THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Communication Sciences Program

Technical Report

ON SOME CATEGORICAL ALGEBRA ASPECTS OF AUTOMATA THEORY: THE CATEGORICAL PROPERTIES OF TRANSITION SYSTEMS

Yehoshafat Give'on

ORA Projects 03105 and 07363

supported by:

U. S. ARMY RESEARCH OFFICE (DURHAM)
GRANT NO. DA-ARO(D)-31-124-G665
DURHAM, NORTH CAROLINA

and

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
CONTRACT NO. Nonr-1224(21)
WASHINGTON, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

February 1966

Engn
UMR
1532

This report was also a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in The University of Michigan, 1966.

RESEARCH PROGRESS REPORT

Title: "On Some Categorical Algebra Aspects of Automata Theory: The Categorical Properties of Transition Systems," Y. Give'on, University of Michigan Technical Report 03105-43-T.

Background: The Logic of Computers Group of the Communication Sciences Program of The University of Michigan is investigating the application of logic and mathematics to the design of computing automata. Applications of the techniques of category theory in the study of automata form a part of this investigation.

Condensed Report Contents:

The ubiquity and usefulness of homomorphisms in various studies of automata lead us to consider the following problem. What can be said on automata by referring only to homomorphisms of automata? In the report we present a study of this problem with respect to a special type of automaton, namely, with respect to transition systems.

Categorical algebra methods are applied to the precise formulation of this problem and to its solution. We find that if W is a monoid belonging to a broad class of monoids, then the categorical abstract properties of transition systems with input W, are determined by the automorphisms of the monoid W. In particular, any property of automata without output is categorical iff it does not depend on the particular labeling of the input alphabet.

This study of the categorical properties of automata has two additional outcomes. First, we realize that categorical algebra methods can be applied to automata with arbitrary input monoids, with results pertinent to the theory of monoids. On the other hand, it indicates a possible usefulness in the study of automata, in particular, in getting a better understanding of the mathematical ideas employed in automata theory.

In order to support this point of view with respect to automata theory, we show that many actually studied properties of automata are categorical. And we give an example of a categorical examination and formulation of a particular study of perfect automata.

For Further Information: The complete report is available in the major Navy technical libraries and can be obtained from the Defense Documentation Center. A few copies are available for distribution by the author.

ACKNOWLEDGEMENTS

I am deeply indebted to my thesis advisor, Professor Lawrence C. Eggan of the Pacific Lutheran University, Tacoma, Washington. I feel inept in expressing my gratitude to him for his guidance, encouragement and for his thorough and critical reading of the manuscript in all phases of its preparation.

I am also indebted to Professor John H. Holland for his constant support and aid during the three years that I worked with the Logic of Computers Group. I especially wish to thank him for administrating the necessary formalities without which this dissertation could not have been submitted.

Professor Hehoshua Bar-Hillel introduced me to automata theory and guided my early work in this field, at the Applied Logic Branch of the Hebrew University of Jerusalem. I feel honored by his readiness to serve on my committee, and I wish to thank him for his encouragement and help.

Dr. Armand Brumer and Dr. Richard M. Karp (of IBM Research Center, Yorktown Heights, New York) helped me in two polar aspects. Dr. Brumer guided me in the work on the extension of the generality of my results to a maximally possible broad class of input monoids, while Dr. Karp guided me in the applicational aspects of my work. I wish to thank both of them for the many invaluable and fruitful discussions that I had with them concerning these two aspects of my work.

I also wish to thank Professor Saunders MacLane of the University of Chicago, and Professor Peter Freyd of the University of Pennsylvania for their encouraging interest in the categorical aspects of my work.

I am deeply grateful to the staff of the Logic of Computers Group at The University of Michigan and especially to its director, Professor Arthur W. Burks, for the privilege and pleasure of working with them. Warm thanks are extended to Mrs. Ann M. Jacobs and Mrs. Kuniko Misawa for their faithful work on the typing and duplicating of the manuscript.

The research reported in this dissertation was conducted at the Logic of Computers Group and was supported by the Department of the Navy, Office of Naval Research, under contract no. Nonr-1224(21), Washington, D.C. and by the U.S. Army Research Office (Durham) under grant number DA-ARO(D)-31-124, Durham, North Carolina. In particular, I wish to thank Dr. Gordon D. Goldstein, of the Information Systems Branch of the U.S. Office of Naval Research for his generous and understanding support.

TABLE OF CONTENTS

TABLE OF CONTENTS (Concluded)

INTRODUCTION

Homomorphisms of automata were defined almost at the beginning of the development of "algebraic" automata theory. To the best of my knowledge, Büchi [2] was the first student of automata who defined homomorphisms of automata and, in an unpublished paper, drew attention to the importance of homomorphisms as a tool for the study of automata. Recent developments in automata theory show that the notion of homomorphisms of automata is, in fact, an efficient tool for solving various problems concerning structure and behavior of automata. This work is directed towards the study of the following problem which seems to be natural in this context. How much information about automata can be derived from the properties of their homomorphisms? We study this problem with respect to automata without output or designated states (i.e., with respect to <u>transition systems</u>) but with an arbitrary input monoid. Homomorphisms of such transition systems can be defined in various ways, and we choose the simplest type of homomorphism which is the type that is commonly used in automata theory. Informally speaking, we discuss homomorphisms of transition systems which are defined as mappings on the sets of states of transition systems with the <u>same</u> input monoid, and which preserve the effect of the input on the states (i.e., which preserve the transition functions). Thus, for any monoid W, we have the class of all transition systems with input W, and the class of homomorphisms of such systems. Our problem can be described now as follows. Given an arbitrary input monoid W, what can be said about the transition systems with input W by referring only to their homomorphisms and the algebraic nature of their composition?

Before we turn to solve this problem, we have to make precise what is meant by "referring only to homomorphisms and the algebraic nature of their composition". Since a similar problem can be raised with respect to any type of mathematical system for which homomorphisms are defined, it is clear that we have to explicate the meaning of such a problem in its most general setting. The most natural framework for this explication is categorical algebra. However, in spite of the fact that one can find in the literature some references to particular phenomena which are related to our problem (e.g. [8; pp. 9, 28]), nobody has examined before the general idea of properties of mathematical systems which are "categorical", i.e., which are definable "in terms of properties of objects in an abstract category".

Our intuitions with respect to such properties of mathematical systems lead us to consider two explications for the notion of "categorical" predicates. In Part I of our work, we explain the motivation of each one of these two explications. We give a proof which shows that these two explications are equivalent. We regard this equivalence as a justification for our intuitions and we state as a thesis that the categorical predicates in any category are precisely the predicates which are of the type defined in either one of the explications. To be specific, our thesis is that the categorical predicates in a category $\mathbb{C}$ are those predicates which are invariant under all the automorphism functors of $\mathbb{C}$. Our problem concerning the homomorphisms of transition systems with input W can be stated now as follows. What are the categorical predicates in the category of transition systems with input W, where W is an arbitrary monoid?

Before we turn to the study of the categorical predicates of transition systems, we elaborate on the general notion of categorical predicate. It is clear now that the automorphisms of a category $\mathbb{C}$ determine the categorical predicates in $\mathbb{C}$. Following some work of Freyd

[8; pp. 28–34] with relation to automorphisms of categories, and which is motivated by the intuitive idea of categorical predicates, we derive several general methods for the determination of the categorical predicates in various types of categories. We indicate the applicability of these methods to various well known categories of mathematical systems.

Provided with the results established in Part I, we turn in Part II to the study of the categorical predicates of transition systems. After a preliminary study of the categories of transition systems, we establish a reduction of a characterization of all the categorical predicates in the category of transition systems with input W to the categoricity of a single class of transition systems. We find that for the transition system M_W, which is W itself regarded as a transition system with W as input, if we can characterize it up to an isomorphism by means of a categorical predicate, then the categorical predicates of transition systems and their homomorphisms have a very simple characterization. In particular, in this case, a class of transition systems with input W which is closed under isomorphisms is categorical iff it is closed under modifications of transition systems which are induced by automorphisms of the monoid W.

Our attention is now drawn to the problem of characterizing M_W by means of a categorical predicate. We end Part II with a further study of the categories of transition systems which is directed towards the solution of this problem. We find that under very broad conditions on W, the object M_W is characterized categorically as a special kind of a projective generator. In particular, if W satisfies the following property:

$$\text{if } uv = 1 \text{ in } W \text{, then } vu = 1 \text{,}$$

then M_W is characterized up to an isomorphism by means of a categorical predicate. We note that if W belongs to either one of the following types of monoids: finite monoids, groups, abelian monoids, free monoids or cancellative monoids, then W has the above mentioned property. Thus, for a very broad class of monoids, a class which includes all the types of input monoids employed in automata theory, the categorical predicates of transition systems with input monoid belonging to this class is completely characterized by the automorphisms of the input monoid. The problem whether M_W is categorically characterizable for any W is still open.

We note that in order to derive this result concerning the characterization of M_W, we derive several interesting properties of the category of transition systems with input W. They are interesting not only from the point of view of categorical algebra, but also because they indicate a possible application of such a study as the one presented in this work to the development of the theory of monoids. In addition to the possible general mathematical attractiveness of these categories, we believe that an application of categorical algebra methods to studies of automata can be proven to be as useful as it has proved to be in obtaining a better understanding of many mathematical ideas. Because of this belief, we add to this work an additional part which gives indications of the relevance of our results to actual automata theory. In Part III, therefore, we review quite a variety of properties which are actually studied in automata theory, and prove that, when we restrict our attention to transition systems, they turn out to be categorical. We end Part III, and the whole work, with a categorical reformulation and examination of a specific study of a certain type of automata which is originated by Fleck's [7] study of the automorphism groups of automata.

In order to make our work self-contained, we introduce in the beginning of Part I some of the basic notions of categorical algebra. We assume however some elementary knowledge of group theory, monoid theory, set theory and the fundamental concepts of algebra (see, for example, [3]). A certain familiarity with automata theory is helpful especially for understanding the motivation of our work and the discussions presented in Part III. In general, we use the ordinary notation and conventions used in algebra. We make a frequent use of the following scheme for explicit descriptions of functions. By

$$f : T_1 \longrightarrow T_2 : t_1 \longrightarrow t_2$$

we always mean that f is a function from T_1 to T_2 such that for all $t_1 \in T_1$ we have $f(t_1) = t_2$. Furthermore, if $f : T_1 \longrightarrow T_2$ is a function, then T_1 is the <u>domain</u> of f, T_2 is the <u>range</u> of f, and the class of all elements of T_2 of the form $f(t_1)$ for all $t_1 \in T_1$ is the <u>image</u> of f. The function $f : T_1 \longrightarrow T_2$ is said to be <u>injective</u> iff $f(t_1) = f(t_2)$ always implies $t_1 = t_2$. It is <u>surjective</u> iff the image of f is equal to its range, and it is <u>bijective</u> iff it is both injective and surjective. We will refer to functions from classes to classes as <u>assignments</u>, while the term function will be kept to functions from sets to sets.

PART I

PRELIMINARIES OF CATEGORICAL ALGEBRA

INTRODUCTION TO PART I

Part I contains the categorical algebra material which is necessary for the fomulation and the study of the problems stated in intuitive terms in the introduction. It consists of two chapters. In Chapter I we present the basic notions of categorical algebra. For a more extensive discussion of categorical algebra the reader is referred to the literature (especially, Freyd [8] and MacLane [13]).

In Chapter II we give two explications for the notion of categorical predicates, which as we prove, turn out to be equivalent. The main tool for the determination of the categorical predicates in a given category $\mathbb{C}$ is the class of the equivalence classes of the automorphisms of $\mathbb{C}$ under natual equivalence. Since this class has a group structure, we call it the <u>automorphism class group</u> of $\mathbb{C}$, and we denote it by $\mathrm{Aut}(\mathbb{C})$. We end this part of the work by providing a method for the computation of $\mathrm{Aut}(\mathbb{C})$ which is applicable to many well known categories of mathematical systems.

CHAPTER I: SOME BASIC NOTIONS OF CATEGORICAL ALGEBRA

1. Categories

The notion of category is defined in order to present a unified framework in which certain methods universally applicable to many categories can be formulated and developed in the most economical and abstract setting. In its most abstract definition, a category $\mathbb{C}$ is a class of morphisms in which a binary operation, the composition of the morphisms of $\mathbb{C}$, is partially defined and is subject to the following axioms.

Axiom I. Associativity:

(i) The composition $h(gf)$ is defined iff $(hg)f$ is defined;

(ii) if $h(gf)$ is defined, then $h(gf) = (hg)f$; we write $h(gf) = hgf$;

(iii) hgf is defined iff both hg and gf are defined.

Axiom II. Identities: A morphism i of $\mathbb{C}$ is said to be an identity morphism (in $\mathbb{C}$) iff $if = f$ whenever if is defined and $gi = g$ whenever gi is defined. We require:

(iv) for each morphism f of $\mathbb{C}$, there exist identity morphisms i and i' in $\mathbb{C}$ such that both $i'f$ and fi are defined.

In spite of the seemingly abstract remoteness of this notion of a category, it is not difficult to verify the following examples of categories.

The category of sets, $\mathbb{S}$, is the category of the triples (f,T_1,T_2), where T_1 and T_2 are any sets and $f : T_1 \longrightarrow T_2$ is any function from T_1 and T_2. The composition rule of $\mathbb{S}$ is determined by the ordinary composition of function with the additional restriction that the composition is defined iff the appropriate range is identical with the appropriate domain, i.e., $(f,T_1,T_2)(g,S_1,S_2)$ is defined iff $S_2 = T_1$, and then it is equal to (fg,S_1,T_2). Once this restriction is accepted, we write

$f : T_1 \longrightarrow T_2$ instead of (f, T_1, T_2).

The reader is referred to the literature (MacLane [13; p.48]) for a discussion on the role of this restriction imposed on the composition rule of $\mathbb{S}$ and other categories. In ordinary terms, this restriction amounts to the requirement that every morphism in any category has a unique domain and a unique range.

The category of monoids, $\mathbb{M}$, is the category of all triples (H, M_1, M_2), written as $H : M_1 \longrightarrow M_2$, where M_1 and M_2 are arbitrary monoids and H is a monoid homomorphism from M_1 to M_2. The composition rule of $\mathbb{M}$ is defined similarly to the composition rule of $\mathbb{S}$, with the analogous restriction imposed on the ranges and the domains of the composed morphisms.

Similarly we have the category $\mathbb{G}$ of groups and their homomorphisms, and the category $\mathbb{T}$ of topological spaces and continuous maps.

In order to realize that the morphisms of a category need not be functions or related to functions, consider the following example. Let π be a reflexive and transitive relation defined on a class P. We define a category $\mathbb{P}_\pi$ with morphisms (p_1, p_2) whenever $(p_1, p_2) \in \pi$. The composition $(p_1, p_2)(q_1, q_2)$ is defined iff $p_2 = q_1$, and then it is equal to (p_1, q_2). $\mathbb{P}_\pi$ thus defined is a category.

Usually, a category is defined in mathematics when a certain kind of mathematical object is studied and the morphisms of the desired category are defined as maps from objects to objects. Thus, historically speaking, groups preceded group homomorphisms and linear spaces preceded linear transformations. Categorical algebra, however, is the mathematical manifestation of the point of view that, logically, functions are more basic than sets and group homomorphisms are more basic than groups. Our previous definition of categories expresses this attitude by referring to the morphisms of a category and their composition rule as primitive notions. There is another definition of categories in which the objects under study seem to play an equally important role as the maps.

