form; in general nothing is revealed of the real processes of development by which a single cell becomes an organized multicellular organism.

Also, in such systems, the numbers of re-writable symbols must be in the same range as the numbers of cells in the entire completed organism. If the final organism form has billions of cells then there must be a

comparable number of re-writable symbols (and rules for them). The complexity of the developmental system becomes thus purely a function of the *numbers of cells* of an organism, rather than of the *complexity* of the structure of the organism. For any very large organism the sheer size of the production system required to model it, may make it impossible for the mind to grasp, and thus useless as an explanatory model of development.

In addition, there are other very particular objections to this use of production rule systems for explaining development. This scheme cannot be generally applicable, for example, since if we take a cell from a 4-cell sea urchin embryo, we know it may develop into a complete, though small, animal, not into a quarter urchin as the above scheme (the process explicated in Lemma 1) would imply. That is, such a developmental scheme cannot account for specific known biological developmental phenomena.

We may of course not limit ourselves to accounting for biological development of form by this simple-minded "absolute-addressing" use of autonomous-component rules.

Autonomous-component systems (with context-free rules of growth and development) can be made considerably more sophisticated. Unfortunately, there is still a very definite limit to the complexity of structure produceable by autonomous-component systems, as a lemma of Bar-Hillel, Perles and Shamir (1961) shows.

LEMMA 2: Let p be the length of the longest string that can be derived in a context-free grammar G using a tree of height at most n, the number of non-terminal symbols in G. Then there exists a number $q \leq p$ such that every sentence z of L(G) with (z) $\geq$ p can be decomposed into the form

such that
$$z = xuwvy \quad \text{with } u \neq \Lambda, \; v \neq \Lambda \text{ and } \ell(uwv) \leq q$$
$$z_k = xu^k wv^k y \quad \text{is in } L(G) \quad \text{for } k = 1,2,\ldots$$

In effect, what the Bar-Hillel *et al.* result says for biology is that for an algorithmic re-write system (in which the use of the rules is not structured by the environment) the complexity of structures produced by context-free rules alone is still closely bound to the numbers of significant cell states of the developing organism. Thus, there can be no very compact algorithms for development if only autonomous-component re-write rules are employed.

This suggests the following:

Observation: When the number of different developmentally significant states of components of an organism structure is much less than the number of components in the structure, only limited organism structures (e.g., simple tissue) can be grown without *inter-cell communication.*

There are also some sound empirical biological reasons which will preclude the use of autonomous-component systems as models for organism development in general. For example, embryos often develop correctly despite damage; this means that cells differentiate not simply on the basis of their individual past histories (as in autonomous-component systems) but also on the basis of information on the states of their neighboring components.

Thus since communication between components is necessary (in models of regulating or plastic systems) we have the important observation:

LEMMA 3: *Autonomous-component systems (context-free grammars) are inadequate for representing all developmental phenomena.*

Systems in which the developmental rules are applicable to components as a function not only of the internal states of the components but also the states of neighbors of the components we will call *sensitive-component* systems. Although generally in sensitive-component systems the neighbors of a particular component may be components at any fixed finite distance from the given component, we will here usually employ a neighborhood consisting only of the immediately contiguous components.

[It should be noted that the tessellation structures of von Neumann and our Part I (Arbib, 1967) can be seen as a special case of two-dimensional sensitive component systems. The reader may find it a worthwhile exercise to replace our rules there for state-change in a fixed array of cells, all but finitely many of which are quiescent, by rewriting rules acting on a finite symbol array corresponding to the active part of the tessellation. Of course, the neighborhood dependence of our state transitions implies that the resulting rewriting rules will not be context-free. Sumarizing, then we have the:

FACT: Template-based tessellation arrays are a special case of a sensitive-component multidimensional rewriting scheme.]

Sensitive-component systems are sometimes restricted so as to exclude regressive rules of growth. Such *restricted* sensitive-component systems parallel the *context-sensitive* grammars of the linguist. We will call such restricted sensitive-component systems, *monotonic-systems* (since their developing structures never decrease in size).

Are such monotonic systems adequate to express all significant developmental phenomena? It is known (from results in formal grammar theory) that (the parallel) context-sensitive systems have expressive powers not possessed by autonomous-component (context-free) systems. Yet there are clearly cases where monotonic systems are not adequate. Monotonic systems,

since they lack "regressive" rules, have no means of expressing the dissolu-
tion of cells or cell connections. But in biological development cells
are released (blood cells, germ cells), cells are dissolved (in shaping
of limbs) etc. .

LEMMA 4: Monotonic systems will in some cases be inadequate to express
all developmental phenomena; in some cases general sensitive-component
(unrestricted re-writing) systems will be required.

[Saunders (1966) has emphasized the importance of cell-death in
embryogenesis, so in simulating "real" growth programs, cell deletions
can be expected to play an important role. However, an interesting two-part
question is: (i) "Are there realistic rules which grow all cell structures
by addition alone (monotonic rules)? (ii) Is there, then, some complexity
result which proves that presence of deletion rules makes for more efficient
growth?" The result may be evolutionary - i.e., it "pays" to be able to
evolve in "both" directions.]

