

THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Computer and Communication Sciences Department

Technical Report

Formalisms for Biology:
A Hierarchy of Developmental Processes

Richard A. Laing

with assistance from:

Department of Health, Education, and Welfare
National Institutes of Health
Grant No. GM-12236
Bethesda, Maryland

and

National Science Foundation
Grant No. GJ-519
Washington, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

April 1971

Introduction

[This paper was presented at the Coral Gables Conference on the Physical Principles of Neuronal and Organismic Behavior, December 1970.]

Automaton theory is the study of all effective (machine-like) procedures.

Much of the original impetus for automaton theory research came from biology. For example, most of the early algebraic treatments of finite state machine theory have their ultimate origin (by way of Kleene (1956)) in the attempts of McCulloch and Pitts (1943) to use mathematical logic to define precisely the capabilities of idealized nerve nets; the theory of tessellation automata has its origin in von Neumann's demonstration (1966) of a means by which the process of self-reproduction can be carried out by a machine.

[This result by von Neumann was obtained (following some suggestions of Ulam) in 1952].

To many researchers, however, automaton theory's relevance to biology seemed difficult to demonstrate, and there appeared to be no reasonable hope of any practical applications. In contrast, the relevance of automaton theory to problems of the design and use of computing machines has always seemed quite obvious.

With the tremendous growth of the computing machine field, there was a concomitant huge growth in automaton theory research.

The result has been that there now exists a vast body of automaton theory results (which is being augmented constantly) but the potential relevance of this material for problems of biology has largely been ignored.

In this paper we attempt to show how some of the results of automaton theory can now be brought fruitfully to bear on problems of biology.

In particular we propose that formal grammar theory and automaton

theory can be employed in the creation of a theory of biological development of form.

[In Arbib (1967) an approach to this problem is made, employing an automaton system based on the von Neumann cellular system.]

Formal Grammar Theory

[Formal grammar theory is a part of automaton theory. Formal grammar theory (in this sense) has its origin in Chomsky's program to change the direction of contemporary American linguistic theory. Good surveys of formal grammar theory can be found in Chomsky (1963) and in Hopcroft and Ullman (1969)].

A formal grammar is usually defined as consisting of:

- 1) a finite set of symbols which can be rewritten; these are called the *nonterminal symbols*.
- 2) a finite set of symbols which can *not* be rewritten; these are called the *terminal symbols*.
- 3) a finite set of *production rules*, which express the conditions under which nonterminal symbols may be rewritten.
- 4) a (usually) single distinguished symbol (which is in all interesting cases one of the nonterminal symbols) called the *initial* (or "start") *symbol*.

For example, a formal grammar might consist of the single nonterminal symbol S (which will also serve as our initial symbol), the two terminal symbols a, b , and the two production rules: 1) $S \rightarrow aSb$, 2) $S \rightarrow ab$.

Starting with the initial symbol S , we may apply rule 1) and by this

means replace S with aSb . We might then apply rule 2) and by it convert aSb to $aabb$. At this point no further rule applications are possible, for $aabb$ contains no nonterminal symbols (it contains no symbols appearing on the left hand side of a production rule). The generative process thus stops at this point. Note however, that we could have repeatedly applied rule 1) before applying rule 2). It will be seen then, that this formal grammar defines the (infinite) set of finite strings of the form $a^n b^n$ (that is to say, the strings composed of initial segments of any number of a 's followed by a final segment of the same number of b 's).

Types of Formal Grammars

Several types of such formal grammars have been defined.

If we allow arbitrary replacement of one string by another, (an "unrestricted re-writing system") we get a *type 0 grammar*.

If we impose the requirement that the string on the right hand side of the rule always be at least as long as the string on the left hand side (that is, that the application of the rule never *shortens* the resulting string) we get a *type 1 grammar*. (If we restrict a *type 0 grammar* by employing production rules which call for replacing of only *one* symbol at a time by one or more symbols *and* the applicability of the production rule may be conditioned by the presence of strings of symbols adjacent to the symbol to be re-written, we again get a *type 1 grammar*. This alternative definition motivates the alternative name *context-sensitive*, for type 1 grammars.)

If we impose on a type 1 grammar the requirement that at the application of a production rule precisely one nonterminal is to be replaced by one or more symbols (and the context of the replaced symbol does not condition

the applicability of the production rule) we get a *type 2 grammar*. Type 2 grammars are often called *context-free* grammars. (Note that our example in the last section is a context-free grammar.)

If we further impose the requirement that every production rule must be one of the two forms $A \rightarrow aB$ or $A \rightarrow a$ (where capital letters denote nonterminals and lower case letters denote terminals) then we get a *type 3 grammar*. Type 3 grammars are sometimes called *regular grammars* or finite state grammars. (The names "regular" and "finite state" reflect associations which can be made between these grammars and the properties of finite state automata.)

Discussion

We now discuss the use of these grammars in linguistics and introduce a biological interpretation of formal grammars.

[A biological interpretation of formal grammars has also been made by Lindenmayer (1971) and is being further developed by him and his colleagues at the Rijksuniversiteit at Utrecht. Lindenmayer had earlier (1968) suggested a biological interpretation of some finite automaton concepts.]

In a formal grammar we are usually given a single initial "start" symbol. For linguistics this is usually interpreted as the unanalyzed "sentence" symbol. A production rule defines the "re-writing" of the initial symbol. The resulting symbol string may then be further re-written by use of production rules. Eventually a string is produced, the symbols of which express in detail the structure of a sentence.

In a *biological* interpretation of formal grammars we begin with a single initial entity one interpretation of which would be a fertilized cell. Production rules (rules of growth and development) permit us to replace

this initial entity with other entities. If the rules genuinely mirror biological "rules" of development, then the successive entities produced will represent successive stages of organism development.

Let us now try to make clear some of the differences in point of view between the linguistic and the biological interpretation of formal grammars.

In formal grammars there is the assumption that the entity being produced is a *string* of symbols. In biological production systems we may also wish to produce entities which are strings of cells. It is clear though that a biological formalism ought to permit the production of entities which are conjoined in other than a linear fashion.

Our formal biological production systems will thus contain production rules for the direct formation of entities other than the linear.

[Lindenmayer (1971) accepts the constraint that his production rules directly yield only linear entities. This means that in his accounting for tree-like and other non-linear graphical structures a string (which might include both symbols for cell-states and symbols for "punctuation" signifying, for example, branch points) is first produced, and then is "decoded" to yield the final multi-dimensional structure. By this means Lindenmayer is able to produce (among other phenomena) the alternate side branching structure of the red alga *Callithamnion roseum*.