A category $\mathbb{C}$ with objects is defined as a class of objects $A,B,C,\ldots$; together with a family of disjoint classes of morphisms $\mathbb{C}(A,B)$, one for each ordered pair (A,B) of objects of $\mathbb{C}$, and a family of functions

$$\mathbb{C}(A,B)\times\mathbb{C}(B,C) \longrightarrow \mathbb{C}(A,C) ,$$

one for each ordered triple (A,B,C) of objects of $\mathbb{C}$. We denote the image of the ordered pair $(f,g) \in \mathbb{C}(A,B)\times\mathbb{C}(B,C)$ under this function by gf, and call it the composition of f and g in $\mathbb{C}$. Also we write $f : A \longrightarrow B$ for $f \in \mathbb{C}(A,B)$, and call A the domain of f, and B the range of f. Thus the composition gf is defined iff the range of f is the domain of g. We further require the following axioms:

Associativity: If $f : A \longrightarrow B$, $g : B \longrightarrow C$ and $h : C \longrightarrow D$ then $h(gf) = (hg)f$.

Identities: For any object B, there exists a morphism $i_B : B \longrightarrow B$ such that for any $f : A \longrightarrow B$ and $g : B \longrightarrow C$ we have

$$i_B f = f \quad \text{and} \quad g i_B = g .$$

The formal difference between categories with objects and categories as defined before is obvious; categories in general do not need to have objects. From the point of view of categorical algebra this difference is only formal. It is clear that the class of all the morphisms of category $\mathbb{C}$ with objects satisfies the axioms specified in our first definition of a category (with morphisms only). On the other hand, the identity morphisms of any category $\mathbb{C}$ can be regarded as objects for $\mathbb{C}$ in such a way that with these objects $\mathbb{C}$ becomes a category with objects according to our second definition.

In order to do so, note that in a (morphism) category $\mathbb{C}$, for any morphism f there exists precisely a single identity morphism i such that if is defined. Denote this identity morphism by $i_{R(f)}$. Similarly, there exists precisely a single identity morphism i such that fi is defined, and denote it by $i_{D(f)}$. We define $\mathbb{C}(i_{D(f)}, i_{R(f)})$, for each morphism f of $\mathbb{C}$, as the class of all morphisms g of $\mathbb{C}$ such that $i_{R(f)} g i_{D(f)}$ is defined in $\mathbb{C}$. It is easy to verify that for any category $\mathbb{C}$, the class of identity morphisms of $\mathbb{C}$ as objects together with the classes $\mathbb{C}(i_{D(f)}, i_{R(f)})$ for all the morphisms f of $\mathbb{C}$, and the given composition rule of $\mathbb{C}$ form a category with objects.

From the point of view of categorical algebra, objects of a category are but indices of their identity morphisms, and the main attention is thus directed towards the morphisms and their composition rule. For a given category $\mathbb{C}$, we will refer to the algebraic structure determined on the class of the morphisms of $\mathbb{C}$ by the rule of the composition of the morphisms of $\mathbb{C}$ by the term the <u>partial monoid of $\mathbb{C}$</u>. On the other hand we will always assume that categories have objects.

Set theoretic conscience sometimes imposes an additional axiom on categories; namely, that for any objects A and B, the class $\mathbb{C}(A,B)$ must be a set. Note that the categories $\mathbb{S}$, $\mathbb{M}$, $\mathbb{G}$, $\mathbb{T}$ and $\mathbb{P}_\pi$ all satisfy this axiom. We will not, however, require this of the categories that we intend to mention or study.

2. The basic morphism types

The basic morphism types are defined for an arbitrary category $\mathbb{C}$ according to their properties as elements in the partial monoid of $\mathbb{C}$. Whenever one encounters an associative binary operation one is curious to determine the invertible elements and the cancellable elements with respect to the given operation.

For example, a morphism $f : T_1 \longrightarrow T_2$ of $\mathbb{S}$, the category of sets, is invertible iff f is a bijection; it is left cancellable iff f is an injection, and it is right cancellable iff f is a surjection. The category $\mathbb{G}$ of groups and the category $\mathbb{T}$ of topological spaces are similar to $\mathbb{S}$ in this respect.

Following the terminology of group theory, we call an invertible morphism $j : A \longrightarrow B$ of $\mathbb{C}$ (i.e., such that there exists a morphism j' of $\mathbb{C}$ such that both jj' and $j'j$ are identity morphisms) an isomorphism in $\mathbb{C}$, and A and B are said to be isomorphic in $\mathbb{C}$.

A morphism $j : A \longrightarrow B$ of $\mathbb{C}$ is called monic iff it is left cancellable in $\mathbb{C}$ (i.e., iff $jf_1 = jf_2$ always implies $f_1 = f_2$). In a dual manner $e : A \longrightarrow B$ of $\mathbb{C}$ is called epic iff it is right cancellable in $\mathbb{C}$.

Note that in $\mathbb{M}$, the category of monoids, the isomorphisms have bijective underlying functions, and the monic morphisms are injective. In $\mathbb{M}$ every surjective morphism is epic, but not conversely. For example, the inclusion morphism of the additive monoid of natural numbers into the additive monoid (group) of integers is both monic and epic.

3. Functors

Categories are introduced whenever homomorphisms are more important than objects. Thus categories being the objects studied in categorical algebra give rise to functors which are nothing but homomorphisms of categories.

Before we give an explicit definition of functors, we discuss a typical example. In the category $\mathbb{G}$ of groups, every object G has a carrier $S(G)$, which is the set of elements of G, and every morphism $f : G_1 \longrightarrow G_2$ is determined by an underlying function $S(f) : S(G_1) \longrightarrow S(G_2)$ that has some properties. The common notation does not distinguish between G and $S(G)$ and between f and $S(f)$. The passage from groups to their carriers and from group homomorphisms to their underlying function is therefore a twofold assignment S from $\mathbb{G}$ to $\mathbb{S}$ which has the following properties:

(i) for any object G of $\mathbb{G}$, $S(G)$ is an object of $\mathbb{S}$;

(ii) for any morphism $f : G_1 \longrightarrow G_2$ of $\mathbb{G}$, $S(f) : S(G_1) \longrightarrow S(G_2)$ is a morphism of $\mathbb{S}$;

(iii) $S(i_G) = i_{S(G)}$;

(iv) $S(gf) = S(g)S(f)$.

Functors as assignments on morphisms are homomorphisms of partial monoids; i.e., they map identity morphisms on identity morphisms and preserve the composition of morphisms. For categories with objects, say $\mathbb{C}_1$ and $\mathbb{C}_2$, a functor $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ is a twofold assignment, one from the objects of $\mathbb{C}_1$ to those of $\mathbb{C}_2$, and one from the morphisms of $\mathbb{C}_1$ to those of $\mathbb{C}_2$, both of them are denoted by F and satisfy:

(i) if $f \in \mathbb{C}_1(A,B)$ then $F(f) \in \mathbb{C}_2(F(A),F(B))$;

(ii) $F(i_A) = i_{F(A)}$;

(iii) $F(gf) = F(g)F(f)$.

Thus, if we denote by $\mathbb{C}^{\mathbb{B}}$ the class of all functors $F : \mathbb{B} \longrightarrow \mathbb{C}$ from a category $\mathbb{B}$ to a category $\mathbb{C}$, then with the evident definition of composition of functors

$$\mathbb{B}^{\mathbb{A}} \times \mathbb{C}^{\mathbb{B}} \longrightarrow \mathbb{C}^{\mathbb{A}},$$

we get a category $\mathbb{C}\mathrm{at}$ whose objects are arbitrary categories and whose morphisms are functors. Note that since $\mathbb{S}^{\mathbb{S}}$ is not a set, the category $\mathbb{C}at$ of all categories does not satisfy the additional set-theoretic axiom.

Again, since we are dealing with homomorphisms of a certain kind of mathematical system, a classification of some types of functors is in place. The identity assignment $I_{\mathbb{C}}$ on the category $\mathbb{C}$ (i.e., $I_{\mathbb{C}}(A) = A$ and $I_{\mathbb{C}}(f : A \longrightarrow B) = (f : A \longrightarrow B)$,) is a functor $I_{\mathbb{C}} : \mathbb{C} \longrightarrow \mathbb{C}$, which is an identity morphism in $\mathbb{C}at$ and called the <u>identity functor</u> of $\mathbb{C}$. Obviously every identity morphism of $\mathbb{C}at$ is an identity functor. An invertible functor is called an <u>isomorphism of categories</u>, and an isomorphism $F : \mathbb{C} \longrightarrow \mathbb{C}$ is called an <u>automorphism</u> of $\mathbb{C}$. The monic morphisms of $\mathbb{C}at$ are the functors which are injective assignments, and the epic morphisms are the surjective functors.

More important and useful than the notions of injective and surjective functors are the notions of "locally injective" and "locally surjective" functors. A functor $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ is said to be an <u>embedding</u>, or <u>locally injective</u>, iff for any objects A and B of $\mathbb{C}_1$, the assignment

$$F : \mathbb{C}_1(A,B) \longrightarrow \mathbb{C}_2(F(A),F(B))$$

is injective. Similarly, a functor $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ is said to be <u>full</u>, or <u>locally surjective</u>, iff F restricted to $\mathbb{C}_1(A,B)$ is onto $\mathbb{C}_2(F(A),F(B))$. It must be clear that embedding functors need not be injective nor that full functors need be surjective.

The <u>forgetful</u> functor $S : \mathbb{G} \longrightarrow \mathbb{S}$, which was discussed at the beginning of this section, is obviously an embedding functor which is not injective; several groups may have the same carrier. Closely related to $S : \mathbb{G} \longrightarrow \mathbb{S}$ is the functor $Fr : \mathbb{S} \longrightarrow \mathbb{G}$ defined as follows.

For any set T we define $Fr(T)$ to be the free group generated by T, and for $f : T_1 \longrightarrow T_2$ a morphism of $\mathbb{S}$ define $Fr(f) : Fr(T_1) \longrightarrow Fr(T_2)$ to be the unique group homomorphism whose underlying function is an extension of f.

The relation between $S : \mathbb{G} \longrightarrow \mathbb{S}$ and $Fr : \mathbb{S} \longrightarrow \mathbb{G}$ is expressed by the universality property of the free groups. For any set T, denote by $g_T : T \longrightarrow SFr(T)$ the function which identifies every element of T as a generator of $Fr(T)$ (i.e., the inclusion of T in the carrier of $Fr(T)$). Now for any function $f : T \longrightarrow S(G)$, from T into the carrier of any arbitrary group G, there exists a unique group homomorphism $f^* : Fr(T) \longrightarrow G$ such that

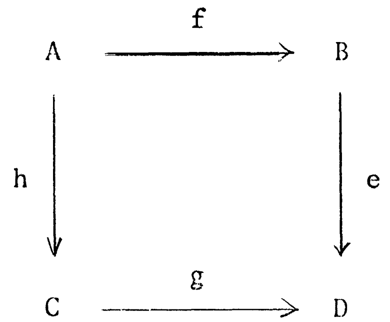
$$S(f^*)g_T = f .$$

An analogous pair of functors exists for the category $\mathbb{M}$ of monoids. $S : \mathbb{M} \longrightarrow \mathbb{S}$ is the forgetful functor of $\mathbb{M}$ which assigns to each monoid M its carrier $S(M)$, and $Fr : \mathbb{S} \longrightarrow \mathbb{M}$ is the free object functor of $\mathbb{M}$ which assigns to each set T the free monoid $Fr(T)$ generated by T. The free monoids also have the universality property.

Note that the free object functors of $\mathbb{G}$ and of $\mathbb{M}$ are both injective. We will return to discuss their properties and in particular their relationship to the forgetful functors in the sequel.

Another example of functors is the well known "commutative diagrams" analyzed as follows. A category in which the class of all morphisms, and therefore the class of all objects, are sets is called small. Small categories can be viewed as diagram schemes and specific diagrams in a category $\mathbb{C}$ with the scheme $\mathbb{D}$ are specific functors from $\mathbb{D}$ to $\mathbb{C}$.

For example consider the commutative square



of morphisms of $\mathbb{C}$ where $ef = gh$. The scheme of this square is a category $\mathbb{S}q$ with four objects a, b, c and d . $\mathbb{S}q$ has the following morphisms in addition to the identity morphisms:

$$\phi : a \longrightarrow b \ , \ \varepsilon : b \longrightarrow d \ , \ \eta : a \longrightarrow c \ , \ \gamma : c \longrightarrow d$$

and $\delta : a \longrightarrow d$, with $\delta = \varepsilon\phi = \gamma\eta$. The particular commutative square in $\mathbb{C}$ is thus given by the functor $F : \mathbb{S}q \longrightarrow \mathbb{C}$ with $F(a) = A$, $F(b) = B$, etc.

The category $\mathbb{C}at$ has an important automorphism D which maps every category $\mathbb{C}$ on its opposite (or dual) category $\mathbb{C}^{op}$. The objects of $\mathbb{C}^{op}$ are the objects of $\mathbb{C}$, but the morphisms $f^{op} : A \longrightarrow B$ are in one-one correspondence with the morphisms $f : B \longrightarrow A$ of $\mathbb{C}$. The composition $f^{op}g^{op}$ is defined in $\mathbb{C}^{op}$ iff gf is defined in $\mathbb{C}$ and then we set $f^{op}g^{op} = (gf)^{op}$.

Sometimes the opposite category of a given category $\mathbb{C}$ is isomorphic to $\mathbb{C}$. For example $\mathbb{S}q$ and $\mathbb{S}q^{op}$ are isomorphic categories, but this is not the general case. Every assignment from a category $\mathbb{C}_1$ to a category $\mathbb{C}_2$ induces in a one-one manner an assignment from $\mathbb{C}_1^{op}$ to $\mathbb{C}_2$. If $\mathbb{C}_1^{op}$ is isomorphic to $\mathbb{C}_1$ then every functor from $\mathbb{C}_1$ to $\mathbb{C}_2$ induces in this manner a functor from $\mathbb{C}_1^{op}$ to $\mathbb{C}_2$. In order to cover the case where $\mathbb{C}_1^{op}$ is not isomorphic to $\mathbb{C}_1$, we define an assignment F from $\mathbb{C}_1$ to $\mathbb{C}_2$ to be a contravariant functor iff the induced assignment from $\mathbb{C}_1^{op}$ to $\mathbb{C}_2$ is a functor. For example, the obvious assignment $\mathbb{C} \longrightarrow \mathbb{C}^{op}$ is a contravariant functor induced by the identity functor of $\mathbb{C}^{op}$. The contravariant functors are the "anti-homomorphisms" of categories.

We conclude our discussion of functors with the hom-functors. These are the functors which are defined for any category $\mathbb{C}$ which satisfies the set-theoretic axiom by considering the sets $\mathbb{C}(A,B)$ in the following manners. First, every object A of $\mathbb{C}$ determines a functor

$$H_A : \mathbb{C} \longrightarrow \mathbb{S}$$

with

$$H_A(B) = \mathbb{C}(A,B) ,$$

for any object B of $\mathbb{C}$. Any morphism $f : B \longrightarrow C$ of $\mathbb{C}$ induces a function

$$f_* : H_A(B) \longrightarrow H_A(C)$$

defined by

$$f_*(g : A \longrightarrow B) = (fg : A \longrightarrow C) ,$$

and $H_A(f)$ is defined as f_*.

Thus H_A is the functor derived from fixing A in $\mathbb{C}(A,B)$ and regarding B as a variable. It is therefore often denoted by $\mathbb{C}(A,-)$. If instead of B we regard A as a variable in $\mathbb{C}(A,B)$, we get a contravariant functor

$$H^B : \mathbb{C} \longrightarrow \mathbb{S} ,$$

with

$$H^B(A) = \mathbb{C}(A,B)$$

and

$$H^B(g : A \longrightarrow C) = (g^* : H^B(C) \longrightarrow H^B(A)) ,$$

where g^* is defined by

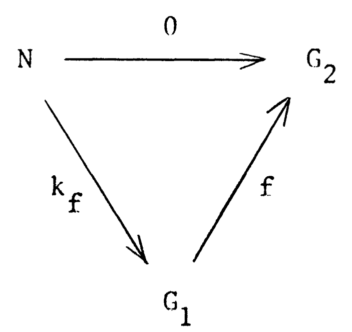
$$g^*(f : C \longrightarrow B) = (fg : A \longrightarrow B) .$$

Similar to H_A, H^B is often denoted by $\mathbb{C}(-,B)$.