6. Problems of Spatial Differentiation

In developing organisms, cells which at their time of origin may be
very similar often eventually come to be very different. Cells which
come to lie in distinct spatial locations often take on very different
and specific characteristics. A cell in the head becomes a Purkinje cell,
say, and a cell in the arm becomes striated muscle. The origin of cell
difference and its relation to cell position is the "problem of spatial
differentiation."

Can we employ our artificial organism systems to explain how this
might come about? Our modelling system is flexible enough to permit us
to get anything at all in the way of organism development in a single rule

(e.g., S → [giraffe]) but the "pre-knowledge" implicit in such a theory of development renders it useless for any adequate accounting for the developmental process. If we limit ourselves to employing only component-by-component (e.g., cell-by-cell) rules of development in our artificial organism systems can we account for all the phenomena of biological development? We know that the properties of unrestricted string rewriting systems (which are a subset of our general-sensitive component artificial organism systems) guarantee that *any* effective process can be represented. (An effective process is one that can be stated unambiguously and carried to its conclusions mechanically.) Thus if we accept that real biological developmental processes do not go beyond the mathematically algorithmic, we are assured that these systems can (with an unspecified degree of direct expressive adequacy) account for any process of real biological development. It is incumbent on us though to try to be more specific and to try to isolate and display some of the particular artificial organism processes by which the spatial differentiation of organisms could be explained.

A developing embryo, early in its life, comes to be composed of three distinct layers of cells, the endoderm, the mesoderm, and the ectoderm. Can an artificial organism system produce this tripartite cell structure?

The problem has been recast in a highly abstract form, the biological "French flag" problem (Wolpert, 1968). A biological "French flag" is a connected system of three kinds of cells, "red", "white", and "blue", separated into three distinct bands, each band being equal in size. The simplest form of the "problem" is merely to display the developmental rules which will produce the "flag".

The problem can be further abstracted into that of displaying the rules of growth which will yield a string of cells such that the cells of the

first third of the string are in state a, those in the second third are
in state b, and those of the last third are in state c; that is, the
numbers of the different kinds of cells must be the same or in fixed pro-
portions   (the cell strings are of the form $a^n b^n c^n$ or in general
$a^{kn} b^{\ell n} c^{mn}$, where k, ℓ, and m are fixed). It is this paradigmatic version
of the embryo cell differentiation (or cell segregation) problem we will
consider in the next pages. The numerous examples which follow are meant to
bring out the capabilities and limitations of artificial organism
systems, and to introduce the notion of multiple formal solutions to a
problem, each of which must be evaluated for its adequacy in explaining
developmental phenomena.

If we wish to produce a string of cells of the form $a^k b^k c^k$ (where k
is some *fixed* number) this can be accomplished by a very simple artificial
organism production system: an artificial organism production system all
of whose rules are of the restricted autonomous sort (that is where at
each application one newly differentiated cell is produced). (Indeed we
have earlier seen than *any* particular string whatsoever can be produced
employing only a restricted autonomous artificial organism system.)

For example, (where k = 2) the highly abstract artificial organism
"embryo" aabbcc can be produced by the following rules (where we assume
throughout the discussion of this section that we are always given an
initial cell S):

$$S \rightarrow aS_1$$
$$S_1 \rightarrow aS_2$$
$$S_2 \rightarrow bS_3$$
$$S_3 \rightarrow bS_4$$
$$S_4 \rightarrow cS_5$$
$$S_5 \rightarrow c$$

Note that for a six-celled organism six rules are required. If we do not restrict ourselves to the use of restricted autonomous production rules we can reduce the number of rules required; we can also employ rules which may be biologically more "natural". For example, aaaabbbbcccc can be produced by the following set of rules:

$$S \rightarrow A\ D$$

$$A \rightarrow A_1 A_1$$

$$A_1 \rightarrow a\ a$$

$$D \rightarrow B\ C$$

$$B \rightarrow B_1 B_1$$

$$B_1 \rightarrow b\ b$$

$$C \rightarrow C_1 C_1$$

$$C_1 \rightarrow c\ c$$

Here, an organism of twelve cells is produced by a system of eight rules.

We have already pointed out (in Section 5) that even with the numbers of rules reduced by this means, the number of rules would, for many interesting cases, remain huge. A system of description which requires millions of distinct statements is extremely difficult to manipulate, even with computer assistance, and certainly close to worthless as far as explanatory power goes. The difficulty of course lies in part in the "absolute addressing" autonomous-cell form of rule system which we have been here employing, in which the state and place of every new cell must be carried explicitly in separate rules. Although repetition of organism structure allows us to employ some rules in several places, the number of rules is still a function of the size of the completed organism. This iron relationship

between number of developmental rules and number of stages of growth,

mechanically "unfolding", requires that we know ahead

of time the type and position of each cell of the organism. Such sorts

of organism developmental system provide little explanation. Absolutely

specified growth may be a good model of some very local morphogenetic

processes but it does not help much in explaining over-all organism

development.