There has been some research on "grammars" for other than linearly connected graphs. See Pfaltz and Rosenfeld (1969) and Montanari (1970)].

From the linguistic point of view (as customarily enunciated) the formal grammar is a machine-like system which is operated "from the outside". That is, the system is viewed as a device which (often in a rather weak sense) "accounts" for the generation or analysis of sentences. Generally

no claims are made that the brain or the language faculty of the mind *must* handle its sentence creation or analysis in precisely this fashion. A formal grammar which satisfactorily accounts for a body of actual linguistic data may be assumed to be a *possible* way in which the linguistic faculty provides its services to us, but only the most daring psycholinguist or neurophysiologist would at present assert that the specific processes of a formal grammar constitute firm evidence for the "real" operation of the human linguistic faculty.

Under the biological interpretation we wish to pursue here, our organism system can be viewed not as merely "accounting for" biological phenomena or as being a system manipulated by an outside faculty. Instead we can view it as biologically "real" and as *self-operated*. That is, we can view the initial symbol and the symbols of all the succeeding assemblages, not as letters being written and re-written (possibly by an outside intellectual faculty) but as (finite automaton) cells capable of being in various states, and capable themselves of carrying out the operations (change of state, reproduction) that the production rules define.

We will generally think of each finite automaton cell of our formal biological system as having an identical state transition system (the same genetic system); although we will generally think of the state transition systems of each automaton of a single organism as being identical we generally will not require all the separate cell automata constantly to remain in the same internal state. We will also want our cell automata to possess as primitive properties such capabilities as self-reproduction, and the creating and dissolution of communication and connection to other cells.

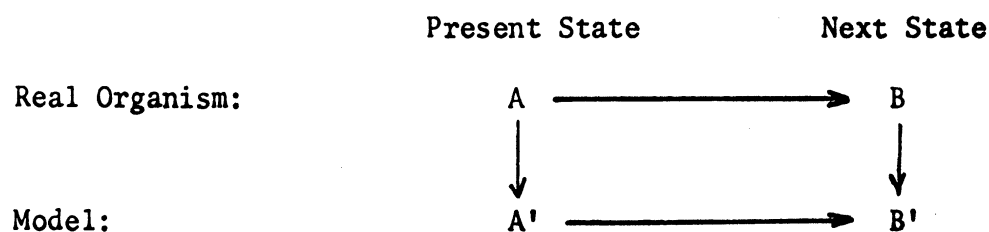
A major consequence of this biological interpretation is that not

only does the ultimate assemblage (in the linguistic case, the actual *terminal* sentence with its dictionary words inserted) have a "real" interpretation as a physical object, but each of the successive intermediate objects produced also has a useful interpretation as a physical biological object.

Thus our goal might be to define organism grammars having the property that the initial objects have a direct physical correlate in biology, and that the production rules mirror biological rules of development so that each successive stage of development of the model corresponds to a stage of development of the real organism.

More precisely, we might ask that our model be a homomorphic image of the real biological system. This means that for every present developmentally significant state of the real system, if we proceed to the next developmentally significant state of the real system, and then find the counterpart state in the model, we will arrive at this same model state if we were to go from the present developmentally significant state of the real system, to the corresponding state of the model, and then calculate the next state of the model.

This idea is expressed graphically below where the path across and down takes us to the same point as the path going down and across. (See Zeigler and Weinberg (1970) for a more complete discussion of the modelling criterion.)



[This "strong" accounting for the development of a biological organism should not be taken as an absolutely necessary goal of our system. As in linguistics a less complete accounting might still be very useful. In a "weak" accounting, we might be able to show how, given certain initial conditions and rules of procedure we could "calculate" an organism, in the sense of produce a description of the structure of a complete organism. In this "weak" accounting, it might be that none of the intermediate stages of calculation would correlate in any very direct fashion with a particular stage of the physical development of the organism. Although a "strong" accounting would be the ideal (and more difficult) result, even the "ultimate calculation" of an organism would be a worthwhile goal.]

In the formal grammars as so far described, the production rules may be employed freely, at any time and place they may apply; this means, that in general the systems are *polygenic* rather than *monogenic*; each system defines the generation of a *set* of entities, not merely a single entity. Thus, under the linguistic interpretation, a generative grammar produces a set of sentences, the *language* defined by the grammar; under the biological interpretation such a generating system produces a *population* of related organisms (one interpretation of which is the set of possible phenotypic consequences of a single genotype).

In the linguistic instance, any particular sequence of application of production rules, yields some particular single sentence. In the biological interpretation, any particular sequence of application of production rules yields some particular single organism. The optionality of the production rules, reflects the physical potential of the organism system (which includes both the organism *per se*, and the pertinent features of the organism environment). The "environment" of the organism can often

be viewed as the selector of those growth rules which are actually applied among the many growth rules which at any moment might be applicable.

What we have just described is a very general view of our formal biological systems for development. By modifying this general definition, we can produce other formal systems which may have biological relevance.

We might for example require that *some* rule must always be applied to each and every cell simultaneously. If this requirement is imposed, a different set of organisms is usually thereby defined.

[Lindenmayer (1971), for example, limits his formal development systems in this manner.]

We might wish to consider *deterministic* formal organism systems. In these systems there would be no optionality of rule application; this would signify that there are no alternative pathways in the developing system at all (this might be interpreted to mean that there is no point where the environment can enter and participate in determining the direction of growth of the organism, or that the organism and the environment are together so completely defined as to bar optionality). In other cases we might wish to make the application of production rules probabilistic, allowing a specified element of chance to enter into the determination of the particular production rule to be applied.

Of course we may very well wish to consider mixed systems for which at some points in time and place of development there are alternative rules (indeterministic or probabilistic) which can be applied (and which may be intended to reflect ways in which the environment can affect the developing organism at that time and place) and at other times no option is possible and the organism, if it continues to survive at all, must follow a pre-determined path.

Environmental Structure

A vast number of developmental phenomena can be described in terms of production rule systems. In expressing these phenomena, we often need not distinguish those rules and rule applications which reflect properties of the *organism* from those rules and rule applications which reflect properties of the *environment* of the organism.

Two points should be made however. The first is that the organism versus environment distinction is repeatedly made and humans find this natural and useful. Secondly, the very phenomena of interest may require, for their proper explication and ultimate explanation, that the organism-environment distinction be made explicit. The observation that patterned (structured) light must fall on a kitten's eyes during a critical period of growth, if proper vision is to be attained, is perhaps an example of this.

We shall therefore want to explore ways of partitioning the rules and their uses to reflect the organism-environment distinction, and to explore ways of expressing, in informative fashion, the "structure" of the environmental rule application system, and the "structure" of the organism rule application system.