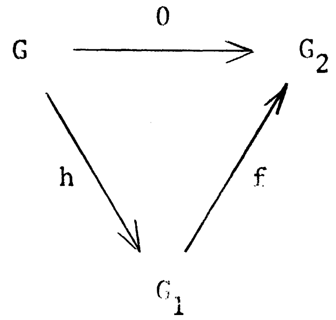
4. Universal constructions in categories

We start with an example from group theory. Kernels of group homomorphisms are usually defined "element-wise"; the kernel of $f : G_1 \longrightarrow G_2$ is the group N of all the elements g of G_1 for which $f(g)$ is the identity element of G_2. However, in practice, kernels are employed in group theory because of their universality properties expressed by certain group homomorphisms.

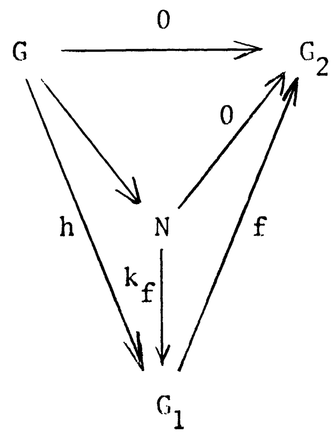
Let $k_f : N \longrightarrow G_1$ be the inclusion monomorphism of N in G_1, where N is the kernel of $f : G_1 \longrightarrow G_2$, and let $0 : N \longrightarrow G_2$ be the trivial homomorphism which maps all of N on the identity element of G_2. We clearly have the commutative triangle



which is universal in the following sense. For any commutative triangle of the form



where $0 : G \longrightarrow G_2$ is again the trivial homomorphism, there exists a unique homomorphism $G \longrightarrow N$ which connects the two triangles into a commutative diagram



Put intuitively, every homomorphism k into G_1 which is "killed" by $f : G_1 \longrightarrow G_2$ is factored uniquely through $k_f : N \longrightarrow G_1$. This property follows immediately from the fact that $fh = 0$ implies that the image of h is included in the kernel of f.

The basic idea which underlies this property, and as we will see later also the universality propety of the free objects in categories (e.g., the free groups in $\mathbb{G}$ and the free monoids in $\mathbb{M}$), is very simple. An object T of a category $\mathbb{C}$ is said to be <u>terminal</u> in $\mathbb{C}$ iff for any object A of $\mathbb{C}$ there exists exactly one morphism $A \longrightarrow T$

in $\mathbb{C}$. Dually, an object J of $\mathbb{C}$ is <u>initial</u> in $\mathbb{C}$ iff for any object A of $\mathbb{C}$ there exists exactly one morphism $J \longrightarrow A$ in C (i.e., iff J is terminal in $\mathbb{C}^{op}$).

We show now that kernels of group homomorphisms are characterizable as terminal objects in especially constructed categories. For any group homomorphism $f : G_1 \longrightarrow G_2$, we define the category $\mathbb{K}(f)$ whose objects are the group homomorphisms $h : G \longrightarrow G_1$ for which $hf : G \longrightarrow G_2$ is the trivial homomorphism. There is a morphism in $\mathbb{K}(f)$ from $h : G \longrightarrow G_1$ to $h' : G' \longrightarrow G_1$ for each group homomorphism $j : G \longrightarrow G'$ for which $h = h'j$. By our previous observation we know that $k_f : N \longrightarrow G_1$ is a terminal object in $\mathbb{K}(f)$.

On the other hand, the free groups (and in general the free objects in categories that have free objects) give rise to initial objects. Let T be a fixed set. We define the category $\mathbb{F}r(T)$ whose objects are the pairs (u_T, G) where G is any group and $u_T : T \longrightarrow S(G)$ is any function from T to the carrier of G. There is a morphism in $\mathbb{F}r(T)$ from (u_T, G) to (v_T, H) for each group homomorphism $f : G \longrightarrow H$ for which $S(f)u_T = v_T$ holds in $\mathbb{S}$. The universality property of the free group $Fr(T)$ generated by T (as expressed in the previous section) means precisely that $(g_T, Fr(T))$ is an initial object in $\mathbb{F}r(T)$.

In general, if T is a terminal object in $\mathbb{C}$, then the only morphism $T \longrightarrow T$ in $\mathbb{C}$ must be the identity morphism. Hence, if T_1 and T_2 are terminal objects in $\mathbb{C}$, the morphisms $T_1 \longrightarrow T_2$ and $T_2 \longrightarrow T_1$ must be inverse to each other, and so any two terminal objects in $\mathbb{C}$ are isomorphic in $\mathbb{C}$. Similarly, initial objects are isomorphic in $\mathbb{C}$.

The recognition that certain constructed objects in categories are in fact terminal or initial in some constructed categories, usually captures the significant and essential properties of the objects.

5. Products and sums in categories

Usually, in categories of mathematical systems whose objects are sets with some additional structure, the products of objects are defined "element-wise" by means of cartesian products of the carriers. The notions of products and sums of objects in general categories is reduced to terminal and initial objects in constructed categories. We present here this reduction due to MacLane [13; pp. 50-51].

Let $\{A_j\}$ be a family of objects of $\mathbb{C}$ indexed by a set J. Consider the category $\mathbb{C}\{A_i\}$ whose objects are families $\{f_j : B \longrightarrow A_j\}$ of morphisms of $\mathbb{C}$ indexed by J with an arbitrary but common domain B. $\mathbb{C}\{A_j\}$ has a morphism from $\{f_j : B \longrightarrow A_j\}$ to $\{f'_j : B' \longrightarrow A_j\}$ for each morphism $h : B \longrightarrow B'$ of $\mathbb{C}$ with $f'_j h = f_j$, for all $j \in J$. A terminal object in $\mathbb{C}\{A_j\}$, if it exists, is called a <u>product diagram</u> of $\{A_j\}$. Thus a product diagram of $\{A_j\}$ consists of an object P of $\mathbb{C}$, a <u>product object</u> of $\{A_j\}$, together with a family $\{p_j : P \longrightarrow A_j\}$ of morphisms of $\mathbb{C}$, indexed by J, such that any family $\{f_j : B \longrightarrow A_j\}$ is factored uniquely and "uniformly" through $\{p_j : P \longrightarrow A_j\}$, namely $f_j = p_j h$ for some unique morphism $h : B \longrightarrow P$ of $\mathbb{C}$ but for all $j \in J$. The product diagram, being a terminal object in $\mathbb{C}\{A_j\}$, is unique up to an isomorphism in $\mathbb{C}\{A_j\}$. Since isomorphisms in $\mathbb{C}\{A_j\}$ are determined by isomorphisms in $\mathbb{C}$, as it can be easily verified, it follows that the product object P of the product diagrams of $\{A_j\}$ is unique up to an isomorphism in $\mathbb{C}$.

With an abuse of notation we identify isomorphic objects of $\mathbb{C}$ and we write $P = \Pi\{A_j\}$; while the product object of a two object family $\{A_1, A_2\}$ is written as $A_1 x A_2$.

For example, in the category $\mathbb{S}$ of sets, every family $\{T_j\}$ of sets indexed by a set J has a product diagram $\{p_j : \Pi\{T_j\} \longrightarrow T_j\}$ where $\Pi\{T_j\}$ is the cartesian product of $\{T_j\}$ and p_j is the projection of $\Pi\{T_j\}$ on T_j.

Note that the definition of the product diagram $\{p_j : P \longrightarrow A_j\}$ of $\{A_j\}$ as a terminal object in $\mathbb{C}\{A_j\}$ implies not only that any family $\{f_j : B \longrightarrow A_j\}$ factors through $\{p_j : P \longrightarrow A\}$ via a unique morphism $h : B \longrightarrow P$, but also that the correspondence $\{f_j\} \longrightarrow h$ determines a bijection of sets

$$\psi : \Pi\{\mathbb{C}(B,A_j)\} \longrightarrow \mathbb{C}(B,\Pi\{A_j\}) ,$$

where the left-hand "Π" denotes the product in $\mathbb{S}$; i.e., the cartesian product of sets. For the definition of $\{p_j : P \longrightarrow A_j\}$ with $P = \Pi\{A_j\}$ implies that ψ is injective. On the other hand, for any $h \in \mathbb{C}(B,\Pi\{A_j\})$, we obviously have $\psi(\{p_j h\}) = h$.

The dual notion of product is that of a <u>sum</u> (or <u>coproduct</u>) of the family $\{A_j\}$. The <u>sum diagram</u> of $\{A_j\}$ is an initial object (if it exists) in a category of families of the form $\{g_j : A_j \longrightarrow C\}$. Here again, the sum diagram $\{q_j : A_j \longrightarrow Q\}$ is unique up to an isomorphism in the associated category, and it follows that the <u>sum object</u> Q is unique up to an isomorphism in $\mathbb{C}$. We write $Q = \Sigma\{A_j\}$, and in particular $\Sigma\{A_1,A_2\} = A_1+A_2$.

In $\mathbb{S}$ every family of sets $\{T_j\}$ indexed by a set J has a sum diagram $\{q_j : T_j \longrightarrow \Sigma\{T_j\}\}$ where the sum object $\Sigma\{T_j\}$ is the disjoint union of $\{T_j\}$ and $q_i : T_i \longrightarrow \Sigma\{T_j\}$ is the obvious injection of the summand T_i into $\Sigma\{T_j\}$.

Again, the definition of $\{q_j : A_j \longrightarrow Q\}$ as a sum diagram of $\{A_j\}$ implies a bijection of sets

$$\Pi\{\mathbb{C}(A_j,C)\} \longrightarrow \mathbb{C}(\Sigma\{A_j\},C) .$$

Not every category yields the existence of product diagrams and sum diagrams for any family of objects. We say that $\mathbb{C}$ admits products (or sums, respectively) iff every set indexed family of objects of $\mathbb{C}$ has a product diagram (or sum diagram, respectively). Thus $\mathbb{S}$ admits both products and sums. The category $\mathbb{G}$ of groups and the category $\mathbb{M}$ of monoids admit products and sums. The product object of a family of groups or of monoids is their direct product and their sum object is their free product. $\mathbb{C}$at , the category of all categories also admits products and sums (even in a wider sense than the other categories). The evident cartesian product of categories yields a product diagram of any family of categories; their disjoint union yields a sum diagram.

In particular, if $\mathbb{C}_1$ and $\mathbb{C}_2$ are categories, then the objects of $\mathbb{C}_1 x \mathbb{C}_2$ are the ordered pairs (A_1,A_2) where A_i is an object of $\mathbb{C}_i$. The morphisms of $\mathbb{C}_1 x \mathbb{C}_2$ are of the form

$$(f_1,f_2) : (A_1,A_2) \longrightarrow (B_1,B_2) ,$$

where $f_i : A_i \longrightarrow B_i$ are morphisms of $\mathbb{C}_i$. The projections $P_i : \mathbb{C}_1 x \mathbb{C}_2 \longrightarrow \mathbb{C}_i$ are the functors defined by

$$P_i(A_1,A_2) = A_i \text{ and } P_i(f_1,f_2) = f_i .$$

The fact that $\{P_1,P_2\}$ is indeed a product diagram of $\{\mathbb{C}_1,\mathbb{C}_2\}$ is equivalent to the fact that for any pair $F_i : \mathbb{B} \longrightarrow \mathbb{C}_i$ of functors with a common domain $\mathbb{B}$, there exists a unique functor $G : \mathbb{B} \longrightarrow \mathbb{C}_1 x \mathbb{C}_2$ with $P_i G = F_i$; set

$$G(x) = (F_1(x),F_2(x)) ,$$

Suppose now that every two objects in $\mathbb{C}$ have a product. So let $\{p_i : A_1 x A_2 \longrightarrow A_i\}$ and $\{p_i' : B_1 x B_2 \longrightarrow B_i\}$ be product diagrams of $\{A_1, A_2\}$ and $\{B_1, B_2\}$ respectively. Then for any pair of morphisms $f_i : A_i \longrightarrow B_i$ of $\mathbb{C}$, there exists a unique morphism $h : A_1 x A_2 \longrightarrow B_1 x B_2$ defined by $p_i' h = f_i p_i$. By denoting $h = f_1 x f_2$, we find that the product of pairs of objects of $\mathbb{C}$ determines a functor

$$\Pi : \mathbb{C} x \mathbb{C} \longrightarrow \mathbb{C} ,$$

with $\Pi(A_1, A_2) = A_1 x A_2$ and $\Pi(f_1, f_2) = f_1 x f_2$.

A dual treatment for a category $\mathbb{C}$ in which every pair of objects $\{A_1, A_2\}$ has a sum $A_1 + A_2$ yields another functor

$$\Sigma : \mathbb{C} x \mathbb{C} \longrightarrow \mathbb{C} .$$

6. Natural transformations and operations on functors

We are concerned here with relationships that may exist between two functors that are defined on the same category and have values in a possibly different but common range category. Let $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ and $G : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ be two functors from $\mathbb{C}_1$ to $\mathbb{C}_2$; when can we say that F and G are essentially identical? First, in order to be able to use F and G interchangeably, we must have at least that for any object A of $\mathbb{C}_1$, $F(A)$ and $G(A)$ are isomorphic in $\mathbb{C}_2$. Furthermore, we would like to have some relationship which binds $F(f)$ and $G(f)$ for any morphism f of $\mathbb{C}_1$. This can be achieved by requiring that for any morphism f of $\mathbb{C}_1$, the morphism $F(f)$ and $G(f)$ of $\mathbb{C}_2$ are <u>similar</u> in $\mathbb{C}_2$ in the sense that $F(f) = pG(f)q$ for some <u>regular</u> morphisms p and q of $\mathbb{C}_2$, i.e., p and q are isomorphisms of $\mathbb{C}_2$. Note that this requirement implies our first requirement. Next we would like to require that the relationships $F(f) = pG(f)q$ will be "uniform" in the sense that p and q will depend only on the range and domain of f .

Let us examine a specific example. The universality property of the free objects in a category, say $\mathbb{M}$, the category of monoids, yields a relationship of this kind between two families of functors from $\mathbb{M}$ to $\mathbb{S}$. For a fixed set T , we have the following two functors:

$$\mathbb{M}(Fr(T),-) : \mathbb{M} \longrightarrow \mathbb{S} ,$$

with $\mathbb{M}(Fr(T),-)(M) = \mathbb{M}(Fr(T),M)$, the set of all monoid homomorphisms from Fr(T) to M , where M is any given monoid; and

$$\mathbb{S}(T,-)\circ S : \mathbb{M} \longrightarrow \mathbb{S}$$

with $[\mathbb{S}(T,-)\circ S](M) = [\mathbb{S}(T,-)](S(M)) = \mathbb{S}(T,S(M))$, the set of all functions from T to the carrier of M . We write $\mathbb{S}(T,S(-))$ for $\mathbb{S}(T,-)\circ S$.