If we are to explicate the developmental processes of organisms,

we must define artificial organism systems in which the number and kinds of

rules is a function of the *complexity* of the organism developmental process

not just the *size* of the completed organism. Depending on the control

conventions for the way we use our rules there are many solutions to this

problem.

The rule system below, for example, employs only seven rules, all of

them of the restricted autonomous sort.

1. S → A

2. A → a A

3. A → B

4. B → b B

5. B → C

6. C → c C

7. C → c

Despite the few rules, and their weakness, this system generates all

the equally tripartite string embryos of any size; unfortunately, the

rules, if applied *ad libitum* (which is here the control assumption by

default) also generate all the *unequally* tripartite string embryos.

An inspection of these rules reveals that the system begins by permitting the generation of any number of a s whatsoever (including none at all), next generates any number of b s and then any number of c s. There is nothing in the system which constrains the numbers of the three types of cells to be the same, or in fixed proportions or even approximately the same proportion.
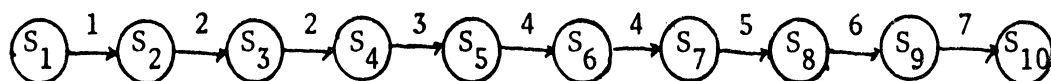
If, however, we make the assumption that some timing and switching control property of the environment of the growing embryo determined that the rule producing new a-type cells would be applied some particular number of times, and that then there would be a switch to producing b-type cells and then to c-type cells, so that each of the three cell-type production rules was applied the same number of times, then the above system would perhaps satisfy our requirements.

This might be imposed by a very highly structured intrusive environment. (cf. our "hippo" example of Section 5).

For example, the environment might be expressed as a program of instructions to the developing organism. The rudimentary program of rules to be applied, given below, will produce the string organism aabbcc.

1. Rule 1.

2. Rule 2.

3. Rule 2.

4. Rule 3.

5. Rule 4.

6. Rule 4.

7. Rule 5.

8. Rule 6.

9. Rule 7.

10. Stop

The identical environmental instructional information could be expressed as a simple finite-state automaton sequence generator.

$$\boxed{S_1} \xrightarrow{1} \boxed{S_2} \xrightarrow{2} \boxed{S_3} \xrightarrow{2} \boxed{S_4} \xrightarrow{3} \boxed{S_5} \xrightarrow{4} \boxed{S_6} \xrightarrow{4} \boxed{S_7} \xrightarrow{5} \boxed{S_8} \xrightarrow{6} \boxed{S_9} \xrightarrow{7} \boxed{S_{10}}$$

The same environmental effect could also be expressed as a computer program flow chart diagram. The examples given of environmental effect are extremely simple, but very much more complex and sophisticated examples are possible of (1) *tables* giving time and place of rule application, or (2) *programs* or (3) *automata* which specify time and place of rule applications.

By such means, the environment effect could be expressed with varying degrees of specificity (from complete "hands-off" the re-write rules, to complete determination of each time and place of rule use). Thus, the control might leave alternative developmental routes open, and thus define a *class* of consequences, or foreclose all but one developmental route and thus define a single consequence of the re-write rules.

There are of course many other ways in which the impact of the environment can be expressed; these would include ordinary natural language statements of the way the environment affects the organism rule use.

In our example we have had our environment impose the sequential production of equal numbers of cell types. This particular environmental effect seems a very improbable and *ad hoc* property however, and a developmental theory which requires this assumption is not likely to survive empirical scrutiny.

Perhaps more naturally the environment should impose the production of the three cell types in *parallel*, thus eliminating the necessity for separate sequential countings or timings.

We can begin by first creating a string of one each of the three

cell-types, either directly, in one step.

S → A B C

or (perhaps more naturally) by several steps

S → A B_1

B_1 → B C_1

C_1 → C

We then apply three separate growth rules in parallel, to the three

cell-types.

For example,

A → a A		A → a A a		A → A A
B → b B	or	B → b B b	or	B → B B
C → c C		C → c C c		C → C C

Any of the above sets of three rules, if employed in synchrony, to

all applicable cells, would assure that the numbers of a-type, b-type,

and c-type, cells will remain the same, and that the tripartite embryo

form will be retained. (Notice that the different sets of rules produce

different numbers of new cells per generation and they thus simulate the

developmental process differently.)

To get any particular size "embryo" we must finally introduce some

sort of "stop" rule. For example:

A → a		A → aa
B → b	or	B → bb
C → c		C → cc

Such "stop rules" produce symbols for which there are no further

applicable rules (and thus might define an interpretable stage or inter-

esting stage of development). Here also we must consider how we want

to apply such rules. We could for example apply these rules *simultaneously*

to *all* applicable cells; the equal tripartite form would then obviously be

retained. A "broadcast" hormonal signal to all cells would be of the proper

type, and would serve to preserve the equal number of cell-types. If

however we permitted the piecemeal application of these rules, the equality

of the tripartite partern might be broken up (since some of the cells could

by the earlier rules continue for a time to reproduce their type).