One thing we would undoubtedly want to do is to see if by giving complete environmental histories (complete rosters of the time and place of application of rules) we could produce some particular organism forms of interest. Such simple listings of rules with their time and place of application is not likely to be very informative though. We will want some more compact and revealing way of describing environmental histories. This means that we will want to develop precise ways of expressing when and where the growth rules are to be applied.

Very little formal work has been done in this area: there have been some schemes for establishing priority among rules, and the notion of a "programmed" or "controlled" production rule system has been introduced. These notions may be of some interest since the form a program or a control unit takes could be said, in our treatment, to reflect the form the environment takes. (For a discussion of rule selection systems, see Salomaa (1970).)

We would also want to correlate properties of environmental histories with characteristics of real organism environments. (Certainly we will need to make this correlation if we ever hope to subject our theory to an empirical testing.)

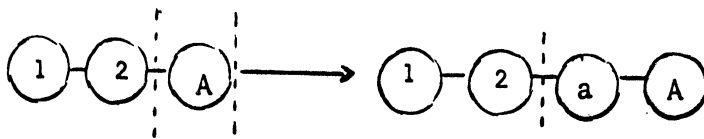
Production Rules for Formal Organisms

In order that we not merely beg the question of how development takes place, we can not merely have rules which, in a "single leap" take an initial fertilized cell and produce from it a completed organism. Our formal biological systems will be informative and explanatory only if our rules are quite simple and local and reflect simple and local real biological processes. A fundamental problem of interest is how simple local rules of growth and development produce the global properties of organisms.

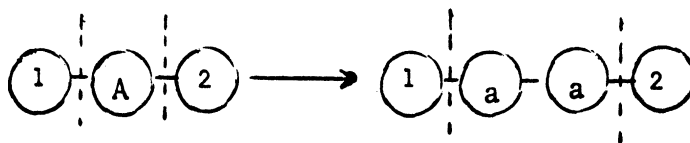
[For a discussion of the importance of local representations in biological modeling see Zeigler and Weinberg (1970).]

In our discussion it will be useful to distinguish two kinds of growth rules: *differentiation rules* and *propagation rules*. Differentiation rules cause the state of a cell to be changed. Propagation rules *may* cause the state of a cell to be changed but *always* change the number of cells or of cell connections of an organism. If the propagation rule causes the number

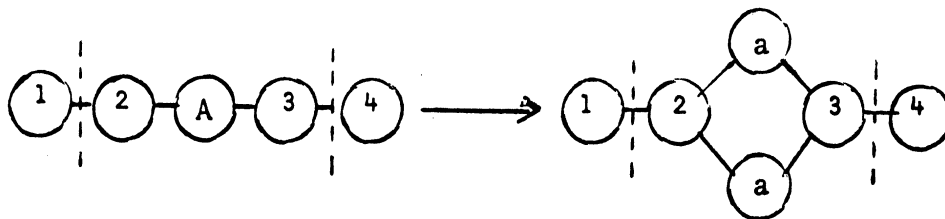
of cells or of cell connections to be increased, we call it a *progressive* propagation rule; if the propagation rule causes the number of cells or cell connections to be decreased, we call it a *regressive* propagation rule. Thus a rule falls into exactly one of the differentiation, progressive, or regressive rule categories. Some examples are given in Table 1. These represent a very few among the many rules which might be employed.



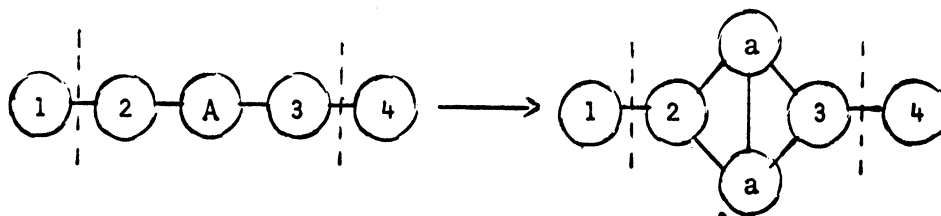
Linear Tip Growth



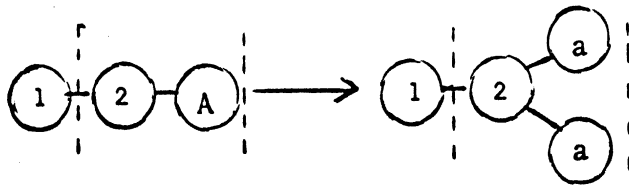
In-line Binary Fission



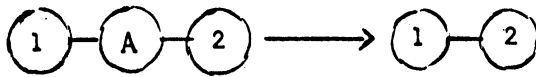
Medial Splitting



Medial Splitting (with Daughter Cell Adhesion)



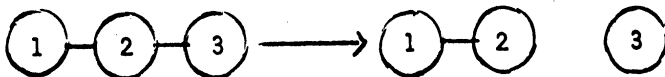
Bifurcation of Growth Tip



Cell Deletion and "closing up"



Cell Deletion and Consequent Separation



Contiguity Deletion

Some Formal Organism Growth Rules

Types of Production Systems for Organisms

Formal linguists, by specifying different kinds of production rules, have distinguished several kinds of grammars; we have earlier noted that the *regular*, the *context-free*, the *context-sensitive*, and the *unrestricted*, are the principal grammar types distinguished.

As we also have already seen, many of the general characteristics of formal grammars and their languages may be given a biological interpretation. We now suggest that the principal types of linguistic production rules can also be given reasonable biological interpretation.

The *regular production rules* correspond to the biological case where a propagation rule causes single new (non-propagating) cells to be "budded-off" by a propagating cell, as for example, at a growing tip; the *context-free rules* correspond to the biological case where the application of a propagation rule is contingent upon the state of the parent cell only (that is, there is no inter-cell communication, the cell is autonomous); the *context-sensitive rules* correspond to the case where propagation *may* depend on the states of cells adjacent to the parent cell; and *unrestricted rules* correspond to the case where regressive propagation rules (that is, rules for the death or dissolution of cells or of cell connections) are permitted.

The four types of grammars we have introduced define a hierarchy of languages. That is, the regular languages are properly included in the context-free, the context-free are properly included in the context-sensitive and the context-sensitive are properly included in the unrestricted. This classifying and ranking of languages has been useful in attempts to distinguish adequate from inadequate models for natural languages. This determination of adequacy is generally done by first establishing empirically a characteristic of natural languages, and then showing that the characteristic can

adequately be expressed by a model (a language) possessing some particular ranking in the hierarchy, but not by any language ranking lower in the hierarchy.