The universality property of the free monoids implies directly the existence of a one-one correspondence (i.e., an isomorphism of $\mathbb{S}$)

$$\sigma(M) : \mathbb{M}(Fr(T),M) \longrightarrow \mathbb{S}(T,S(M))$$

given explicitly by

$$[\sigma(M)](f) = S(f)g_T ,$$

where $g_T : T \longrightarrow SFr(T)$ is the inclusion of T into the carrier of Fr(T) , the free monoid generated by T . We prove now in detail that for any monoid homomorphism $h : M_1 \longrightarrow M_2$ the diagram

$$\begin{array}{ccc}
\mathbb{M}(Fr(T),M_1) & \xrightarrow{\sigma(M_1)} & \mathbb{S}(T,S(M_1)) \\
{\scriptstyle [\mathbb{M}(Fr(T),-)](h)}\downarrow & & \downarrow{\scriptstyle [\mathbb{S}(T,S(-))](h)} \\
\mathbb{M}(Fr(T),M_2) & \xrightarrow{\sigma(M_2)} & \mathbb{S}(T,S(M_2))
\end{array}$$

is commutative and therefore the two functors $\mathbb{M}(Fr(T),-)$ and $\mathbb{S}(T,S(-))$ are related to each other in the sense described above. The technical term for this type of relationship is <u>natural equivalence</u>.

So let f be an arbitrary element of $\mathbb{M}(Fr(T),M_1)$; i.e., $f : Fr(T) \longrightarrow M_1$ is a morphism of $\mathbb{M}$ from $Fr(T)$ to M_1 . We chase the diagram along its left edge and get:

$$\begin{aligned}
[\sigma(M_2)][[\mathbb{M}(Fr(T),-)](h)](f) &= [\sigma(M_2)](hf) \\
&= S(hf)g_T \\
&= S(h)S(f)g_T \ ;
\end{aligned}$$

and in the other way we get:

$$\begin{aligned}
[[\mathbb{S}(T,S(-))](h)][\sigma(M_1)](f) &= [[\mathbb{S}(T,-)](S(h))](S(f)g_T) \\
&= S(h)S(f)g_T \ .
\end{aligned}$$

Hence the diagram is commutative and $\mathbb{M}(Fr(T),-)$ and $\mathbb{S}(T,S(-))$ are naturally equivalent.

The notion of natural transformation is a generalization of the notion of natural equivalence. We define a <u>transformation</u> τ from $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ to $G : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ to be a family of morphisms $\tau(A) : F(A) \longrightarrow G(A)$ of $\mathbb{C}_2$, one for each object A of $\mathbb{C}_1$. A transformation τ from F to G is an <u>equivalence</u> (of F and G) iff for any object A of $\mathbb{C}_1$, $\tau(A)$ is an isomorphism of $\mathbb{C}_2$. A transformation τ from F to G is said to be *natural* (in symbols $\tau : F \longrightarrow G$) iff any morphism $f : A \longrightarrow B$ of $\mathbb{C}_1$ yields a commutative square

$$\begin{array}{ccc} F(A) & \xrightarrow{\tau(A)} & G(A) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(B) & \xrightarrow[\tau(B)]{} & G(B) \end{array}$$

of morphisms in $\mathbb{C}_2$. A *natural equivalence* is obviously an equivalence which is natural.

We realized that in $\mathbb{C}at$, the category of all categories, the functors were regarded as morphisms of categories. Now it must be clear (and our notation reveals it) that natural transformations of functors can be regarded as *morphisms of* functors in suitable categories whose objects are functors.

Let $\mathbb{B}$ and $\mathbb{C}$ be any categories, the *functor category* $\mathbb{C}^{\mathbb{B}}$ has objects all functors $F : \mathbb{B} \longrightarrow \mathbb{C}$ and its morphisms are natural transformations of such functors $\tau : F \longrightarrow G$ with the evident composition of transformations. If $\mathbb{B}$ is a small category then $\mathbb{C}^{\mathbb{B}}$ satsifies the set-theoretic axiom on categories.

An interesting amalgamation of the category $\mathbb{C}at$ with the categories of the form $\mathbb{C}^{\mathbb{B}}$ is suggested in an unpublished work of J. Bénabou. We give here his ideas as presented by MacLane [13; pp.49-50].

LEMMA 6.1. Let $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ be categories, $F, G : \mathbb{A} \longrightarrow \mathbb{B}$ and F' , $G' : \mathbb{B} \longrightarrow \mathbb{C}$ functors, and $\tau : F \longrightarrow G$ and $\tau' : F' \longrightarrow G'$ be natural transformations of functors. Then for any object A of $\mathbb{A}$

we have

$$G'(\tau(A))\tau'(F(A)) = \tau'(G(A))F'(\tau(A)) .$$

PROOF. By applying the commutative square of the natural transformation $\tau' : F' \longrightarrow G'$ to the morphism $\tau(A) : F(A) \longrightarrow G(A)$ of $\mathbb{B}$, we get the desired equation.

LEMMA 6.2. With the same notation as in Lemma 6.1, the transformation $\tau' * \tau$ from $F' \circ F$ to $G' \circ G$ defined by

$$[\tau' * \tau](A) = \tau'(G(A))F'(\tau(A))$$

is a natural transformation $\tau' * \tau : F' \circ F \longrightarrow G' \circ G$.

PROOF. Let $f : A \longrightarrow B$ be any morphism of $\mathbb{A}$. We want to prove the equality

$$[\tau' * \tau](B)[F' \circ F](f) = [G' \circ G](f)[\tau' * \tau](A) .$$

Indeed:

$$\begin{aligned} [\tau' * \tau](B)[F' \circ F](f) &= \tau'(G(B))F'(\tau(B))F'(F(f)) \\ &= \tau'(G(B))F'(\tau(B)F(f)) \\ &= \tau'(G(B))F'(G(f)\tau(A)) \\ &= \tau'(G(B))F'(G(f))F'(\tau(A)) \\ &= G'(G(f))\tau'(G(A))F'(\tau(A)) \\ &= [G' \circ G](f)[\tau' * \tau](A) . \end{aligned}$$

LEMMA 6.3. If we add to the hypothesis of Lemma 6.1 that $F'' , G'' : \mathbb{C} \longrightarrow \mathbb{D}$ are functors with a natural transformation

$\tau'' : F'' \longrightarrow G''$ then we have the associativity

$$(\tau''*\tau')*\tau = \tau''*(\tau'*\tau) .$$

PROOF. Let A be any object of $\mathbb{A}$, then we have

$$\begin{aligned}[(\tau''*\tau')*\tau](A) &= [\tau''*\tau](G(A))[F''\circ F](\tau(A)) \\ &= \tau''([G'\circ G](A))F''(\tau'(G(A)))F''F(\tau(A)) \\ &= \tau''([G'\circ G](A))F''(\tau'(G(A))F(\tau(A))) \\ &= \tau''([G'\circ G](A))F''([\tau'*\tau](A)) \\ &= [\tau''*(\tau'*\tau)](A) .\end{aligned}$$

As a combined result of Lemmata 6.1, 6.2 and 6.3 we infer:

THEOREM 6.4. The collection $\mathbb{N}\mathrm{at}$ of all categories as objects and of morphisms of the form

$$(\tau,F,G) : \mathbb{A} \longrightarrow \mathbb{B} ,$$

where $F,G : \mathbb{A} \longrightarrow \mathbb{B}$ are functors and $\tau : F \longrightarrow G$ is a natural transformation, is a category with the composition rule

$$(\tau',F',G')*(\tau,F,G) = (\tau'*\tau,F'\circ F,G'\circ G)$$

PROOF. All we have to add to the previous lemmata is the proof of the existence of identity morphisms in $\mathbb{N}\mathrm{at}$. We leave it for the reader to verify that for any category $\mathbb{A}$, the morphism

$$(i_{I_{\mathbb{A}}},I_{\mathbb{A}},I_{\mathbb{A}}) : \mathbb{A} \longrightarrow \mathbb{A}$$

of $\mathbb{N}\mathrm{at}$, where $i_{I_{\mathbb{A}}} : I_{\mathbb{A}} \longrightarrow I_{\mathbb{A}}$ is the identity natural transformation of the identity functor $I_{\mathbb{A}}$ of $\mathbb{A}$, is in fact an identity morphism in $\mathbb{N}$at .

In order to realize that $\mathbb{N}$at combines in its structure both the structure of $\mathbb{C}$at and indirectly the structures of the categories of the form $\mathbb{C}^{\mathbb{B}}$, we observe the following properties of $\mathbb{N}$at . First, we define an assignment i from $\mathbb{C}$at into $\mathbb{N}$at . For any category $\mathbb{A}$ set $i(A) = A$. For any functor $F : A \longrightarrow B$ (i.e., a morphism of $\mathbb{C}$at) we define $i(F)$ to be the morphism

$$(i_F, F, F) : \mathbb{A} \longrightarrow \mathbb{B}$$

of $\mathbb{N}$at , where $i_F : F \longrightarrow F$ is the identity natural transformation of F .

PROPOSITION 6.5. $i : \mathbb{C}\text{at} \longrightarrow \mathbb{N}\text{at}$ is an injective functor.

PROOF. By its very definition, the assignment i is injective, and it maps identity morphisms of $\mathbb{C}$at on identity morphisms of $\mathbb{N}$at .

Let $F : \mathbb{A} \longrightarrow \mathbb{B}$ and $F' : \mathbb{B} \longrightarrow \mathbb{C}$ be morphisms of $\mathbb{C}$at and let A be any object of $\mathbb{A}$, then we have

$$i_{F'}(F(A)) = F'(i_F(A)) = i_{F' \circ F}(A) ,$$

and therefore

$$i_{F' \circ F}(A) = [i_{F'} * i_F](A) ,$$

which implies

$$i(F' \circ F) = i(F') * i(F) .$$

Thus, if we identify a functor F with its image i(F) in $\mathbb{N}$at we find that the composition rule $*$ of the morphisms of $\mathbb{N}$at yields the ordinary composition of functors when it is applied to the image of $\mathbb{C}$at in $\mathbb{N}$at . The following proposition describes the manner in which the morphisms of the form $i(F)$ are combined with any other morphisms of $\mathbb{N}$at.

PROPOSITION 6.6. With the notation of Lemma 6.1, for any object A of $\mathbb{A}$ we have

$$[i(F')*\tau](A) = F'(\tau(A)) \quad \text{and} \quad [\tau'*i(F)](A) = \tau'(F(A)) \ ;$$

hence we can write

$$i(F')*\tau = F'\circ\tau \quad \text{and} \quad \tau'*i(F) = \tau'\circ F \ .$$

PROOF. For the first equation we have

$$\begin{aligned}[i(F')*\tau](A) &= [i(F')](G(A))F'(\tau(A)) \\ &= i_{F'}(G(A))F'(\tau(A)) \\ &= i_{F'\circ G}(A)F'(\tau(A)) \\ &= F'(\tau(A)) \ .\end{aligned}$$

Similarly, for the second equation we have

$$[\tau'*i(F)](A) = \tau'(F(A))F'(i_F(A)) = \tau'(F(A)) \ .$$

As an immediate but important corollary to this proposition, we find that natural equivalence of functors is a congruence relation with respect to the ordinary composition of functors.

COROLLARY 6.7. With the notation of Lemma 6.1, if $\tau : F \longrightarrow G$ is a natural equivalence, then

$$F'\circ\tau : F'\circ F \longrightarrow F'\circ G$$

is also a natural equivalence. If $\tau' : F' \longrightarrow G'$ is a natural equivalence then

$$\tau' \circ F : F' \circ F \longrightarrow F' \circ G$$

is also a natural equivalence.

Finally, the categories of the form $\mathbb{C}^{\mathbb{B}}$ appear in $\mathbb{N}$at as follows. The partial monoid of the morphisms of $\mathbb{C}^{\mathbb{B}}$ is precisely the class $\mathbb{N}$at($\mathbb{B}$,$\mathbb{C}$) provided with the ordinary composition of natural transformations. Tedious, but straightforward, applications of the properties of $\mathbb{N}$at as expressed by the previous lemmata, yield that the composition rule $*$ of $\mathbb{N}$at determines a functor

$$* : \mathbb{N}\mathrm{at}(\mathbb{A},\mathbb{B}) \times \mathbb{N}\mathrm{at}(\mathbb{B},\mathbb{C}) \longrightarrow \mathbb{N}\mathrm{at}(\mathbb{A},\mathbb{C}) ,$$

for any categories $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$.

CHAPTER II: CATEGORICAL PREDICATES IN CATEGORIES

The notion of categorical predicates of morphisms in a given category is understood intuitively as those classes of morphisms that can be defined "categorically" or "in terms of their properties as morphisms in an abstract category" (Freyd [8; p.28]). However, this notion was never before explicated or subjected to mathematical analysis. The process by means of which a definition is given to a notion that has been in use before but without a specific definition is called explication. An explication of a given intuitive notion must yield a certain justification to the effect that the intuitive meaning of the given notion is well captured by the suggested definition. This justification is not a matter of proof and, therefore, it is open to informal criticism. The best well known example for an explication in modern mathematics is Turing's explication of the intuitive notion of effective computation. Church's Thesis is the declaration that Turing's explication is well justified (Kleene [11; pp.300-301, 376-381]).

In this part of the work we give an explication for the notion of categorical predicates in categories. First, we study some examples of "obviously categorical" predicates in some specific categories. This study leads us to some preliminary approximations of the notion of categorical predicates by means of an interpretation of certain expressions of some formal languages. An examination of these approximations yields the first suggestion for a definition in terms of certain diagrams; i.e., functors from an arbitrary fixed category $\mathbb{D}$ into the given category $\mathbb{C}$. We call the type of the classes of morphisms of $\mathbb{C}$ defined by this suggestion diagramatical. Next, following a generally accepted intuition in general algebra, we suggest that the categorical classes of morphisms of $\mathbb{C}$ be categorical iff they are closed under all the automorphisms of $\mathbb{C}$. We call such classes auto-invariant. Now, as a justification for both suggestions, we prove that a class of morphisms in $\mathbb{C}$ is diagramatical iff it is auto-invariant. Our thesis is that categorical predicates in $\mathbb{C}$ are precisely the auto-invariant, or equivalently, the diagramatical classes of morphisms in $\mathbb{C}$. Henceforth we use only the term "categorical".

We conclude this chapter with a classification of categories according to the dispersion of their categorical predicates, and with a discussion of some methods for the determination of the categorical predicates in certain types of categories. In particular, we define $\mathrm{Aut}(\mathbb{C})$ to be the class group (i.e., a class with a group structure) of the equivalence classes of the automorphisms of $\mathbb{C}$ under nutural equivalence. Obviously, $\mathrm{Aut}(\mathbb{C})$ determines the categorical predicates in $\mathbb{C}$. We develop a method for computing $\mathrm{Aut}(\mathbb{C})$ which is applicable to many well known categories of mathematical systems, as we will show by some examples.

7. Examples of "obviously categorical" predicates

Consider the category $\mathbb{N}$ whose objects are the positive integers. The morphisms of $\mathbb{N}$ are the form

$$(m,n) : m \longrightarrow n$$

whenever $m \leq n$. The composition rule of $\mathbb{N}$ expresses the transitivity of the relation $\leq$; i.e., $(m_2,n_2)(m_1,n_1)$ is defined iff $m_2 = n_1$, and then it is equal to (m_1, n_2). As usual in categorical algebra, we identify (m,m), the identity morphism of m, with m itself.

It is obvious that every morphism of $\mathbb{N}$ can be characterized uniquely by means of its properties as an element in the partial monoid of the morphisms of $\mathbb{N}$. First 1 is characterized as the initial object of $\mathbb{N}$. Then, define a morphism (m,n) to be primitive iff it is not an identity morphism and it does not factor in a non-trivial manner as a product of two morphisms. Thus, (m,n) is primitive precisely when $(m,n) = fg$ in $\mathbb{N}$ for some morphisms f and g of $\mathbb{N}$, iff exclusively either f or g is an identity morphism. Clearly (m,n) is primitive iff n is the successor of m. Now 2 is characterized as the only object n of $\mathbb{N}$ such that $(1,n)$ is primitive. And, in general, $m+1$ is characterized as the only object n of $\mathbb{N}$ such that (m,n) is primitive. By induction, every object or identity morphism of $\mathbb{N}$ is uniquely characterized. Hence we can characterize uniquely every morphism (m,n) of $\mathbb{N}$ as the only morphism f of $\mathbb{N}$ such that nfm is defined in $\mathbb{N}$ (i.e., (m,n) is the only morphism of $\mathbb{N}$ from m to n). From this it follows that every set of morphisms of $\mathbb{N}$ can be characterized uniquely by means of a set of conditions, all of them expressed "categorically".