One possible solution to this is to have our control convention single

out *obligatory* rules which when *one* cell-type ceases reproduction of

its kind, it affects the others to shut-off their production of their

type, thus bringing this phase of development to a close.

For example, if the initial shut-off occurred in the region of a-type

cells the following obligatory rules would cause the complete shut-off

of this growth phase.

aA → aa

Aa → aa

aB → ab

bB → bb

etc.

Notice that in order to produce the desired effect, an obligatory

rule must not only be applied if it is applicable, but it must also be

required that *no other rules* be applied until the obligatory rule *ceases*

to be applicable. By such techniques of obligatory rule use, we can, in

a one-rule-at-a-time system get some of the *effect* of parallel simultaneous

action (Notice that we may by such techniques produce many intermediate

nonsignificant stages of development.)

We may distinguish two facets of this problem: that of the *origin*

of the "shut-off" signal, and that of the *agent* that actually causes the

halt in further increase in cells at this stage and/or the completion of

the tripartite differentiation of cell types.

The *origin* of the "shut-off" signal (the "condition" satisfied) could

be outside the developing embryo entirely; it could have its origin in changes in the maternal organism, or in the larger environment beyond. Although "insult" intervention from the outer environment might be the source of the "shut-off" signal (and such phenomena could be expreseed in these artificial organism systems) such phenomena can not *directly* give us the information we desire on "natural" developmental behavior. If the origin of the shut-off signal to the embryo is in the maternal organism, then an adequate explanation of development must incorporate selected features of the maternal host.

[We can for example imagine a timing apparatus in the mother organism, running separately but parallel to the growing embryo which when it reaches a certain state, sends a "signal" to the embryo to cease new cell production and complete the initial differentiation.]

The condition to be satisfied to initiate the shut-off signal, could come from the developing embryo itself. The *size* reached by the embryo might be the condition (detected by interaction with the maternal organism, or by the developing embryo alone). The *state* reached by one or all cells of the embryo, might be the condition to be satisfied. The cells of each new generation might differ slightly from the parent cells, and the embryo might, by means of a sort of internal counting, determine its own shut-off time.

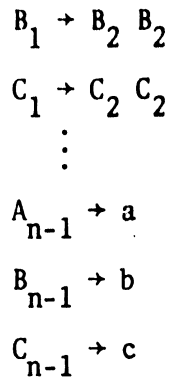This last might be accomplished in the following fashion :

$S \rightarrow A \ B \ C$

$A \rightarrow A_1 \ A_1$

$B \rightarrow B_1 \ B_1$

$C \rightarrow C_1 \ C_1$

$A_1 \rightarrow A_2 \ A_2$

$$B_1 \rightarrow B_2 \; B_2$$

$$C_1 \rightarrow C_2 \; C_2$$

$$\vdots$$

$$A_{n-1} \rightarrow a$$

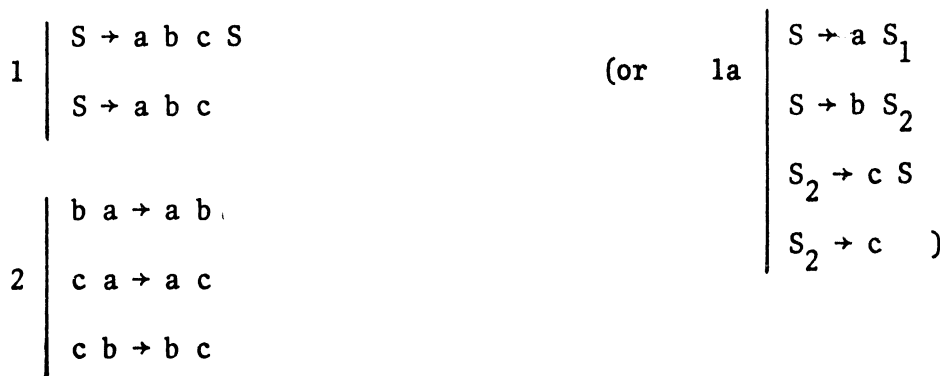$$B_{n-1} \rightarrow b$$

$$C_{n-1} \rightarrow c$$

The "counting structure" of this system is easily understood.

The number of rules is, unfortunately, linked to the number of cell *generations*, which may or may not be a desirable condition. (We can for example imagine that some substance may be accumulating until it reaches a threshold concentration that inhibits further divisions or otherwise produces the effect of a "switch",)

Two principal difficulties of the above schemes for producing a tripartite embryo have been the "synchrony" rule and the "stop" rule. We have discussed the "stop" rule at some length.

Let us now try to incorporate into the explicit developmental system the effect of the synchrony rule by which the tripartite form of the embryo was retained despite its variable over-all size.