It would be useful to be able to classify formal biological organisms on the basis of *their* allowable production rules, and to rank the classes of organisms obtained. If we can then bring to bear some facts about real biological systems we might then perhaps be able to speak of adequate or inadequate models for classes of real biological organisms.

Our first step in doing this will be to try to make a more precise association between the formal linguistic and formal biological entities. If we do this, many of the mathematical results already obtained for formal linguistics (and automaton theory) would become immediately available to us for use in formal biological theory.

Organism Systems

Of the vast numbers of kinds of organism systems which can be defined by varying the growth rules we will select several systems for further discussion, trying to ascertain their properties and limitations, and trying to rank them. We will begin by introducing some terminology and definitions.

The organism system composed of autonomous-cells only, which permits differentiation rules and in which all of the propagation rules are progressive, and in which new cells are created from existing cells by "budding off" (especially at "growth tips") only, we call the *regular organism systems*.

If such a regular organism system produces organisms consisting of strings of cells, only, we call it a *linear regular organism system*.

The sets of organisms produced by regular organism systems we call the *regular organisms*.

The sets of organisms produced by *linear* regular organism systems we call the *linear regular organisms*.

A system in which all the propagation rules are progressive, and in which differentiation and propagation rules are applicable to any existing cells as a consequence of the state of that cell only (i.e., all cells are autonomous-cells) is an *autonomous-cell organism system*. If any autonomous-cell organism system produces organisms consisting of strings of cells only, we call it a *linear autonomous-cell organism system*.

The sets of organisms produced by an autonomous-cell organism system are the *autonomous-cell organisms*.

The sets of organisms produced by a *linear* autonomous-cell organism system are the *linear autonomous-cell organisms*.

A system in which all propagation rules are progressive, and in which the differentiation and propagation rules are applicable to an existing cell possibly as a function of the states of other cells is a *sensitive-cell organism system*.

If a sensitive-cell organism system produces organisms consisting of strings of cells only, then we call such a system a *linear sensitive-cell organism system*.

The sets of organisms produced by a sensitive-cell organism system are called the *sensitive-cell organisms*.

The sets of organisms produced by a *linear* sensitive-cell organism system are called the *linear sensitive-cell organisms*.

A system which permits the application of differentiation and propagation rules to an existing cell to be mediated by the states of other existing

cells, and in which regressive propagation rules may be employed, is called a *general-cell organism system*.

If a general organism system produces organisms consisting of strings of cells only, then the system is called a *linear general-cell organism system*.

The sets of organisms produced by a general-cell organism system are called the *general-cell organisms*.

The sets of organisms produced by a linear general-cell organism system are called the *linear general-cell organisms*.

Some Theorems of Formal Organism Systems

The theorems listed in this section are mostly rather straightforward translations of results well known in recursive function theory, automaton theory, or (especially) theory of formal grammars.

[See Hopcroft and Ullman (1969)].

In many cases no formal proof is called for; the associations are usually quite obvious.

[The results in this section have, in themselves, almost no significance for biology; they merely tell us the power and properties of the biological *formalisms* we are constructing. In the next section, however, we begin to relate our formal organism systems to the properties of real organisms; by this link, the contents of the present section can be brought to bear on problems of genuine interest.]

Theorem: Given an organism system with no regressive propagation rules, and given an organism, there exists an algorithm by which we can always tell if the organism is a consequence of the system.

(Since only regressive propagation rules decrease the size of the organism being produced, we need only generate all organisms which have as many cells as the given organism and check if our given organism is in this (finite) set of organisms.)

Theorem: The regular grammars and the linear regular organism systems define the same sets.

(All regular languages are sets of linear regular organisms; this is so because all the regular grammars are linear regular organism systems.

Also all the linear regular organism systems are regular grammars. Although the regular linear organism systems have the additional property that differentiation rules may be employed (cell-states may be "re-written" without propagation) this additional property does not take us outside the regular, for for every differentiation rule that might be applied to change the state of a cell, the (regular) propagation rule which originally gave rise to this cell could have placed the cell in the post-differentiation state in the first place.)

[Actually, for most of our applications in this paper we can define the sets of states into which cell may differentiate (without propagation) as terminal *classes* of states, each class equivalent to individual terminal symbols of formal grammars.]

Theorem: The linear regular organisms are accepted by nondeterministic (or deterministic) finite-state automata.

An automaton *accepts* the strings of a set by examining the symbols of a string one-by-one, and upon completion of the examination goes into an acceptance or a rejection state; in effect the machine sorts objects into members and non-members of a given set. The regular sets (finite state languages) are accepted by finite-state automata (known result); we have

already shown (above) that the regular sets and the sets of linear regular organisms are the same.)

Conjecture: The regular organisms can be "checked" by finite-state automata.

(We have used the term "checked" rather than "accepted" since the notion of a device receiving as input an entity which is not a string does not have a generally understood meaning. We here mean by it that a discrete device could, given a set of regular organism rules, be employed to check whether the production of organism could be a consequence of the rules. Our claim is that the checking device needs only a fixed finite number of states to carry out its job.

The recently developed notion of "tree automata" -- automata which accept tree-shaped symbol structures, not just symbol strings - will undoubtedly be useful in defining appropriate checking devices. For a review of this new field see Thatcher (1969).)

Theorem: The context-free grammars and the linear autonomous-cell organism systems define the same sets.

(This follows from the way we have defined the linear autonomous-cell organism systems.)

Theorem: There are sets of linear autonomous-cell organisms which can not be accepted by finite state devices.

(There are context-free languages which can *not* be accepted by finite state devices; this is a known result.)

Theorem: The sets of linear autonomous-cell organisms are accepted by non-deterministic push-down store automata.

(The context-free languages are known to be accepted by non-deterministic push-down store automata.)