This study of $\mathbb{N}$ leads us to consider the following elementary formal explication of categorical prediates. We define a language ECL (i.e., the Elementary Categorical Language) as follows. ECL has a single type of individual variables denoted by $x, y, \ldots$; . A term of ECL is any non-empty finite string of individual variables. An atomic formula of ECL is of the form $t_1 = t_2$, where t_1 and t_2 are any terms of ECL. The compound formulae of ECL are derived from atomic formulae by means of the propositional connectives, and quantifiers over individual variables. We refer to the compound formulae of ECL as the elementary categorical formulae. The interpretation of the elementary categorical formulae in a given category $\mathbb{C}$ is evident. Namely, we interpret individual variables as (variables of the) morphisms of $\mathbb{C}$ and a string of variables is interpreted as the composition of the interpretation of the variables (i.e., of morphisms).

An elementary categorical definition is an elementary categorical formula $\pi(x)$ with a single free individual variable x. An elementary categorical formula $\pi(x)$ is said to define the class $\mathbb{C}(\pi(x))$ of morphisms of $\mathbb{C}$ in case $f \in \mathbb{C}(\pi(x))$ iff there eixsts an interpretation of $\pi(x)$ in $\mathbb{C}$ which turns out to be a true statement when x is interpreted as f. Thus our first approximation for the notion of categorical predicates in $\mathbb{C}$ is the family of all classes of morphisms of $\mathbb{C}$ which are defined by elementary categorical definitions. In other words, it should be intuitively clear that in any category $\mathbb{C}$, the classes of morphisms of $\mathbb{C}$ which are defined by means of elementary categorical definitions are "categorical". We will call such classes of morphisms of $\mathbb{C}$ elementary categorical classes in $\mathbb{C}$. The generalization to many place predicates of morphisms of $\mathbb{C}$ definable by means of elementary categorical formulae is straightforward.

It is obvious that in any category $\mathbb{C}$, the class of all elementary categorical classes is countable. This shows that the notion of elementary categorical classes is insufficient as an explication of the notion of "categorical" classes. For example, we have just agreed that the class of "categorical" classes in $\mathbb{N}$ is not countable. Employing higher order predicate calculi similar to ECL would not help us as long as we use the interpretations of single finite formulae as the definitions of classes of morphisms.

8. The diagramatical classes of morphisms

One may consider the possibility of employing infinite formulae (e.g. the infinite formulae derived from the atomic formulae of ECL by means of infinite applications of propositional connectives and quantifiers) for the definition of classes in given categories. In fact, the same intuition which leads us to believe that the elementary categorical classes are "categorical", may lead us to believe that classes defined by means of infinite formulae of the kind suggested above are also "categorical". For example, it is easy to verify that in this fashion we can define all the "categorical" classes in $\mathbb{N}$ (i.e., all the sets of morphisms of $\mathbb{N}$). Furthermore, it is possible that if we add also unrestricted higher order quantifiers over variables over classes of morphisms, we may get a sufficiently powerful explication for the desired notion of categorical predicates. The unfeasability of such an approach in this formal manner is obvious. Fortunately, categorical algebra is powerful enough to provide us with a rigorous definition of this procedure. In this section we will discuss a categorical algebra formulation of the "general type of definition of predicates in categories by means of categorical formulae."

We begin with two observations that connect the notion of formulae and their interpretations in a given category $\mathbb{C}$ with the notion of categories $\mathbb{D}$ and functors from $\mathbb{D}$ to $\mathbb{C}$. Let L be any calculus in which the only non-logical predicate is a binary function, and let $\psi : L \longrightarrow \mathbb{D}$ denote a certain interpretation of the formulae of L in a category $\mathbb{D}$. Then for any functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ the obvious composition $T \circ \psi$ is also an interpretation of L, but now in $\mathbb{C}$. Thus the functors $\mathbb{D} \longrightarrow \mathbb{C}$ transform the interpretations of L in $\mathbb{D}$ into interpretations of L in $\mathbb{C}$. On the other hand, any category $\mathbb{D}$ can be regarded as a formula in an appropriate calculus L and then functors from $\mathbb{D}$ into $\mathbb{C}$ are precisely the interpretations of $\mathbb{D}$ in $\mathbb{C}$. A canonical way of regarding $\mathbb{D}$ as a formula is to regard it as the, possibly infinite, conjunction of the atomic formulae (of ECL!) of the form

$$fg = h$$

for all f, g and h such that $fg = h$ is true in $\mathbb{D}$.

Thus we are going to consider the following general scheme of definition of predicates in $\mathbb{C}$ by means of functors. We restrict our discussion to single place predicates; i.e., to classes of morphisms in $\mathbb{C}$. Let $\mathcal{J}$ be any class of functors from categories to categories (i.e., a class of morphisms of **Cat**), let $\mathbb{D}$ be any category and let X be a class of morphisms of $\mathbb{D}$. The class K of morphisms of $\mathbb{C}$ is said to be <u>definable by X in $\mathbb{D}$ via $\mathcal{J}$</u> (in symbols $K = \langle \mathcal{J}, \mathbb{D}, X, \mathbb{C} \rangle$) iff $K = (\mathcal{J} \cap \mathbb{C}^{\mathbb{D}})(X)$; i.e., f belongs to K iff $f = T(x)$ for some functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ which belongs to $\mathcal{J}$, and for some x which belongs to X . Thus, under our previous interpretation, $\mathbb{D}$ is regarded as a formula, the elements of X , each one in its turn, as free variables in $\mathbb{D}$, and $\mathcal{J}$ is the class of admissible interpretations that may provide interpretations of $\mathbb{D}$ in $\mathbb{C}$. Our main concern now is to restrict the types of the classes $\mathcal{J}$ of the "admissible interpretations" in a manner which suits our intuition with respect to "categorical" predicates.

It should be clear that if we do not restrict the type of $\mathcal{J}$, then every class of morphisms in $\mathbb{C}$ is definable by means of such a scheme. The restriction of $\mathcal{J}$ to "categorical" classes in **Cat** is not only circular but also does not serve our purpose. For example, if $\mathcal{J}$ is the class of all identity functors, which is "obviously" categorical in **Cat** , then every class K of morphisms is definable by K itself in $\mathbb{C}$ via $\mathcal{J}$. On the other hand, if $\mathcal{J}$ is not restricted to functors which are surjective whenever their range happens to be $\mathbb{N}$, then the only definable classes in $\mathbb{N}$ via $\mathcal{J}$ are either empty or infinite.

The need for a certain kind of surjectiveness of the functors in has a deeper reason than the ad-hoc consideration of the "categorical" classes in $\mathbb{N}$. Our guiding intuition here is the following. Given a category $\mathbb{D}$ and a morphism x of $\mathbb{D}$, we want to consider a functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ so that $T(x)$ will be included in the class defined by x in $\mathbb{D}$. In order that this class can be regarded as "categorical" in $\mathbb{C}$, the functor $T : \mathbb{D} \longrightarrow \mathbb{C}$, in some sense, must specify the "neighborhood" of $T(x)$ in $\mathbb{C}$ according to the structure of $\mathbb{D}$; i.e., according to the "neighborhood" of x in $\mathbb{D}$. One way to make this requirement precise is to require that every true interpretation of any elementary categorical formula $\pi(x)$ in $\mathbb{D}$ will be transformed by $T : \mathbb{D} \longrightarrow \mathbb{C}$ into a true interpretation of $\pi(x)$ in $\mathbb{C}$, where $T(x)$ is now the interpretation of x as a variable of $\pi(x)$. Hence we have to consider as "admissible" functors only those functors which preserve true interpretations of at least the elementary categorical formulae.

For example, let $\pi(x)$ have the form $(\forall y)(\pi(x,y))$, where $\pi(x,y)$ is an elementary categorical formula with two free variables x and y. If we do not know anything about $\mathbb{C}$, and $\pi(x)$ has a true interpretation in $\mathbb{D}$, then we have to require that a functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ will be surjective in order to make it possible that T will yield a true interpretation of $\pi(x) = (\forall y)(\pi(x,y))$ in $\mathbb{C}$. For it is clear that "for all y <u>in the image of T in $\mathbb{C}$</u>, $\pi(x,y)$ " represents the transformation of the interpretation of $\pi(x)$ from $\mathbb{D}$ to $\mathbb{C}$ via T. Hence, if the image of T is all of $\mathbb{C}$, that is, if T is surjective, we get a true interpretation of $\pi(x)$ in $\mathbb{C}$.

The following consideration shows that we do not really need surjective functors. In order that "for all y in the image of T in $\mathbb{C}$, $\pi(x,y)$ " will be equivalent to "for all y in $\mathbb{C}$, $\pi(x,y)$ ", it is sufficient to require that all the morphisms of $\mathbb{C}$ that occur outside of the image of T in $\mathbb{C}$ will not be "relevant" to $\pi(x,y)$ as interpreted in $\mathbb{C}$ via $T : \mathbb{D} \longrightarrow \mathbb{C}$. One way to accomplish it is to require that for any morphism g of $\mathbb{C}$ which occurs outside of the image of T and any morphism f of $\mathbb{C}$ which occurs inside the image of T, both fg and gf will be undefined. Clearly, such a functor into $\mathbb{N}$ must be surjective but in general T does not have to be surjective, as we will explain later. In addition to this requirement on the image of T of $\mathbb{C}$, it is natural to require that $T : \mathbb{D} \longrightarrow \mathbb{C}$ be injective. For injective functors $T : \mathbb{D} \longrightarrow \mathbb{C}$ are those functors for which $\mathbb{D}$ can be regarded as a faithful description of the image of T in $\mathbb{C}$. We are ready now for precise definitions.

A class J of morphisms of a category $\mathbb{C}$ is an <u>ideal</u> in $\mathbb{C}$ iff for any morphism f of $\mathbb{C}$ and any morphism j in J, if fj is defined in $\mathbb{C}$ then fj belongs to J, and if jf is defined in $\mathbb{C}$ then jf belongs to J. A class J is said to be <u>bi-ideal</u> in $\mathbb{C}$ iff both J and the complement of J in $\mathbb{C}$ are ideals in $\mathbb{C}$.

Clearly, if J is a bi-ideal in $\mathbb{C}$, then for any j in J and any morphism f of $\mathbb{C}$ outside of J, both jf and fj are undefined in $\mathbb{C}$. Furthermore, if $T : \mathbb{D} \longrightarrow \mathbb{C}$ is an injective functor with the property that for any morphism g in the image of T in $\mathbb{C}$ and any morphism f of $\mathbb{C}$ outside of the image of T in $\mathbb{C}$, both fg and gf are undefined in $\mathbb{C}$, then the image of T in $\mathbb{C}$ is a bi-ideal. By the <u>image</u> of T in $\mathbb{C}$ we mean the class of all T(d) in $\mathbb{C}$ for all morphisms d in $\mathbb{D}$. In order to see this, note that for an injective functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ with the above mentioned property, both the image of T and its complement in $\mathbb{C}$ are closed under the morphism composition in $\mathbb{C}$. Now, if a partition of a category $\mathbb{C}$ into two classes which are closed under the morphism composition has the property that for any pair of morphisms f and g , one from each component of the partition, both fg and gf are undefined, then each component is a bi-ideal in $\mathbb{C}$.

We can give a graph-theoretic characterization of the bi-ideals in $\mathbb{C}$ as follows. First note that every ideal J in $\mathbb{C}$ assumes a category structure by restricting the composition rule of $\mathbb{C}$ to J . Now the binary relation "f and g are <u>connected</u> in $\mathbb{C}$ ", defined on the morphisms of $\mathbb{C}$ as the reflexive, symmetric and transitive closure of the relation "the composition fg is defined in $\mathbb{C}$ ", is obviously the minimal equivalence relation which includes the relation" the composition fg is defined in $\mathbb{C}$ ". Furthermore, every bi-ideal in $\mathbb{C}$ must be a union of equivalence classes under this connectedness, and every such union is a bi-ideal in $\mathbb{C}$. In particular, we have that every category $\mathbb{C}$ is a sum (i.e., disjoint union) $\Sigma\mathbb{C}_\alpha$ of categories $\mathbb{C}_\alpha$, each one of them is a minimal non-empty bi-ideal in $\mathbb{C}$. The $\mathbb{C}_\alpha$ are the <u>connected components</u> of $\mathbb{C}$; i.e., the equivalence classes of $\mathbb{C}$ under the connectedness relation.

We define now a functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ to be <u>admissible</u> iff T is an injective functor and the image of T in $\mathbb{C}$ is a bi-ideal.

As our first full explication of the notion of categorical (one-place) predicates we suggest the following notion.

DEFINITION 8.1. A class K of morphisms of $\mathbb{C}$ is <u>diagramatical</u> (in $\mathbb{C}$) iff there exists a category $\mathbb{D}$ and a class X of morphisms of $\mathbb{D}$ such that every morphism f in K is of the form $f = T(x)$ for some admissible functor $T : \mathbb{D} \longrightarrow \mathbb{C}$ and some x in X . We write $K = \langle \mathbb{D}, X, \mathbb{C} \rangle$.

It is easy to verify that every set K of morphisms of $\mathbb{N}$ is diagramatical : $K = \langle \mathbb{N}, K, \mathbb{N} \rangle$; for the only admissible functor $\mathbb{N} \longrightarrow \mathbb{N}$ is the identity functor of N . In general, since the identity functor of any category $\mathbb{C}$ is admissible, we have $K \subseteq \langle \mathbb{C}, K, \mathbb{C} \rangle$ for any class K of morphisms of $\mathbb{C}$. Since every automorphism of $\mathbb{C}$ is admissible, the class $\langle \mathbb{C}, K, \mathbb{C} \rangle$ is closed under all the automorphisms of $\mathbb{C}$. Hence if K is not closed under all the automorphisms of $\mathbb{C}$, $K \neq \langle \mathbb{C}, K, \mathbb{C} \rangle$.

9. Automorphism invariant predicates

Experience in general algebra leads one to consider another notion as a suitable explication for the notion of categorical predicates.

DEFINITION 9.1. A class K of morphisms of a category $\mathbb{C}$ is auto-invariant (in $\mathbb{C}$) iff for any morphism f that belongs to K , and any automorphism F of $\mathbb{C}$, $F(f)$ also belongs to K .

The motivation for choosing the notion of auto-invariant as an explication of the notion of categorical predicates is derived from the following considerations. In a group G , if $g \in G$ and α is an automorphism of G , then g and $\alpha(g)$ are indistinguishable simply from their "algebraic" properties in G . On the other hand, every one free variable formula in the (formal) theory of groups defines in G a set of elements which is closed under all the automorphisms of G . Note that any group is a category with a single identity morphism and only isomorphisms.

In general, it is obvious that all the elementary categorical classes in an arbitrary category $\mathbb{C}$ are auto-invariant in $\mathbb{C}$. Furthermore, since for any automorphism F of $\mathbb{C}$ and any admissible functor $T : \mathbb{D} \longrightarrow \mathbb{C}$, the functor $F \circ T : \mathbb{D} \longrightarrow \mathbb{C}$ is also admissible, it follows that all the diagramatical classes in $\mathbb{C}$ are auto-invariant. Our goal in this section is to prove the converse, that is, that every auto-invariant class is diagramatical. The equivalence of these two notions can be regarded as a justification for our intuitions in the persuit of an appropriate explication of the notion of categorical predicates. Furthermore, once this equivalence is established and we define categorical classes as either diagramatical classes or, equivalently, auto-invariant classes, the characterization of the categorical classes as auto-invariant will serve as an important tool for the determination of the categorical classes, and even the categorical predicates in any given category.