The following artificial organism system could be employed:

1
| $S \rightarrow a \; b \; c \; S$
| $S \rightarrow a \; b \; c$

(or   1a
| $S \rightarrow a \; S_1$
| $S \rightarrow b \; S_2$
| $S_2 \rightarrow c \; S$
| $S_2 \rightarrow c$)

2
| $b \; a \rightarrow a \; b$
| $c \; a \rightarrow a \; c$
| $c \; b \rightarrow b \; c$

This system produces the infinite set consisting of the equally tripartite strings of all lengths. In this system the equal numbers of cell types, and their separation into distinct regions is an inherent property of the system itself.

Notice that rule-set 1 (or 1a) consists of autonomous-component rules (context-free rules); indeed they may be restricted component rules (finite state rules), which are a subset of the autonomous-component rules. These rules determine the ultimate size of the embryonic organism (the number of cells in the "embryo").

On the other hand, the rules of the second part of the system are sensitive-component rules (more particularly monotonic or linguists' context-sensitive rules); applicability of these rules depends not only on the state of a particular cell, but may also depend on the states of certain neighboring cells.

[The rules in this second part of the system are basically "re-arranging" or "re-shuffling" rules.

Note that there are two biological interpretations of the function carried out by these latter sort of rules. They can be interpreted as rules which have a physical cell transport and sorting effect, i.e., morphogenetic cell movement rules; on the other hand, they can be interpreted as dual cell differentiation rules in which there is no actual physical cell movement, but rather the cell on the left differentiates to take on the characteristics of the cell on its right and the cell on the right takes on the characteristics of the cell on its left. That two biologically striking different processes might have the same morphogenetic effect must be kept in mind when relating artificial organism systems to possible biological counterparts.]

In all systems we have so far discussed, the cells, as they were

created, were assumed already to be distinguishable as a, or b, or c types. There are many reasons why we should not take this to be the only way in which our tripartite embryo could be formed. Indeed, the "embryological problem" is often posed as the problem of an undifferentiated cell somehow discovering what it should differentiate into. This is sometimes posed as a problem of differentiating according to spatial location: how does a largely undifferentiated cell find out *where* it is so that it can become the sort of cell appropriate to its spatial location. The polar point of view, (as taken in some of the above systems) is that the cell knows where it is (or where it is to go) because of what it has already become.

Most embryological processes are probably an interaction of those two situations: a cell knows only *what* it is and what its local neighborhood is like; but some of what it knows of itself and its local neighborhood tells it where it is, and what it should further become.

It is quite clear though that the cells of many embryos go through an uncommitted" stage, such that if parts are removed prior to the committed stage, the remaining cells are competent to produce a complete (though possibly smaller embryo) rather than a structurally incomplete embryo.

How can we express this phenomenon in artificial organism systems?

In a very rudimentary form the artificial organism system below expresses this possibility.

$$
1 \quad \left|
\begin{array}{l}
S \rightarrow u \\
\\
u \rightarrow u\ u
\end{array}
\right.
$$

$$
2 \quad \left|
\begin{array}{l}
u \rightarrow a\ B \\
B \rightarrow b\ C \\
C \rightarrow c
\end{array}
\right.
$$

$$
3 \quad \left| \begin{array}{l} b\ a \rightarrow a\ b \\ c\ a \rightarrow a\ c \\ c\ b \rightarrow b\ c \end{array} \right.
$$

The first set of rules produces some number of undifferentiated cells (the exact number not inherently controlled within this particular organism system). The second group of rules converts each u cell into a triple of differentiated cells, one of each of the final types. The last set of rules re-arranges these differentiated cells so that they assume the correct final tripartite arrangement.

Notice that removal of u-type cells will reduce the ultimate size of the organism but will not affect the development of the proper tripartite arrangement. Once differentiation has begun however (group 2 rules) removal of cells may disrupt the equality of the three final cell types. Notice also that although this artificial organism may not be able to recover from *removal* of cells at later stages it may still be able to recover from *disruption* of cell pattern. If the proper tripartite pattern is disturbed so that, e.g., a-cells are placed in among b-cells, the application of the third group of rules will restore the proper arrangement. Thus, this artificial organism has some capability for self-repair.

We can express the final loss of this capability by adding the differentiation rules.

$a \rightarrow a^*$

$b \rightarrow b^*$

$c \rightarrow c^*$

where there are no rearrangement rules for the differentiated a^*, b^*, c^* cell types.

It is clear that the re-arrangement rules are very powerful (recall

also that they are sensitive-component rules; they reflect action as a consequence of inter-cell communication). Such rules allow us to produce order among cell types which were produced somewhat chaotically. Is there a more orderly way of producing the tripartite arrangement of cells? Is there a way in which the artificial organism goes through an undifferentiated stage, and the final cell types are produced "in place", not produced scattered and then "re-arranged"?