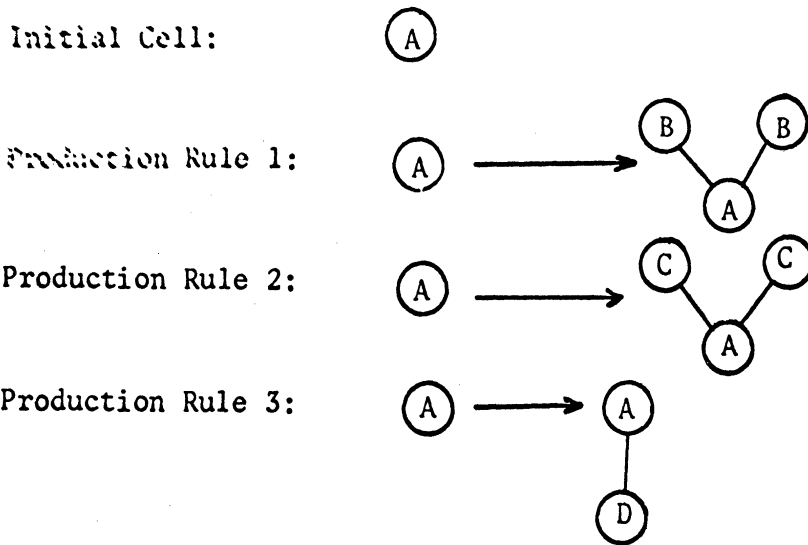
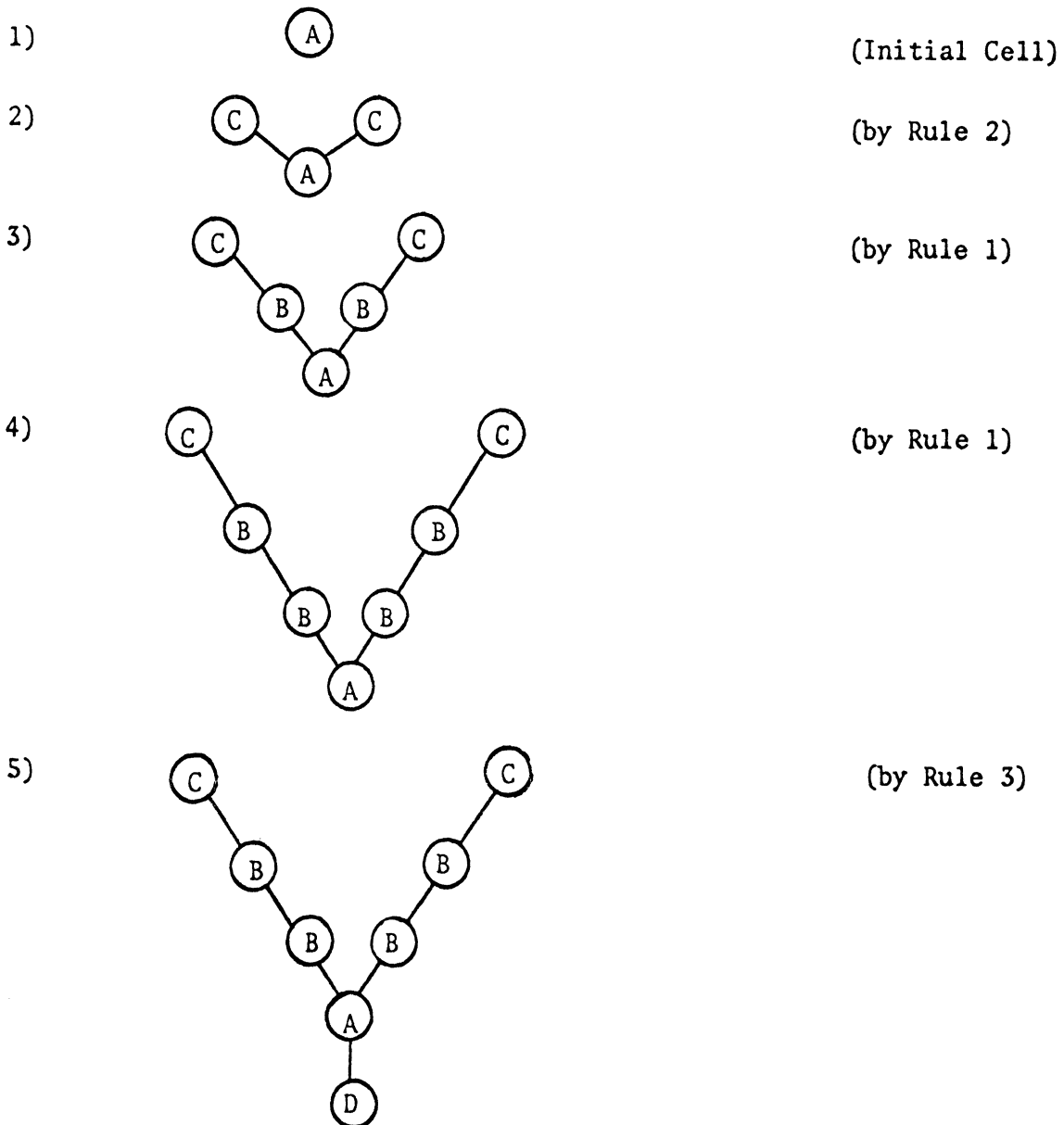
Conjecture: The sets of autonomous-cell organisms can be checked
by (non-deterministic) push-down store automata.

(Here again we have the notion of generalizing automaton behavior to include the idea of checking a graphical structure which is not necessarily a simple string. I conjecture that a push-down store memory system would be sufficient for the device to check whether an organism is a consequence of a given set of autonomous-cell rules.)

Theorem: The sets of linear regular organisms are properly contained
in the sets of linear autonomous-cell organisms.

Theorem: The sets of regular organisms are properly contained in
the sets of autonomous-cell organisms.

(By definition any linear regular organism is also a linear autonomous-cell organism; to show that the inclusion is proper, we must exhibit a set of autonomous-cell organisms which is not also a set of regular organisms. It is known that the " $a^n b^n$ " set of linear autonomous-cell organisms is not also a set of regular organisms. Again by definition a regular organism is also an autonomous organism. We shall show that the set of organisms which is produced by the autonomous-cell system illustrated on the next page is not a regular set of organisms:

A Simple Organism SystemExample Derivation:

These rules will produce "Y" shaped organisms where branches of the "Y" are of equal length and will be composed of cells in the states C and B and where the C,B pattern will be the same in both branches, and where the branches may be indefinitely long.

This precise set of organisms cannot be produced by a regular organism system. Although regular organism systems can yield "Y" shaped organisms the growth of the organism can take place only at the tips of the "Y". Since the cells are autonomous there is no way to "synchronize" the actions at the growth tips so as to insure that the branches will be the same length and possess the same pattern of cell states. The consequence will be that any regular organism system which produces the "symmetric" organisms, will necessarily produce numerous non-symmetric organisms also.

[In an autonomous-cell system (such as in the above example) the growth rules are applied as a function of the states of cells only (and not for example as a function of the states of contiguous cells). Thus, above, a cell in state A will as a consequence of internal conditions only, produce paired cells (both either in state C or in state B) and they will be generated at specific locations, viz. "upper right, and upper left corners". In this simple example the diagrammatic presentation of the production rules suffices to make clear the intended process. There will of course, be cases where more detailed and specific description of the intended effect of growth rules will be necessary; the precise location and orientation of new cells may have to be given; in regressive rules, the post-deletion contiguities may have to be explicitly defined. Most of these issues do not arise in ordinary formal grammar theory, for there the "string" nature of the structures is always assumed, a symbol having at most two neighbor symbols.]