If the category $\mathbb{C}$ under discussion is connected, that is, if $\mathbb{C}$ does not contain any non-trivial bi-ideals, then every admissible functor into $\mathbb{C}$ must be surjective. Hence, in this case, the only admissible functors $\mathbb{C} \longrightarrow \mathbb{C}$ are the automorphisms of $\mathbb{C}$. Therefore, if K is an auto-invariant class in $\mathbb{C}$, then $K = \langle \mathbb{C}, K, \mathbb{C} \rangle$, and thus every auto-invariant class in $\mathbb{C}$ is diagramatical.

The general case is, however, that a category $\mathbb{C}$ is a non-trivial sum $\Sigma\mathbb{C}_\alpha$ of minimal bi-ideals $\mathbb{C}_\alpha$, where α belongs to an indexing class A. Given any auto-invariant class K in $\mathbb{C}$, then $K = \Sigma K_\alpha$, where $K_\alpha = K \cap \mathbb{C}_\alpha$, $\alpha \in A$, is a decomposition of K into a disjoint union of classes K_α, each one of the being a class of morphisms in $\mathbb{C}_\alpha$ which is a connected category. Now, every automorphism of $\mathbb{C}_\alpha$ can be extended into an automorphism of $\mathbb{C}$ say by making it the identity on all $\mathbb{C}_\beta$ with $\beta \neq \alpha$. Hence each K_α is an auto-invariant class in $\mathbb{C}_\alpha$, and therefore, by our previous discussion of auto-invariant classes in connected categories, K_α is diagramatical in $\mathbb{C}_\alpha$. Thus, K is a (class) disjoint union of classes K_α, where for each $\alpha \in A$, K_α is diagramatical in the connected component $\mathbb{C}_\alpha$ of $\mathbb{C}$. Say $K_\alpha = \langle \mathbb{D}_\alpha, X_\alpha, \mathbb{C}_\alpha \rangle$, with the requirement that if K_α is empty, then both $\mathbb{D}_\alpha$ and X_α are empty.

LEMMA 9.2. With the notation introduced above, $K = \langle \Sigma \mathbb{D}_\alpha, \Sigma X_\alpha, \mathbb{C} \rangle$.

PROOF. Let $T : \Sigma \mathbb{D}_\alpha \longrightarrow \mathbb{C}$ be an admissible functor. For any $\beta \in A$, denote by T_β its restriction on $\mathbb{D}_\beta$; i.e., $T_\beta = T \circ J_\beta$, where $J_\beta : \mathbb{D}_\beta \longrightarrow \Sigma \mathbb{D}_\alpha$ is the canonical inclusion functor. We want to show that for all $x \in \Sigma X_\alpha$, $T(x) \in K$. If the range of T_β is $\mathbb{C}_\beta$, then for all $x \in X_\beta$ we have $T_\beta(x) \in K_\beta \subseteq K$. If the range of T_β is $\mathbb{C}_\gamma$, where $\gamma \neq \beta$, then $\mathbb{C}_\beta$ and $\mathbb{C}_\gamma$ are isomorphic categories say by $J_{\beta\gamma} : \mathbb{C}_\gamma \longrightarrow \mathbb{C}_\beta$, and $J_{\beta\gamma} \circ T_\beta : \mathbb{D}_\beta \longrightarrow \mathbb{C}_\beta$ is an admissible functor. The isomorphism functor $J_{\beta\gamma}$ can be extended into an automorphism F of $\mathbb{C}$ which interchange $\mathbb{C}_\gamma$ with $\mathbb{C}_\beta$ under $J_{\beta\gamma}$ and its inverse, and leaves the rest of $\mathbb{C}$ fixed. Now for every $x \in X_\beta$, $(J_{\beta\gamma} \circ T_\beta)(x) \in K_\beta \subseteq K$, and therefore $T_\beta(x) = F(J_{\beta\gamma}(T_\beta(x))) \in K_\beta \subseteq K$. Since for any $x \in \Sigma X_\alpha$ we must have $x \in X_\beta$ for some $\beta \in A$ and $T(x) = T_\beta(x)$, it follows that for any $x \in \Sigma X_\alpha$, $T(x) \in K$. That is $\langle \Sigma \mathbb{D}_\alpha, \Sigma X_\alpha, \mathbb{C} \rangle$ is included in K. But obviously K must be included in $\langle \Sigma \mathbb{D}_\alpha, \Sigma X_\alpha, \mathbb{C} \rangle$. For $f \in K$ implies $f \in K_\beta$ for some $\beta \in A$. Hence $f = T_\beta(x)$ for some admissible functor $T_\beta : \mathbb{D}_\beta \longrightarrow \mathbb{C}_\beta$ and some $x \in X_\beta$. Now pick any admissible functor $T_\gamma : \mathbb{D}_\gamma \longrightarrow \mathbb{C}_\gamma$ for all $\gamma \neq \beta$ such that $\mathbb{D}_\gamma$ is not the empty category, then for the obvious admissible functor $\Sigma T_\alpha : \Sigma \mathbb{D}_\alpha \longrightarrow \mathbb{C}$ we have $(\Sigma T_\alpha)(x) = T_\beta(x) = f$.

In conclusion we have proved

THEOREM 9.3. In any category $\mathbb{C}$, a class of morphisms is diagramatical iff it is auto-invariant.

Note, however, that most of the familiar categories of mathematical systems have either initial or terminal objects. From this follows that they are connected categories. Assume that a category $\mathbb{C}$ has an object

A such that for any object B of $\mathbb{C}$, the class $\mathbb{C}(A,B)$ is not empty. We prove that $\mathbb{C}$ is connected. So let J be any non-empty bi-ideal in $\mathbb{C}$, say $f : B \longrightarrow C$ belongs to J. Now any morphism $A \longrightarrow B$ must belong to J. For if not, it belongs to $\bar{J}$, the complement of J, which is also an ideal in $\mathbb{C}$. Hence $A \longrightarrow B \longrightarrow C$ belongs to $\bar{J}$. But $f : B \longrightarrow C$ belongs to J, and therefore $A \longrightarrow B \longrightarrow C$ belongs also to J, which is impossible. By similar arguments we derive that the identity morphism of A must belong to J, any morphism from A to any other object of $\mathbb{C}$ must belong to J and any morphism of $\mathbb{C}$ must belong to J. Hence every bi-ideal in C is trivial. Dually, if a category $\mathbb{C}$ has an object Z such that for any object B of $\mathbb{C}$, the class $\mathbb{C}(B,Z)$ is not empty, then $\mathbb{C}$ is connected.

The restriction of our discussion to one-place predicates is non-essential, but some caution should be observed in generalizing our notions to relations among morphisms. For example, the "categorical" classes in $\mathbb{C}x\mathbb{C}$ should not be taken as the "categorical" binary relations in $\mathbb{C}$. To illustrate this, we note that the relation " f factors through g " defined as " $f = gx$ for some morphism x " is "obviously" a categorical relation in $\mathbb{C}$. But in general, it is not a "categorical" class in the product category $\mathbb{C}x\mathbb{C}$. The intuitive objection is that "categorically" speaking, the morphisms of any category, and in our case the morphisms of $\mathbb{C}x\mathbb{C}$, are analyzed only as elements in the partial monoid. Thus in general we cannot assume the ability to decompose "categorically" any morphism of $\mathbb{C}x\mathbb{C}$ as a pair of morphisms of $\mathbb{C}$. Furthermore, if we accept, even as a necessary condition, that every "categorical" class must be auto-invariant, then we cannot accept " f factors through g " as categorical in $\mathbb{C}x\mathbb{C}$. Simply, the relation " $f = gx$ for some x " is in general not preserved under all the automorphisms of $\mathbb{C}x\mathbb{C}$. On the other hand, for any automorphism F of C, " $f = gx$ for some morphism x of $\mathbb{C}$ " is equivalent to " $F(f) = F(g)x$ for some morphism x of $\mathbb{C}$ ".

DEFINITION 9.4. A predicate $P(x,y,\ldots)$ is said to be <u>auto-invariant</u> in $\mathbb{C}$ iff for any automorphism F of $\mathbb{C}$ and any morphisms $f,g,\ldots$; of $\mathbb{C}$, $P(f,g,\ldots)$ and $P(F(f),F(g),\ldots)$ are equivalent.

Put differently, the auto-invariant α-place predicates in $\mathbb{C}$ are the "diagonally" auto-invariant classes in $\mathbb{C}^\alpha$. The analogous generalization of the notion of diagramatical classes is immediate. A class K of morphisms of $\mathbb{C}^\alpha$ is a <u>diagramatical</u> <u>α-place</u> <u>predicate</u> <u>in</u> <u>$\mathbb{C}$</u> iff there exists a category $\mathbb{D}$ and a class X of α-tuples of morphisms of $\mathbb{D}$ such that $(f,g,\ldots)$ belongs to K iff $f = T(x)$, $g = T(y)$, ... ; for some α-tuple $(x,y,\ldots)$ in X and for some admissible functor $T : \mathbb{D} \longrightarrow \mathbb{C}$. A similar proof to that of Theorem 9.3 yields

THEOREM 9.5. An α-place predicate $P(x,y,\ldots)$ on the morphisms of $\mathbb{C}$ is auto-invariant iff the class of all α-tuples of morphisms of $\mathbb{C}$ that satisfy $P(x,y,\ldots)$ is a diagramatical α-place predicate in $\mathbb{C}$.

Our thesis is then

THESIS 9.6. An α-place predicate in a category $\mathbb{C}$ is categorical iff it is auto-trivial in $\mathbb{C}$.

<u>10. Degrees of categoricity of categories</u>

In this section we want to classify certain categories according to the distributions of their categorical predicates.

DEFINITION 10.1. A class of morphisms of a category $\mathbb{C}$ is said to be <u>natural</u> iff it is closed under all those automorphisms of $\mathbb{C}$ which are naturally equivalent to $I_{\mathbb{C}}$, the identity functor of $\mathbb{C}$.

Every categorical class in $\mathbb{C}$ must be natural since it is closed under all the automorphisms of $\mathbb{C}$. Note also that every isomorphism type in $\mathbb{C}$ (i.e., a class of identity morphisms of all the objects of $\mathbb{C}$ which are isomorphic in $\mathbb{C}$ to some arbitrary fixed object in $\mathbb{C}$) and every typical class in $\mathbb{C}$ (i.e., a union of isomorphism types in $\mathbb{C}$) are natural in $\mathbb{C}$. The later examples of natural classes are important because typical classes in $\mathbb{C}$ represent abstract algebraic properties of objects in $\mathbb{C}$ and, in particular, the isomorphism type of any given object in $\mathbb{C}$ represents the abstract algebraic structure of the object as an object of $\mathbb{C}$. For example, in the category of groups $\mathbb{G}$, the typical classes are the abstract algebraic properties of groups, and the isomorphism type of any given group G is the algebraic structure of G as a group.

We distinguish therefore between three types of categories.

DEFINITION 10.2. A category $\mathbb{C}$ is said to be:

(i) objective iff every isomorphism type in $\mathbb{C}$ is categorical;

(ii) transparent iff every natural class in $\mathbb{C}$ is categorical;

(iii) auto-trivial iff all the automorphisms of $\mathbb{C}$ are naturally equivalent to $I_{\mathbb{C}}$.

Every autotrivial category is transparent, and every transparent category is objective. It is fairly easy to construct a category $\mathbb{C}$ with three objects A,B and C such that B and C are isomorphic in $\mathbb{C}$ and the only non-identity automorphism of $\mathbb{C}$ leaves A fixed and interchanges B and C, and yet it is not naturally equivalent to $I_{\mathbb{C}}$. That is, the isomorphism $B \longrightarrow C$ is not natural. Hence the natural classes in $\mathbb{C}$ are any classes of morphisms of $\mathbb{C}$, and therefore not every natural class is categorical in $\mathbb{C}$. In particular the class which consists of the identity morphism of A only is natural but not auto-invariant. Hence $\mathbb{C}$ is an objective category which is not transparent. Note that this example also shows that natural classes of identity morphisms need not be typical. While all the familiar categories which are known to be transparent are, in fact, auto-trivial as well, I do not know of a proof to the effect that every transparent category is autotrivial, nor do I know of any example of a transparent category which is not auto-trivial.

If we consider groups as categories, then the automorphism functors of groups are precisely the ordinary automorphisms of groups, and the automorphism functors of a group G which are naturally equivalent to the identity functor of G are precisely the inner automorphisms of G. Hence a group is transparent iff all its automorphisms preserve conjugation classes, and a group is auto-trivial iff all its automorphisms are inner automorphisms. This suggests the following open problem in group theory. Is there a group with only class-preserving automorphisms which have any automorphism which is not inner?

In order to understand the nature of this problem, we have to understand the difference between objective, transparent and auto-trivial categories. The objective categories are those categories in which objects (i.e., identity morphisms) can be characterized up to an isomorphism by means of a one-place categorical predicate. Both transparent and auto-trivial categories are categories in which any morphism can be characterized up to a "natural equivalence" by means of a one-place categorical predicate. From this point of view, there is no difference between transparent and auto-trivial categories. In a transparent category $\mathbb{C}$, if f is any morphism and F is any automorphism of $\mathbb{C}$, then there exists an automorphism F^f of $\mathbb{C}$, naturally equivalent to $I_{\mathbb{C}}$, such that $F(f) = F^f(f)$. This is so because the minimal natural class in $\mathbb{C}$ which contains f must be auto-invariant. Note the logical structure of the previous statement "for all F, for all f, there exists F^f..., such that $F^f(f) = F(f)$". Thus it may be the case that for $f \neq g$ we would have $F^f \neq F^g$. If $\mathbb{C}$ is an auto-trivial category, then trivially the existential quantifier "there exists an automorphism F^f naturally equivalent to $I_{\mathbb{C}}$" can be put before the universal quantifier "for all f", since we take $F^f = F$.

11. The class groups associated with categories

In general a category need not be either auto-trivial or transparent or objective. In order to determine the deviation of a given category $\mathbb{C}$ from being of either one of these types of categories, three class groups associated with $\mathbb{C}$ are defined. A class group is a class which has a group structure. For example, for any category $\mathbb{C}$, the class of all automorphisms of $\mathbb{C}$ has a group structure with respect to the composition of functors, even if the carrier of this group may be not a set. The idea of employing class groups in this context is borrowed, with a certain modification, from Freyd's notion of the automorphism class group [8; p.28]. Namely, we define certain equivalence classes in the total class group of the automorphisms of $\mathbb{C}$ and show that they are congruences. Hence, these relations induce quotient class groups. The choice of the equivalence relations is almost determined by our goal.