One system which has the property of possessing a stage at which removal of cells affects size but not completeness of form, and which does not employ "re-shuffling" rules, is the following (given in part):

0) $S \rightarrow u$

1) $u \rightarrow uu$

2) $u \rightarrow \underline{a}$

3) $\underline{a} \; u \rightarrow \underline{a} \; \underline{a}$

4) $u \; \underline{a} \rightarrow \underline{a} \; \underline{a}$

5) $\# \; \underline{a} \rightarrow \# \; a'$

6) $a' \; \underline{a} \rightarrow a \; \underline{a}'$

7) $a \; \underline{a} \rightarrow a \; a'$

8) $\underline{a}' \; \# \rightarrow \underline{a} \; \underline{b}$

9) $a' \; \# \rightarrow a \; \underline{b}$

10) $\underline{a}' \; b \rightarrow \underline{a} \; \underline{b} \; \underline{b}$

11) $a' \; \underline{b} \rightarrow a \; \underline{b} \; \underline{b}$

(etc. for the b to c segment)

In this system,   Rule 1 produces any number of u-cells. Rule 2) initiates a differentiation stage. Rules 3) and 4) are to be interpreted as obligatory: if they *can* be applied they *should* be, until no longer applicable, no other rule use intervening. This means that once rule 2)

is applied production of further u-cells ceases.

Rules 5), 6), and 7) initiate and propagate a growth signal. The signal starts at the open (#) left end of the string of a-cells and is propagated to the right.

Rules 8) and 9) show the consequences of growth signal: b-cells are produced to the right.

Notice that each cell growth signal initiates only in a-cells (cells with an underline); after each growth signal propagation, the signalling a cell becomes an a-cell. Thus the number of growth signals (which determines the number of b-cells to be formed) is equal to the number of a-cells converted to a-cells.

At this point (although the required further rules have not been set down here) the b cells control the production of c-cells; c-cells (rather than c-cells) can be produced directly since at this time we do not ask that the c-cells control any further growth.

Notice that any removal of cells up to the time the a-cells begin the control of production of b-cells will yield smaller but complete tripartite "embryos".

An example of the processes of this system (beginning after the initial cells have been produced) is given below. Rule 5) is first to be applied.

    # a a a

    # a' a a

    # a a' a

    # a a a'

    # a a' a'

    # a a' a b

    # a a a' b

a a a' b̲ b̲

a a a' b̲ b̲

a a a b̲ b̲ b̲ etc.

In the last example, we made the assumption that the open left end of the string made sufficient "environmental" difference, that the initial (leftend) cell might differentiate and "send growth pulses". (This might be expressed either by making the end cells sensitive to the *absence* of other cells or to the *presence* of explicit neighbor *environmental* components. The effect of the external environment might be introduced by (finite) automaton environment components which feed initial (or perhaps even *constant*) pulses to the end cells.

Employing either of these notions, another way in which we might get the tripartite embryo, is the following.

We begin with an undifferentiated string. Differentiation begins at both (sensitive) ends, the left end cells becoming a-cells, and the right end cells becoming c-cells. If differentiation takes place at the same rate at both ends, then when the differentiation process "meets" at the center, equal numbers of a-cells and c-cells will have been formed. At this point, either the a-cell or c-cell segment could (as in the past example) control the production of an equal number of intervening b-cells.

In the last two examples, we have produced the tripartite form, by first producing part of the structure (either the a-cells, or the c-cells, or both) and then completing the form by producing in addition, the correct number of required new cells of the correct sort.

Another way of posing the tripartite embryo question, is to ask that a block of undifferentiated cells be produced, and that *all* three types of cells be produced from this block, *in place* (without re-shuffling),

and without producing any additional new cells.

If we permit ourselves a control apparatus imposing some synchronous rule application, this can be accomplished in the following fashion.

We begin, (as in the last example) by commencing a differentiation of a-cells from the left end, and c-cells from the right end; the cells are not as yet committed to their roles however. The "proto-differentiations" will meet at the center of the segment. We now assume that the "collision and conflict" interaction of the opposing a-cell and c-cell proto-differentiation initiates the differentiation of b-cells. The differentiation of b-cells begins to diffuse" out from the center. If the completion of differentiation from the ends, and the completion of differentiation from the center proceed at the proper rates, then the desired tripartite embryo will be obtained.

Notice first that we can represent different diffusion rates or communication speed or rapidity of differentiation by sending a cell through a series of internal states, the first of which represents the onset of the process or signal, and the last of which represents the completion of the process or the final onward transmission of the signal. If the process takes long (or the signal delay is to be lengthy) then the number of internal transitional states of the modelling component is large; if the process is quick, then the number of the intervening internal states of the component will be few.

One way of using such timing techniques to get the desired tripartite result is the following.

1. Two signals are started simultaneously at the a-cell end, (and two parallel signals are started at the c-cell end). The first of these signals is passed along at the fastest rate, i.e., one