Theorem: Given two autonomous-cell systems, it will be impossible to tell in general whether the sets of organisms produced by the two systems have any organisms in common.

(This follows from results in theory of context-free languages, and the associations we have already made between autonomous-cell systems and context-free grammars. The result can also be obtained directly by the following: take as the first autonomous-cell system the system producing the set of "Y" shaped organisms of the last theorem. Take as the second autonomous-cell system a system, similar to the above but in which at every step a *string* of cells in states C and D is produced up the left hand branch, and a string of cells in states C and D is produced up the right hand branch. For every Post correspondence system (Minsky, 1967), there is such a system. The organisms of the first system will define every *successful* correspondence. If we could tell of every such given set of organisms of the first sort whether it had any members in common with the organisms of the latter sort then we could solve the Post correspondence problem, which is known to be unsolvable. Thus the sought for intersection algorithm does not exist.)

Theorem: Given two autonomous-cell systems, it is not possible to tell in general if the same set of organisms is produced by both systems.

(This follows from the last result and some additional constructions. It also follows as a direct translation of analogous well-known results for context-free languages.)

Definition: If a cell system can produce an identical organism in two different ways, the system is said to be *ambiguous*.

(Here "identical" means that the cell graphical structures are the same

and the states of cells at corresponding positions are the same. "Different" means not just derivable in some alternative chronological order, but that similarly positioned cells in the two identical final organisms will have developed from different ancestor cells in a derivation tree.)

Theorem: There is no algorithm which will tell of an arbitrary autonomous-cell system whether it is ambiguous.

(This follows directly from results in context-free language theory.)

Definition: An *inherently ambiguous* set of organisms is one in which every organism system generating that set of organisms is ambiguous.

Theorem: There exist inherently ambiguous sets of organisms. (In particular, there exist inherently ambiguous sets of linear autonomous-cell organisms.)

(The common example (translating from results in formal linguistics), is the set of tripartite "linear organisms" $a^i b^j c^k$ where $i = j$ or $j = k$ (that is, where the number of a-cells and b-cells is the same, or the number of b-cells and c-cells is the same.))

Theorem: The sets defined by linear sensitive-cell organism systems include the sets defined by the context-sensitive grammars.

Conjecture: The sets defined by context-sensitive grammars and the linear sensitive-cell organism systems are the same.

Corollary conjecture: The sets of linear sensitive-cell organism systems are accepted by linear bounded automata.

(It is clear that the sets generated by sensitive-cell organism systems include the sets generated by the context-sensitive grammars; this follows from the way these systems are defined. The organism sets may, however,

possibly produce sets which context-sensitive grammars do not; the differentiation of cells may change the context of other cells, permitting propagation rules to be applied. It is not clear whether this will produce any fundamental differences in the systems. I conjecture not.)

Theorem: The sets of autonomous-cell organisms are properly contained in the sets of sensitive-cell organisms.

(This follows of course from results for the linear cases of formal linguistics.)

Theorem: The sets of organisms produced by general organism systems can be generated by a Turing machine.

(With appropriate coding, and accepting the Church-Turing thesis, a Turing machine can carry out all the effective procedures, including, in this case, applying a finite set of rules in all possible allowable ways.)

Theorem: There is no uniform algorithm which will decide, given an arbitrary general organism system and a particular organism, whether the organism is a consequence of the system.

(General formal organism systems (which allow regressive rules) will permit the expression of formal systems known to have universal computational properties. Systems with these universal computational properties possess various unsolvable decision problems. One of these is the derivation problem. Another unsolvable decision problem in such formal systems is that of foretelling whether a particular "word" or putative theorem will ever be a consequence of the system.)

Theorem: The sets of sensitive-cell organisms are properly contained in the sets of general organisms.

(The derivability problem is always solvable for the sensitive-cell systems. This means that the sensitive-cell systems produce only recursive sets. The general organism systems produce the recursively enumerable sets. (It

is known that there are recursively enumerable sets which are not recursive.))

Theorem: Given any general organism system, there is no algorithm which will always decide if the organism being produced by the system will ever stop growing.

(In particular there exist deterministic general organism systems (defining the growth of single organisms) for which the "halting" problem of Turing machine theory is not decidable.)

Theorem: Among the general organism systems are systems which can reproduce themselves.

(We already have some sorts of reproduction present in our weaker organism systems: we have taken the act of a cell producing a cell as a primitive operation in our system. With general organism systems, however, reproduction of a complete multi-cellular organism is clearly possible. (Indeed, in a sense, the reproduction of any and all multi-cellular formal organisms is possible, since the general organism systems can express any effective procedure.)

More particularly, since general organism systems permit the application of regressive rules, we can cause a cell to be separated from an organism. If the released cell is in the initial fertile state, then if the cell is subject to the identical or a similar environmental history as the parent, an identical or a similar offspring organism results.

Also, for some sufficiently complex general organisms, an offspring can be created piecemeal under complete control of the parent organism, the child organism taking on much or all of the adult form of the parent before separation takes place. This is so because the general organism systems have been designed to allow the embedding of such systems as Turing machines, Post Normal Systems, (Tag systems, Lag systems, etc.) and the

von Neumann cellular system, all of which have long been known to possess universal computational properties, universal description properties, and (with varying degrees of expressive power) universal construction properties (and many concomitant unsolvable properties, in particular, unsolvable halting and derivability problems)).

There are numerous further results from automata theory and formal linguistics which could be easily translated to our organism systems. Some of these results may have some bearing on real biological problems; most, however, would not, and so there is no pressing need to re-phrase them in a biological terminology.

We will now try to exploit what we have so far established, trying to get some results of consequence to biology.

Notice that the series of proper inclusions among organism systems we have pointed out, gives us a *hierarchy* of increasingly more powerful systems: the regular organisms are contained in the autonomous-cell organisms, the autonomous-cell organisms are contained in the sensitive-cell organisms, and the sensitive-cell organisms are contained in the general organisms.

We are now in a position to consider particular real biological systems and to try to decide where they fit in this hierarchy.

General Considerations

We have so far discussed some of the relationships that exist among our classes of formal organisms, and between our formal organisms on the one hand and formal languages on the other. We have discussed only briefly the relationship between our formal organisms and real organisms. This is the relationship of most interest to us here, and will be considered in this section.