We start with the deviation of a category from being objective. A category $\mathbb{C}$ is objective iff for any two functors F_1 and F_2 and object A of $\mathbb{C}$, there exists an isomorphism $F_1(A) \longrightarrow F_2(A)$, for the identity functor is one of the automorphisms of $\mathbb{C}$. According to our terminology of transformations of functors (cf. Ch. I, Section 6), a category $\mathbb{C}$ is objective iff for any pair of automorphisms F_1 and F_2, there exists an equivalence from F_1 to F_2. It is natural therefore to inquire about the effect of this equivalence relation (i.e., F_1 and F_2 are equivalent iff there exists an equivalence from F_1 to F_2) on the total class group of the automorphisms of any category $\mathbb{C}$. If F_1 and F_2 are equivalent, then for any functor $F : \mathbb{C} \longrightarrow \mathbb{C}$, $F_1 \circ F$ and $F_2 \circ F$ are equivalent. If in addition F preserves isomorphisms in $\mathbb{C}$, and an automorphism has this property, then it follows that $F \circ F_1$ and $F \circ F_2$ are also equivalent. Hence equivalence among the automorphisms of $\mathbb{C}$ is a congruence relation. We denote by $\mathrm{Ob}(\mathbb{C})$ the quotient class group of the total class group of the automorphisms of $\mathbb{C}$ under equivalence. We clearly have that $\mathrm{Ob}(\mathbb{C})$ is trivial iff $\mathbb{C}$ is objective. If, however $\mathbb{C}$ is not objective, then $\mathrm{Ob}(\mathbb{C})$ determines the categorical typical classes in $\mathbb{C}$ as follows. Let K be a typical class in $\mathbb{C}$ and let F_1 be any automorphism of $\mathbb{C}$. If F_1 preserves K and F_2 is equivalent to F_1 then F_2 also preserves K. Hence we have

THEOREM 11.1. For any category $\mathbb{C}$, let $\mathcal{F}_{ob}$ be a choice class of $\mathrm{Ob}(\mathbb{C})$, (i.e., a maximal class of automorphisms of $\mathbb{C}$ which are mutually not equivalent). Then a typical class K in $\mathbb{C}$ is categorical iff it is closed under all the automorphisms in $\mathcal{F}_{ob}$.

From the properties of $\mathbb{N}$at (cf. Ch. I, Section 6) it follows that natural equivalence of functors determines also a congruence relation in the total class group of all the automorphisms of $\mathbb{C}$. We denote by $\mathrm{Aut}(\mathbb{C})$ the quotient class group determined by natural equivalence of automorphisms. $\mathrm{Aut}(\mathbb{C})$ is trivial iff $\mathbb{C}$ is auto-trivial. Furthermore, since natural equivalence of functors is a particular case of equivalence of functors, the congruence determined by natural equivalence is a refinement of the congruence determined by equivalence. Hence we have a canonical (class) epimorphism

$$N : \mathrm{Aut}(\mathbb{C}) \longrightarrow \mathrm{Ob}(\mathbb{C}) .$$

The kernel of this epimorphism is the quotient class of all natural equivalence classes of automorphisms of $\mathbb{C}$ which are equivalent to $I_{\mathbb{C}}$ under natural equivalence. Hence N is an isomorphism iff every automorphism of $\mathbb{C}$ which is equivalent to $I_{\mathbb{C}}$ is naturally equivalent to $I_{\mathbb{C}}$. Put differently, N is an isomorphism iff for any automorphism F of $\mathbb{C}$ which is not naturally equivalent to $I_{\mathbb{C}}$, there exists an object A of $\mathbb{C}$ which is not isomorphic to F(A) in $\mathbb{C}$.

Similar to Thorem 11.1, we have

THEOREM 11.2. For any category $\mathbb{C}$, if $\mathcal{F}_{aut}$ is a choice class of $\mathrm{Aut}(\mathbb{C})$, then a natural class of K of morphisms of $\mathbb{C}$ is categorical iff it is closed under all the automorphisms in $\mathcal{F}_{aut}$.

Note that the condition specified in Theorem 11.2 (unlike Theorem 11.1) is not the most economical. For example, if $\mathbb{C}$ is transparent and not auto-trivial, then $\mathrm{Aut}(\mathbb{C})$ is not trivial and yet all the natural classes in $\mathbb{C}$ are by assumption categorical. The third congruence relation is introduced in order to derive the most economical condition for natural classes to be categorical. We say that the automorphisms F_1 and F_2 of $\mathbb{C}$ are <u>naturally related</u> iff for any morphism f of $\mathbb{C}$ there exists an automorphism F^f of $\mathbb{C}$, naturally equivalent to $I_{\mathbb{C}}$, such that $F^f(F_1(f)) = F_2(f)$. Thus, F_1 and F_2 are naturally related iff they preserve the same natural classes in $\mathbb{C}$. If F_1 and F_2 are naturally equivalent, then they are naturally related; take $F^f = F_2 \circ F_1^{-1}$. On the other hand, if F_1 and F_2 are naturally related, then they are equivalent, since $F^f(A)$ is isomorphic to A whenever F^f is naturally equivalent to $I_{\mathbb{C}}$. Hence natural relatedness is an intermediate relation between natural equivalence and equivalence of functors. We prove now that natural relatedness is a congruence relation among the automorphisms of $\mathbb{C}$. Assuming that F_1 and F_2 are naturally related, then for any morphism f of C there exists an automorphism F^f , naturally equivalent to $I_{\mathbb{C}}$, with $F^f(F_1(f)) = F_2(f)$. Hence for any morphism f and any automorphism G of $\mathbb{C}$, we have $F^{F(f)}(F_1(G(f))) = F_2(G(f))$; which shows that $F_1 \circ G$ and $F_2 \circ G$ are also naturally related. We also have $[G \circ F^f \circ G^{-1}](G(F_1(f))) = G(F_2(f))$. But since F^f is naturally equivalent to $I_{\mathbb{C}}$, it follows from the fact that natural equivalence is a congruence, that $G \circ F^f \circ G^{-1}$ is naturally equivalent to $I_{\mathbb{C}}$. Hence $G \circ F_1$ and $G \circ F_2$ are also naturally related. We denote by Tran($\mathbb{C}$) the quotient class group of the total class group of all the automorphisms of $\mathbb{C}$ under natural relatedness. We know by now that the epimorphism $N : \text{Aut}(\mathbb{C}) \longrightarrow \text{Ob}(\mathbb{C})$ factors through Tran($\mathbb{C}$) . Furthermore, by the nature of natural relatedness, namely that two automorphisms are naturally related iff they preserve the same natural classes, it follows that $\mathbb{C}$ is transparent iff Tran($\mathbb{C}$) is trivial. By the same argument we derive the following theorem.

THEOREM 11.3. For any category $\mathbb{C}$, if $\mathcal{F}_{tr}$ is a choice class of Tran($\mathbb{C}$) , then a natural class K of morphisms of $\mathbb{C}$ is categorical iff it is closed under all the automorphisms in $\mathcal{F}_{tr}$.

Denote by

$$T : \text{Aut}(\mathbb{C}) \longrightarrow \text{Tran}(\mathbb{C})$$

the canonical epimorphism from Aut($\mathbb{C}$) onto Tran($\mathbb{C}$) . The kernel of T is precisely the quotient class of all automorphisms of $\mathbb{C}$ which are naturally related to $I_{\mathbb{C}}$ under natural equivalence. Hence T is an isomorphism iff for any automorphism F_2 of $\mathbb{C}$ which is not naturally equivalent to $I_{\mathbb{C}}$, there exists a morphism f of $\mathbb{C}$ such that for all the automorphisms F_1 of $\mathbb{C}$ which are naturally equivalent to $I_{\mathbb{C}}$, we have $F_1(f) \neq F_2(f)$.

It turns out that among the three class groups $Ob(\mathbb{C})$, $Tran(\mathbb{C})$ and $Aut(\mathbb{C})$, the largest one (i.e., $Aut(\mathbb{C})$) is the easiest to compute. This is fortunate since, after computing $Aut(\mathbb{C})$, one would hope to be able to determine the other groups by examining $Aut(\mathbb{C})$ and by providing the epimorphisms T and N. In the rest of the chapter we discuss a sufficient condition for a category to be auto-trivial, and then generalize it so that we get a procedure by means of which the class group $Aut(\mathbb{C})$ can be computed for a broad class of categories. The class of categories which is covered by the applications of these procedures contains among many other categories the following categories: the category $\mathbb{C}$at of all categories, the category of small categories, the category of finite categories, the category of partially ordered sets, the category of sets, the category of monoids, the category of abelian monoids, the category of groups, the category of abelian groups, and categories of modules.

12. Auto-trivial categories with reconstruction schemes

We start with an example adapted from Freyd [8; pp.30-31] concerning $Aut(\mathbb{G}^{ab})$, where $\mathbb{G}^{ab}$ is the category of abelian groups. Broadly speaking, we consider the set $\mathbb{G}^{ab}(Z,G)$ of the group homomorphisms from Z, the group of integers, to an arbitrary abelian group G. By means of point-wise addition we define a binary operation + on $\mathbb{G}^{ab}(Z,G)$ which turns $\mathbb{G}^{ab}(Z,G)$ into an abelian group R(G) which is isomorphic to G. Since, as it turns out to be, the construction R(G) is "categorical on

$\mathbb{G}^{ab}$ " , it follows that every isomorphism type in $\mathbb{G}^{ab}$ is categorical, and therefore $\mathbb{G}^{ab}$ is at least objective. A further examination of the nature of the construction $G \longrightarrow R(G)$ reveals that $\mathbb{G}^{ab}$ is auto-trivial.

To be more specific, consider the forgetful functor $S : \mathbb{G}^{ab} \longrightarrow \mathbb{S}$ with $S(G)$ being the carrier of G , and the functor $H_Z : \mathbb{G}^{ab} \longrightarrow \mathbb{S}$ with $H_Z(G) = \mathbb{G}^{ab}(Z,G)$. The family of functions

$$\rho(G) : S(G) \longrightarrow H_Z(G) : g \longrightarrow h_g ,$$

where $h_g : Z \longrightarrow G$ is the unique homomorphism with $h_g(+1) = g$, is an equivalence from S to H_Z . For clearly $h_{g_1} = h_{g_2}$ implies $g_1 = g_2$ (evaluate both h_{g_1} and h_{g_2} at $+1$), and for any $f : Z \longrightarrow G$ we have $h_{f(+1)} = f$.

For any abelian group H and any morphism $h_1, h_2 : Z \longrightarrow H$, we denote by

$$(h_1, h_2) : Z+Z \longrightarrow H$$

the unique morphism guaranteed by the sum diagram of $Z+Z$ and by $h_1, h_2 : Z \longrightarrow H$. Note that the sum in $\mathbb{G}^{ab}$ amounts to the direct product of abelian groups. (In an element-wise manner, the homomorphism $(h_1, h_2) : Z+Z \longrightarrow H$ is given by $(h_1,h_2)(m,n) = h_1(m)+h_2(n)$.) From the properties of the sum diagram in $\mathbb{G}^{ab}$, it follows that for any morphisms $f : H_1 \longrightarrow H_2$ and $h_1, h_2 : Z \longrightarrow H_1$, we have

$$f(h_1, h_2) = (fh_1, fh_2) .$$

(The element-wise verification of this equality is trivial.)

Let $\delta : Z \longrightarrow Z+Z$ be the unique morphism for which both $(0,i)\delta$ and $(i,0)\delta$ are the identity morphism i of Z , where $0 : Z \longrightarrow Z$ is the trivial morphism. (The categorical definition of $0 : Z \longrightarrow Z$ is the morphism $Z \longrightarrow Z$ which factors through a terminal object of $\mathbb{G}^{ab}$.) Note that the element-wise description of $\delta : Z \longrightarrow Z+Z$ is $\delta(m) = (m,m)$. Now we can define the point-wise addition in $\mathbb{G}^{ab}(Z,G)$ as follows. For any $h_{g_1}, h_{g_2} : Z \longrightarrow G$, the morphism $h_{g_1}+h_{g_2} : Z \longrightarrow G$ is defined by

$$h_{g_1}+h_{g_2} = (h_{g_1}, h_{g_2})\delta .$$

Hence $h_{g_1}+h_{g_2}$ is described, element-wisely, by

$$[h_{g_1}+h_{g_2}](m) = h_{g_1}(m)+h_{g_2}(m) ,$$

and therefore we have

$$h_{g_1}+h_{g_2} = h_{g_1+g_2} .$$

This proves simultaneously that with this additon, $\mathbb{G}^{ab}(Z,G)$ enjoys an abelian group structure, that we will denote by $R(G)$, <u>and</u> that the bijection $\rho(G) : S(G) \longrightarrow \mathbb{G}^{ab}(Z,G)$ determines an isomorphism of G with $R(G)$.

Thus, we have a procedure by means of which we can reconstruct an isomorphic copy of any abelian group. Furthermore, once Z is given, the rest of the procedure depends only on the abstract structure of $\mathbb{G}^{ab}$. The precise meaning of this last statement is that for any automorphism F of $\mathbb{G}^{ab}$ for which $F(Z)$ is isomorphic to Z, $R(F(G))$ is isomorphic to $R(G)$, and therefore to G, for any given object G of $\mathbb{G}^{ab}$. We prove now that for any automorphism F of $\mathbb{G}^{ab}$, $F(Z)$ is isomorphic to Z; that is, the isomorphism type of Z is categorical in $\mathbb{G}^{ab}$.

LEMMA 12.1. An object X of $\mathbb{G}^{ab}$ is isomorphic to Z in $\mathbb{G}^{ab}$ iff

(i) for any object A of $\mathbb{G}^{ab}$, if A is not terminal in $\mathbb{G}^{ab}$, then $\mathbb{G}^{ab}(X,A)$ has more than one element, and

(ii) if $e : X \longrightarrow X$ is an idempotent morphism (i.e., $e^2 = e$) then either e is the identity morphism of X or e factors through a terminal object of $\mathbb{G}^{ab}$ (which is the trivial group).

PROOF. Obviously any isomorphic copy of Z in $\mathbb{G}^{ab}$ satisfies these two conditions. On the other hand, assume that X is an abelian group with non-trivial homomorphisms into non-trivial groups and such that every idempotent endomorphism of X is a trivial idempotent. Now, since Z is not a trivial group, there exists a non-trivial homomorphism $h : X \longrightarrow Z$. But this is possible iff there exists an epimorphism $e : X \longrightarrow Z$. Now, if $e : X \longrightarrow Z$ is onto, then there exists a homomorphism $f : Z \longrightarrow X$ such that ef is the identity morphism of Z. For assume $e(x) = +1$ for an element x of X, then define $f : Z \longrightarrow X$ by $f(m) = mx$. Clearly, fe is an idempotent of X which is not a trivial homomorphism (i.e., $fe(y)$ is not 0 for all y), hence it must be the identity morphism of X. Thus $e : X \longrightarrow Z$ is an isomorphism.

We can conclude now and infer that $\mathbb{G}^{ab}$ is objective. An intuitive way to describe this is the following. For any abstract property P of abelian groups, if one wants to determine "categorically" whether a given abelian group G satisfies P, then one has to construct $R(G)$, and G satisfies P iff $R(G)$ satisfies P. The rigorous proof that $\mathbb{G}^{ab}$ is objective is outlined as follows. Let $[G]$ be the isomorphism type of G in $\mathbb{G}^{ab}$ and let F be an arbitrary automorphism of $\mathbb{G}^{ab}$. The nature of the construction $G \longrightarrow R(G)$ implies that $R(F(G))$ is isomorphic to $R(G)$; hence G is isomorphic to $F(G)$. Thus every isomorphism type in $\mathbb{G}^{ab}$ is categorical and from this follows that $\mathbb{G}^{ab}$ is objective.

We will reexamine the definition of $R(G)$ and show that it yields a proof that $\mathbb{G}^{ab}$ is auto-trivial. First, we extend the assignment

$G \longrightarrow R(G)$ into a functor $R : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$. We note that for any morphism $f : G_1 \longrightarrow G_2$ of $\mathbb{G}^{ab}$, the function

$$H_Z(f) : \mathbb{G}^{ab}(Z,G_1) \longrightarrow \mathbb{G}^{ab}(Z,G_2) : h \longrightarrow fh$$

determines a homomorphism $R(G_1) \longrightarrow R(G_2)$ of abelian groups:

$$\begin{aligned}[H_Z(f)](h_{g_1}+h_{g_2}) &= [H_Z(f)]((h_{g_1},h_{g_2})\delta) \\ &= f(h_{g_1},h_{g_2})\delta \\ &= (fh_{g_1},fh_{g_2})\delta \\ &= [H_Z(f)](h_{g_1})+[H_Z(f)](h_{g_2}) .\end{aligned}$$

Thus we define $R(f) : R(G_1) \longrightarrow R(G_2)$ to be the morphism of $\mathbb{G}^{ab}$ which is determined by $H_Z(f)$. In this manner we have a functor $R : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ which satisfies $S \circ R = H_Z = \mathbb{G}^{ab}(Z,-)$.