cell each rule application. The first signals meet in the center

of the segment of undifferentiated cells, and initiate the differ-

entiation of b-cells out from the center. Differentiation of b-

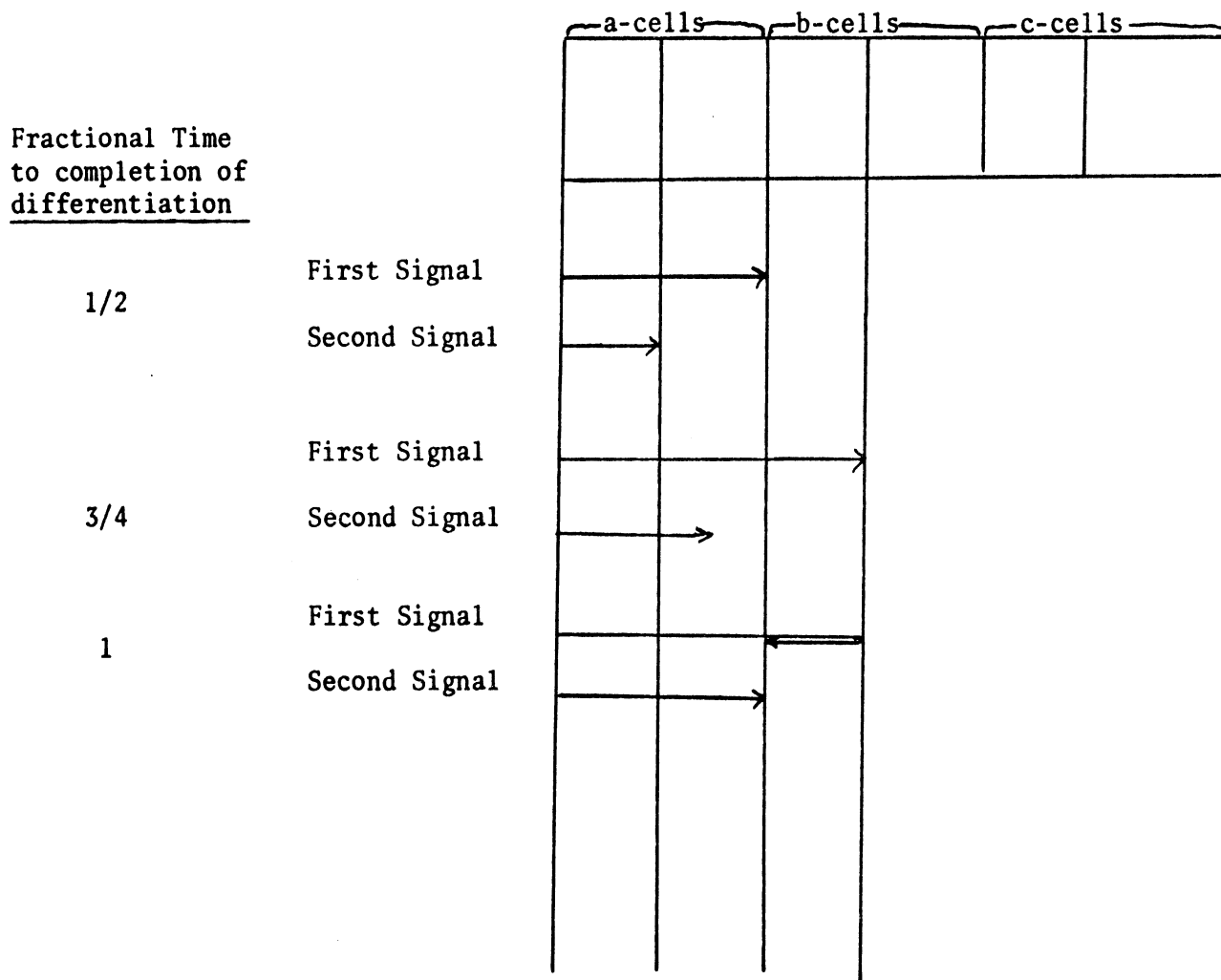cells proceeds at the same rate: one new cell (in each direction)

per rule application.

2) At the same time, the second sort of signal has been proceeding

at the rate of one cell each *two* rule applications; that is, at

half the speed of the first signal.

The second signal represents the passage of differentiation of a-cells

(on the left) and c-cells (on the right).

3) The differentiating processes from the center and from the ends

will meet one third of the way in from both ends; at the meeting,

the processes stop, and the tripartite form is complete.

The timing scheme (outlined in Figure 3) is not the only one that

might be employed. The same ultimate outcome can be obtained with many

other differentiation and timing schemes (cf. Arbib (1969) and Herman

(1969)).

Fractional Time
to completion of
differentiation

1/2   First Signal

            Second Signal

            First Signal

3/4   Second Signal

            First Signal

1   Second Signal

a-cells   b-cells   c-cells

Figure 3. Timing Scheme for Tripartite Differentiation.
Only the differentiation of the "left-hand"
side of the organism is described, the right-
hand side acting symmetrically.

References

Arbib, M.A. (1966) "A Simple Self-Reproducing Universal Automaton" *Information and Control*, *9*, 177-189.

___, (1967) "Automata Theory and Development: Part I" *J. of Theoret. Biol.*, *14*, 131-156.

___, (1970) "Self-Reproducing Automata - Some Implications for Theoretical Biology" Towards a Theoretical Biology, *Sketches*. 204-226.

Banks, E.R. (1971) *Information Processing and Transmission in Cellular Automata*. Mass. Inst. of Tech. thesis. 100 pp.