A form of question we shall repeatedly ask is: can all and only the forms assumed by a certain class or species of biological organisms be accounted for by a particular formal organism system? A rather cosmically vast form of this question is to ask whether we can get *all* (and only) forms of living organisms by means of formal organism systems. Even attacking this problem would first of all require sufficient insight into the structure of organisms to provide a satisfactory formal definition of the notion of "all organism forms". This is a very complex and at best, only intuitively grasped concept. (At times it almost seems that the vast variety of forms of living things would make it impossible to bar anything with finality.)

On the other hand, we should be able to say that all organism forms can be accounted for by use of unrestricted re-writing systems (such as our general organism systems). I believe that all adequate explanations of growth and development can be given in automaton terms, and unrestricted re-writing systems can exhibit all automaton processes.

This of course is not very revealing; we should like to know *which* sub-classes of all possible re-writing systems produce *which* biological forms.

Notice that although unrestricted re-writing systems (of our general organism systems) can in some sense express everything, including all organism forms, these systems may fail to be useful to biology because of the *manner* in which they express biological phenomena. At any moment it may be difficult to tell what is going on, or what the biological relevance may be. Also note that more powerful systems are not necessarily more desirable. In attaining a particular goal, in general, the weaker the

assumptions adopted, the more revealing and significant the result. We should accept the rule of economy and should try to prove as much as possible employing only the weaker rules, and only grudgingly employ stronger rules.

One tactic which has been extensively employed in formalizing linguistic theory, is to try to find the weakest system which can account for some linguistic phenomenon. If a particular system can be proven to be inadequate, we then attempt to show the adequacy of the next stronger system in the hierarchy.

A formal system is shown to be inadequate if a counter-example can be found. Thus in formal grammar theory the regular languages were shown to be inadequate to account for *all* natural language syntactic characteristics, because there is a natural language process which permits indefinitely extended "bilaterally symmetric" sentences, and this process is not adequately described in terms of regular grammars. To resolve this difficulty, a broader class of languages, the context-free languages, was introduced, and it too was examined for adequacy, etc.

Notice that the phenomenon against which regular grammars was tested (the "extended bilateral symmetry") was arrived at through observational and intuitive considerations of real language phenomena, that is, by the processes of empirical investigation.

In the next sub-section we shall examine some of the biological formalisms introduced in the last section and try to show the adequacy or inadequacy of those formalisms for describing the development of real biological organisms.

Tentative Results for Biology

In this section we set forth our principal results in a skeletal

form. The statements are labeled *Empirical Observation* if I believe the statement is supported by biological evidence. A statement is labelled *Observation* if it states a simple logical conclusion, or if it recalls to mind a simple fact, or results obtained earlier. The label *Theorem* is used if the assertion made has mathematical or logical content and is likely to be novel or unexpected.

The "weakest" formal organism system we have defined is the linear regular organism system.

Empirical Observation: There are some simple plant-like organisms (some algae) which are linear and grow only at the tips.

Observation: The linear regular organism systems may be useful in describing development of some simple biological organisms.

Empirical Observation: There are some simple *branching* organisms with tip growth only.

Observation: The *linear* regular organism systems are inadequate for describing the development of branching organisms.

Observation: The regular organism systems can express branching growth.

Empirical Observation: There are linear and branching organisms which inherently possess (potential) indefinitely extendible "internally controlled" bilateral symmetry.

Theorem: The regular organisms cannot express inherent indefinitely extendible "internally controlled" bilateral symmetry.

Theorem: The regular organism systems are inadequate to express all organism developmental phenomena.

Observation: The autonomous-cell systems are capable of expressing indefinitely extended "internally controlled" bilateral symmetries.

Empirical Observation 1: "Communication" between cells is known to exist: cells are affected by the electro-chemical and hormonal activities of other cells.

Empirical Observation 2: Communication between cells *must* exist to account for certain phenomena. (e.g., gastrulation, or repair of a wound *does* occur; the closing tissue stops "at the right place". This phenomenon requires inter-cell communication.)

[I have made two "empirical observations" here in order to bring out some useful distinctions. It is (abstractly) possible that an organism system might possess some property (such as cell-sensitivity) but that the property might not take part in any organism process. This property might be termed a "gratuitous complexity" of the system. (We would of course, in any real system, try to find out how this property was acquired, and what purpose it might have served, and whether it really does still perform some function for the organism.)

It is also possible that an organism system possesses some property, and that it *does* make use of this property, but that an identical ultimate outcome could be achieved even if the property were not employed. For example, a sensitive-cell system might make use of the sensitive cell property in producing some particular organism forms. It is possible that a much weaker system (for example, a regular organism system) could also produce

the same organism form. In such a case, it is clear that the two organism systems might employ different means to the same end; the only advantage the stronger system might possess is the ability to produce the organism faster or more efficiently.

Finally, it is possible that an organism system possesses some property, and the use of this property is *necessary* if certain phenomena or organism forms are to be produced at all. For example, it seems likely that cell-sensitivity is absolutely necessary to proper wound closure. Another property of sensitive-cell systems not possessed by weaker systems is the ability to generate certain cell-structures having a fixed tripartite ratio of *separated* element kinds. The linear form $a^n b^n c^n$ is a simple example.

Many interesting problems arise out of these distinctions. For example, some organisms achieve their mature form in two stages, a first stage of largely progressive growth, followed by a largely regressive growth stage of dissolution of cells and cell connections. A general cell organism system could be employed to mirror this growth and decay process. Can these same mature biological forms be produced by direct, progressive growth rules alone? That is, can sensitive-cell organism systems, with no rules for dissolution of cells and cell connections, produce the same organism forms? What is the significance of these alternative routes to the same phenomena? Since many biological phenomena can undoubtedly be accounted for by several different formal organism systems, how do we decide which formal organism system best accounts for the phenomena?]

Observation: There is no communication between the cells of an autonomous-cell system.

Theorem: Autonomous-cell systems are inadequate to account for all developmental phenomena.

Observation: In sensitive-cell systems there is communication between cells.

Empirical Observation: Cell dissolution occurs as part of a "normal" developmental process.

Empirical Observation: Organisms release cells (e.g., especially germ cells).

Empirical Observation: Organisms "bud-off" and release completed (or nearly completed) organisms.

Observation: The sensitive-cell systems have no "regressive" rules: they have no rules of cell dissolution or detachment.

Theorem: The sensitive-cell systems are inadequate to account for all developmental phenomena.

Observation: The general-cell systems permit detachment and dissolution rules.

Theorem: The general-cell systems can exhibit all developmental processes including reproduction and self-reproduction.

This last rather dramatic theorem has less to it than meets the eye.