Bar-Hillel, Y.; Perles, M.; and Shamir, E. (1961) "On Formal Properties of Simple Phrase-Structure Grammars" *Z. Phonetik, Sprachwiss. u. Kommanikationsforsch*, *14*, 143-172.

Burks, A.W. (1970) *Essays on Cellular Automata*, U. of Illinois Press.

Chomsky, N. (1963) "Formal Properties of Grammars" *Handbook of Mathematical Psychology*, *2*, Wiley, New York, 323-418.

___, (1965) *Aspects of the Theory of Syntax*, MIT Press.

Codd, E.F. (1965) "Propagation, Computation, and Construction in Two-Dimensional Cellular Spaces" U. of Michigan Technical Report. Published as *Cellular Automata*, Academic Press (1968).

DeLong, R.G. and Sidman, R. (1962) "Effects of Eye Removal at Birth of the Mouse Superior Colliculus" *J. Comp. Neurol.*, *118*, 205-224.

Fris, I. (1968) "Grammars with Partial Ordering of the Rules" *Inf. and Cont.*, *12*, 415-425.

Fu, K.-S. and Swain, P.H. (1971) "On Syntactic Pattern Recognition" *Software Engineering*, Volume 2, Academic Press, 155-182.

Ginsburg, S. and Spanier, E. (1968) "Control Sets on Grammars" *Math. Systems Theory*, *2*, 159-177.

Herman, G.T. (1969) "Computing Ability of a Developmental Model for Filamentous Organisms" *J. of Theoret. Biol.*, *25*, 3, 421-435.

___, (1970) "Role of Environment in Developmental Models" *J. of Theoret. Biol.*, *29*, 3, 329-342.

Hopcroft, J. and Ullman, J. (1969) *Formal Languages and Their Relation to Automata*, Addison-Wesley, Reading, Mass.

Kalman, R.E.; Falb, P.L.; and Arbib, M.A. (1969) *Topics in Mathematical System Theory*, McGraw-Hill, New York.

Laing, R.A. (1971) "A Hierarchy of Developmental Provesses" *Inter. J. Neuroscience, 2,* 219

Lindenmayer, A. (1971) "Developmental Systems without Cellular Interactions, Their Languages and Grammars" *J. of Theoret. Biol., 30,* 3, 455-484.

Milgram, D. and Rosenfeld, A. (1970) "Array Automata and Array Grammars" U. of Maryland Computer Science Center, Technical Report 70-141.

Montanari, U.G. (1970) "Separable Graphs, Planar Graphs and Web Grammars" *Inf. and Cont., 16,* 243-267.

Pfaltz, J. and Rosenfeld, A. (1969) "Web Grammars" Proceedings of the *Int. Joint Conf. on Artificial Intelligence,* (ed., Walker and Norton). 609-619.

Rosenfeld, A. (197) "Isotonic Grammars, Parallel Grammars, and Picture Grammars" *Machine Intelligence, VI,* (ed., Michie and Meltzer) Edinburgh, U. Press.

Rosenkrantz, D. (1969) "Programmed Grammars and Classes of Formal Languages" *J.A.C.M., 16,* 107-131.

Rosenstiehl, P. (1966) "Existence D'Automates Finis Capables De S'Accorder Bien Qu' Arbitrairement Connectes Et Nombreux" *ICC Bulletin, 5,* 245-261.

Salomaa, Arto (1970) "Periodically Time-Variant Context-Free Grammars" *Inf. and Cont., 17,* 294-311.

Saunders, J.W., Jr. (1966) "Death in Embryonic Systems" *Science, 154,* 604-612.

Smith, A.R. (1971a) "Cellular Automata Complexity Trade-Offs" *Inf. and Cont.*

___, (1971b) "Two-Dimensional Formal Languages and Pattern Recognition by Cellular Automata" *Conference Record 1971 Twelfth Annual Symposium on Switching and Automata Theory,* Inst. Elect. and Electron. Engineers, New York, 144-152.

Steen, W.J.van der, and Jager, J.C. (1971) "Biology, Causality and Abstraction, with Illustrations from a Behavioral Study of Chemoreception" *J. of Theoret. Biol., 33,* 265-278.

Thatcher, J. (1969) "Universality in the von Neumann Cellular Model" U. of Michigan Technical Report. Also, *Essays on Cellular Automata*, (ed., A.W. Burks), U. of Illinois Press. 132-186 (1970).

Von Neumann, J. (1966) *Theory of Self-Reproducing Automata*, (edited and completed by A.W. Burks), U. of Illinois Press.

Wolpert, L. (1968) "The French Flag Problem: A Contribution to the Discussion on Pattern Development and Regulation" *Towards a Theoretical Biology: I, Prolegomena.* (ed., C.H. Waddington), Edinburgh U. Press, 125-133.

___, (1969) "Positional Information and the Spatial Pattern of Cellular Differentiation" *J. of Theoret. Biol., 25,* 1-47.

Zeigler, B.P. and Weinberg, R. (1970) "System Theoretic Analysis of Models: Computer Simulation of a Living Cell" *J. of Theoret. Biol., 29,* 35-56.