It merely affirms my contention that all applications of the developmental processes of real organisms will be mechanistic in nature and that the general-cell systems can (with varying degrees of naturalness) exhibit all mechanistic procedures. To exhibit a developmental phenomenon of a real organism that could not be also exhibited in a general-cell system would be either to discredit (which is unlikely) the Church-Turing thesis (which asserts that all such general systems do indeed capture all of the informal notion of *effective*) or to show that explanation of development in biology entails growth rules which are essentially non-machine-like (non-effective) in nature. This latter is possible, though I personally doubt it.

It should also be remembered that throughout this section we have been considering the properties of organisms possessing rule application optionality. If we were to require of our systems that every cell at every instant have some rule applied to it we would get a different set of organisms. For example, imposing this requirement, there is a regular organism system which could yield the $a^n b^n$ organisms, and an autonomous-cell organism system which could yield the $a^n b^n c^n$ set of organisms.]

Conclusion

We have reached the top of our hierarchy of formal organism systems. Although what has been presented here has been more of an outline of a research program than a report of completed research, I hope it is clear that automaton theory has considerable relevance for biological theory. (I also hope it is clear that very much more complicated formal organisms and processes could have been exhibited.) A tremendous amount remains to be done though.

We must point out once more that while our formalisms may be able, theoretically, to represent any developmental process of biology, they may so distort the picture of these processes that the formalizations will be worthless as explanatory scientific models. There undoubtedly will be cases where a general-cell system can indeed exhibit a developmental process, but the picture of it is thereby so distorted as to fail to satisfy the intuition of the reasonable skeptic.

We should perhaps at this point discuss briefly but more specifically the expressive capabilities and limitations of the formal organism systems we have been considering here.

In this paper we have concentrated on showing how these formal organism systems might express the morphogenetic processes of differentiation, and creation and dissolution of cells and cell contacts. The ultimate form assumed by organisms is also however a function of such physical properties as the orientation of cells, size and shape of cells, the adhesive properties of cells, and the movement of cells and cell materials. The systems we have proposed here can of course express all these processes computationally; they can also express some of these processes fairly directly. For example, the state of a cell-automaton in the formal model could (in addition to recording the internal state of a real cell) record the orientation, size and shape, and adhesive properties of the real cell represented. With this broader interpretation of the state of a cell we can define production rules which take these additional cell properties into account when new cells or cell complexes are to be generated.

For example, if the rules say in effect that a cell in a particular state can have at most five contiguous cell neighbors, and the neighbors must impinge at certain particular surface locations, then such rules could be said to take into account the particular size and shape possessed by a cell.

The movements of cells and cell materials can also be expressed by, for example, employing rules (similar to Chomsky's grammatical transformation rules) which produce re-arrangements of cells or blocks of contiguous cells.

A major weakness of these formal organism systems in their present state of development, then, is not that they can not express the morphogenetic processes of interest, but that we have not yet developed within these formal systems adequate ways of clearly specifying and distinguishing the particular morphogenetic processes.

If we wish to say something specific about the global effect of particular morphogenetic processes, then we shall have to somehow clearly distinguish these processes as they are embodied in the local production rules.

As we may want to distinguish the rule uses which reflect the effect of inherent organism environment from the rule uses which reflect the effect of the environment external to the organism, we may also want to distinguish rules which reflect the effect of cell size from rules which reflect, for example, the cell shape.

Also there is a genuine question whether if we incorporate more and more features into our organism systems in a more or less *ad hoc* fashion, we will not end by blurring much of the direct expressive power of our systems, and with this any intuitive feeling for the underlying nature of the processes we are investigating.

For this reason, it is important to explore ways in which these, and other formal organism systems which might be devised, can express, in an informative and revealing fashion, specific complex developmental phenomena of real biological systems of interest. Our "empirical observations" were of an extremely general and naive sort. Before the resources of automaton theory can be brought fully to bear on problems of developmental biology, automaton theorists will have to acquire more biological facts, and more biological sophistication. In continuing this program, the assistance of interested biologists will not only be very helpful, it will be absolutely indispensable.

Acknowledgements

Much of what I have said in this paper is a consequence of discussions, over recent years, with Michael Arbib (now of the Computer and Information Science Department of the University of Massachusetts), and with my colleagues at The University of Michigan (especially Arthur W. Burks and John H. Holland).

The research upon which this paper is based was supported in part by grants to The University of Michigan Logic of Computers Group from the National Institutes of Health and the National Science Foundation.

Bibliography

- Arbib, M.A. (1967) "Automata Theory and Development," *J. Theoret. Biol.*, 14, 131-156.
- ___, (1969) *Theories of Abstract Automata*, Prentice-Hall.
- Burks, A.W. (1970) (editor) *Essays on Cellular Automata*, University of Illinois Press.
- Chomsky, N. (1963) "Formal Properties of Grammars," *Handbook of Mathematical Psychology, Volume II*, Wiley, 323-418.
- Codd, E.F. (1968) *Cellular Automata*, Academic Press.
- Hopcroft, J. and Ullman, J. (1969) *Formal Languages and Their Relation to Automata*, Addison-Wesley.
- Kleene, S. (1956) "Representation of Events in Nerve Nets and Finite Automata," *Automata Studies*, Princeton, 3-42.
- Laing, R. (1969) "Formalisms for Living Systems," University of Michigan Report 08226-8-T.
- Lindenmayer, A. (1968) "Mathematical Models for Cellular Interactions in Development," *J. Theoret. Biol.*, 18, 280-299, 300-315.
- ___, (1971), "Developmental Systems without Cellular Interactions, their Languages and Grammars," *J. Theoret. Biol.*, 30, 455-484.
- McCulloch, W. and Pitts, Walter (1943) "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bull. Math. Biophysics*, 5, 115-133.
- Minsky, M. (1967) *Computation: Finite and Infinite Machines*, Prentice-Hall.
- Montanari, G.U. (1970) "Separable Graphs, Planar Graphs, and Web Grammars," *Information and Control*, 16, 243-267.
- Pfaltz, J. and Rosenfeld, Azriel (1969) "Web Grammars," *Proceedings of the International Joint Conference on Artificial Intelligence*, (ed. Walker and Norton), 609-619.
- Salomaa, A. (1970) "Periodically Time-Variant Context-Free Grammars," *Information and Control*, 17, 294-311.
- Smith, A.R. (1968) "Simple Computation-Universal Cellular Spaces and Self-Reproduction," *Record of Ninth Annual Symposium on Switching and Automata*, 269-277. Inst. Electrical and Electronic Engineers. New York.

Thatcher, J. (1969) "Transformations and Translations from the Point of View of Generalized Finite Automata Theory," *ACM Symp. on Theory of Computing*, 129-142.

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

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

UNIVERSITY OF MICHIGAN



3 9015 03466 4147