LEMMA 12.2. The functor $R : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ is naturally equivalent to the identity functor of $\mathbb{G}^{ab}$. In particular, $\rho^* : I_{\mathbb{G}^{ab}} \longrightarrow R$ is given by $\rho : S \longrightarrow H_Z$; that is, for any abelian group G , the isomorphism $\rho^*(G) : G \longrightarrow R(G)$ is the isomorphism determined by $\rho(G) : S(G) \longrightarrow \mathbb{G}^{ab}(Z,G)$.

PROOF. We have shown above that $\rho(G)$ determines an isomorphism $\rho^*(G) : G \longrightarrow R(G)$. We prove now that $\rho : S \longrightarrow H_Z$ is a natural equivalence, and this implies that $\rho^* : I_{\mathbb{G}^{ab}} \longrightarrow R$ is a natural equivalence. So let $f : G_1 \longrightarrow G_2$ be an arbitrary morphism of $\mathbb{G}^{ab}$, then we have to show that the diagram

$$\begin{array}{ccc} G_1 & \xrightarrow{\rho^*(G_1)} & R(G_1) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle R(f)} \\ G_2 & \xrightarrow{\rho^*(G_2)} & R(G_2) \end{array}$$

is commutative. That is, we have to show that the underlying diagram

$$\begin{array}{ccc} S(G_1) & \xrightarrow{\rho(G_1)} & \mathbb{G}^{ab}(Z,G_1) \\ f \big\downarrow & & \big\downarrow H_Z(f) \\ S(G_2) & \xrightarrow{\rho(G_2)} & \mathbb{G}^{ab}(Z,G_2) \end{array}$$

(where here f denotes the underlying function of $f : G_1 \longrightarrow G_2$) is commutative. So let $g \varepsilon S(G_1)$, and we have

$$[\rho(G_2)](f(g)) = h_{f(g)} , \quad \text{and}$$

$$[H_Z(f)][\rho(G_1)](g) = fh_g .$$

By evaluating both $h_{f(g)}$ and fh_g at $+1$ we get $h_{f(g)} = fh_g$. Hence $\rho^* : I_{\mathbb{G}^{ab}} \longrightarrow R$ is natural equivalence.

Now we make precise and detailed the meaning of the "categoricity" of the construction $R : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$. In this particular context, it is sufficient to show that for any automorphism F of $\mathbb{G}^{ab}$, R and $R \circ F$ are naturally equivalent. If this is the case, then it follows from Lemma 12.2 that every automorphism F of $\mathbb{G}^{ab}$ is naturally equivalent to the identity functor of $\mathbb{G}^{ab}$. Hence $\mathbb{G}^{ab}$ is auto-trivial. Here we give only some hints for the construction of the natural equivalence $R \longrightarrow R \circ F$. We define an intermediate functor $R_{F(Z)} : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ described informally as following the construction $G \longrightarrow R(G)$ but instead of Z we take $F(Z)$. The automorphism F itself determines a natural equivalence $\sigma_F : R \longrightarrow R_{F(Z)} \circ F$. The transformation σ_F is given by the obvious isomorphism $R(G) \longrightarrow R_{F(Z)}(F(G))$ which is determined by

$F : \mathbb{G}^{ab}(Z,G) \longrightarrow \mathbb{G}^{ab}(F(Z),F(G))$. The isomorphism $F(Z) \longrightarrow Z$, whose existence is guaranteed by Lemma 12.1 and its implications, determines a natural equivalence $R_{F(Z)} \longrightarrow R$. Hence we have a natural equivalence $R_{F(Z)} \circ F \longrightarrow R \circ F$. Now by combining the natural equivalences $R \longrightarrow R_{F(Z)} \circ F$ and $R_{F(Z)} \circ F \longrightarrow R \circ F$, we get the desired natural equivalence $R \longrightarrow R \circ F$.

These discussions lead us to the first condition for auto-trivial categories. We say that a functor R from a category $\mathbb{C}$ (into an arbitrary category) is <u>categorical on $\mathbb{C}$</u> iff for any automorphism F of C , R and $R \circ F$ are naturally equivalent.

PROPOSITION 12.3. A category $\mathbb{C}$ is auto-trivial iff there exists a functor $R : \mathbb{C} \longrightarrow \mathbb{C}$ which is categorical on C and naturally equivalent to $I_{\mathbb{C}}$.

PROOF. If $\mathbb{C}$ is auto-trivial, then $I_{\mathbb{C}}$ is categorical on C and, of course, naturally equivalent to itself. If $R : \mathbb{C} \longrightarrow \mathbb{C}$ is naturally equivalent to $I_{\mathbb{C}}$ and categorical on $\mathbb{C}$, then we have for any automorphism F of $\mathbb{C}$, the natural equivalences $I_{\mathbb{C}} \longrightarrow R$, $R \longrightarrow R \circ F$ and $R \circ F \longrightarrow F$. Hence $\mathbb{C}$ is auto-trivial.

The role that Z plays in the functor $R : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ is emphasized in the following discussion of R . First we realize that any group G determines a functor $R_G : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ with $S \circ R_G = \mathbb{G}^{ab}(G,-)$; i.e., $S(R_G(H)) = \mathbb{G}^{ab}(G,H)$ for any abelian group H . The functor R_G is specified as follows. For any abelian group H , the group $R_G(H)$ is the abelian group with $\mathbb{G}^{ab}(G,H)$ as its carrier and the addition in $R_G(H)$ is the point-wise addition. The "categorical" definition of the addition in $R_G(H)$ is given as follows. The diagonal morphism $\delta_G : G \longrightarrow G+G$ is defined in an analogous manner to $\delta : Z \longrightarrow Z+Z$. For any two morphisms $h_1,h_2 : G \longrightarrow H$ we define $h_1+h_2 : G \longrightarrow H$ by

$$h_1+h_2 = (h_1,h_2)\delta_G ,$$

where $(h_1,h_2) : G+G \longrightarrow H$ is the unique morphism determined by the sum diagram of $G+G$ and by the morphisms $h_1,h_2 : G \longrightarrow H$. Again, since for any morphisms $h_1,h_2 : G \longrightarrow H_1$ and $h : H_1 \longrightarrow H_2$ we have $h(h_1,h_2) = (hh_1,hh_2)$, the function

$$\mathbb{G}^{ab}(G,H_1) \longrightarrow \mathbb{G}^{ab}(G,H_2) : h_1 \longrightarrow hh_1$$

determined by $h : H_1 \longrightarrow H_2$, determines a morphism

$$R_G(h) : R_G(H_1) \longrightarrow R_G(H_2)$$

and $R_G : \mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ is a functor.

Now the assignment $G \longrightarrow R_G$ is an assignment from the objects of $\mathbb{G}^{ab}$ to the objects of $\mathbb{N}at(\mathbb{G}^{ab},\mathbb{G}^{ab})$, the category with functors $\mathbb{G}^{ab} \longrightarrow \mathbb{G}^{ab}$ as objects and natural transformations as morphisms. It is not difficult to see that this assignment can be extended into a contravariant functor

$$R_* : \mathbb{G}^{ab} \longrightarrow \mathbb{N}at(\mathbb{G}^{ab},\mathbb{G}^{ab}) ,$$

where for any morphism $f : G_1 \longrightarrow G_2$ of $\mathbb{G}^{ab}$, $R_*(f)$ is the natural transformation

$$R_*(f) : R_{G_2} \longrightarrow R_{G_1}$$

given by

$$[R_*(f)](H) : R_{G_2}(H) \longrightarrow R_{G_1}(H) : h_2 \longrightarrow h_2 f .$$

$R_*(f)$ is natural just because of the associativity of the morphism composition in $\mathbb{G}^{ab}$. Note that since the isomorphisms in $\mathbb{N}at(\mathbb{G}^{ab},\mathbb{G}^{ab})$

are natural equivalences, $R_*(j)$ is a natural equivalence for any isomorphism j of $\mathbb{G}^{ab}$.

It is intuitively clear that for each object G of $\mathbb{G}^{ab}$, the functor R_G is defined "categorically from G". This intuition expresses the following fact. For any automorphism F of $\mathbb{G}^{ab}$, the restriction of F on $\mathbb{G}^{ab}(G,H)$, i.e., the function

$$\mathbb{G}^{ab}(G,H) \longrightarrow \mathbb{G}^{ab}(F(G),F(H)) \ : \ h \longrightarrow F(h)$$

determines an isomorphism

$$\sigma_F(H) \ : \ R_G(H) \longrightarrow R_{F(G)}(F(H))$$

of $\mathbb{G}^{ab}$.

LEMMA 12.4. The equivalence σ_F is a natural equivalence

$$\sigma_F \ : \ R_G \longrightarrow R_{F(G)} \circ F \ ,$$

for any automorphism F of $\mathbb{G}^{ab}$.

PROOF. Let $f : H_1 \longrightarrow H_2$ and $h : G \longrightarrow H_1$ be any morphisms of $\mathbb{G}^{ab}$; then we have

$$\begin{aligned}
[\sigma_F(H_2)][R_G(f)](h) &= [\sigma_F(H_2)](fh) \\
&= F(fh) \\
&= F(f)F(h) \\
&= [R_{F(G)}(F(f))](F(h)) \\
&= [R_{F(G)}(F(f))][\sigma_F(H_1)](h) \ .
\end{aligned}$$

We are familiar now with the type of argument which yields the proof of the following lemma.

LEMMA 12.5. If for some abelian group G , R_G is naturally equivalent to $I_{\mathbb{G}^{ab}}$, then every automorphism F of $\mathbb{G}^{ab}$ for which $F(G)$ is isomorphic to G is naturally equivalent to $I_{\mathbb{G}^{ab}}$.

Since we know by now that R_Z is naturally equivalent to $I_{\mathbb{G}^{ab}}$ and that the isomorphism type of Z is categorical in $\mathbb{G}^{ab}$, we have the following expected result.

COROLLARY 12.5. The category of abelian groups is auto-trivial.

Note that due to the functor R_* , the categoricity of a single isomorphism type (i.e., that of Z) is sufficient not only for the categoricity of every isomorphism type, typical class or natural class in $\mathbb{G}^{ab}$, but also for the autotriviality of $\mathbb{G}^{ab}$. The general case is now obvious.

DEFINITION 12.6. A <u>Hom-reconstruction scheme</u> for a category $\mathbb{C}$ is a contravariant functor

$$R_* : \mathbb{C} \longrightarrow \mathrm{Nat}(\mathbb{C},\mathbb{C})$$

together with a functor $S : \mathbb{C} \longrightarrow \mathbb{S}$ such that

(i) for any object A of $\mathbb{C}$,

$$S \circ R_*(A) = \mathbb{C}(A,-) ,$$

and

(ii) for any object A of $\mathbb{C}$ and any automorphism F of $\mathbb{C}$, there exists a natural equivalence

$$\sigma_F : R_*(A) \longrightarrow R_*(F(A)) \circ F$$

such that $S \circ \sigma_F : \mathbb{C}(A,-) \longrightarrow \mathbb{C}(F(A),F(-))$ given by $[S \circ \sigma_F](f) = F(f)$ is a natural equivalence.

More generally, a <u>general reconstruction scheme</u> for $\mathbb{C}$ is any (contravariant or covariant) functor

$$R_* : \mathbb{C} \longrightarrow \mathbb{N}at(\mathbb{C},\mathbb{C})$$

such that for any object A of $\mathbb{C}$ and any automorphism F of $\mathbb{C}$, there exists a natural equivalence

$$\sigma_F : R_*(A) \longrightarrow R_*(F(A))\circ F ,$$

The role of Z in the reconstruction scheme for $\mathbb{G}^{ab}$ discussed above, motivates us to give the following definition.

DEFINITION 12.7. Let $\mathbb{C}$ be a category with a general reconstruction scheme R_*. An object M of $\mathbb{C}$ is called an <u>R_*-generator</u> iff $R_*(M)$ is naturally equivalent to $I_{\mathbb{C}}$.

By the usual chain of natural equivalences we derive the following theorem.

THEOREM 12.8. Let $\mathbb{C}$ be a category with a general reconstruction scheme R_*, and let M be an R_*-generator in $\mathbb{C}$. Then $\mathbb{C}$ is auto-trivial iff $\mathbb{C}$ is transparent; in particular, $\mathbb{C}$ is auto-trivial iff the isomorphism type of M in $\mathbb{C}$ is categorical.

Most of the familiar auto-trivial categories have an R_*-generator for some Hom-reconstruction sheme R_*. For example, the category $\mathbb{M}^{ab}$ of abelian monoids has a Hom-reconstruction scheme R_* in a complete analogy with the Hom-reconstruction scheme for $\mathbb{G}^{ab}$. The infinite monogenic monoid is the only R_*-generator in $\mathbb{M}^{ab}$ for this reconstruction scheme, and it is distinguished in $\mathbb{M}^{ab}$ by the same properties as Z is distinguished in $\mathbb{G}^{ab}$ (cf. Lemma 12.1). Hence $\mathbb{M}^{ab}$ is auto-trivial.

Freyd shows [8; pp.29, 32-33] that the category of sets $\mathbb{S}$, the category of topological spaces, the category of T_1 spaces and the category of Hausdorf spaces all have Hom-reconstruction shcemes with generators, and therefore, as he points out, they are all auto-trivial. The category $\mathbb{S}$ of sets deserves special attention because it leads us to consider another but immediate sufficient condition for auto-trivial categories.

We say that a functor R from a category $\mathbb{C}$ (to some arbitrary category) reflects natural equivalences on Aut($\mathbb{C}$) iff for any two automorphisms F_1 and F_2 of $\mathbb{C}$, if $R \circ F_1$ is naturally equivalent to $R \circ F_2$, then F_1 and F_2 are naturally equivalent. It is easily verified that a category $\mathbb{C}$ with a functor R from $\mathbb{C}$, which reflects natural equivalences on Aut($\mathbb{C}$), is auto-trivial iff R is categorical on $\mathbb{C}$. In the example of the category $\mathbb{S}$ of sets, the functor $\mathbb{S}(U,-) : \mathbb{S} \longrightarrow \mathbb{S}$, where U is any singleton set, is categorical on $\mathbb{S}$, naturally equivalent to $I_{\mathbb{S}}$ and it reflects natural equivalences on Aut($\mathbb{S}$). Hence $\mathbb{S}$ is auto-trivial.

13. Non-deterministic reconstruction schemes for the computation of Aut($\mathbb{C}$)

Freyd [8; pp.29-30, 31-32] shows that some categories (namely $\mathbb{C}$at, the category of all categories and some of its subcategories, and $\mathbb{G}$, the category of all groups) have reconstruction schemes similar to Hom-reconstrcution schemes as defined in the previous section. However, for these categories the reconstruction schemes do not yield objects in the given category in a unique manner. These examples lead us to consider "non-deterministic" schemes, and we will show how such schemes lead to the computation of the class groups Aut($\mathbb{C}$) for many familiar categories C.

We start with an illustration based on Freyd's discussion of the category of small categories ([8]; pp.29-30). We will however apply it to $\mathbb{C}$at , the category of all categories. Let $[\longrightarrow]$ be the category with two objects and precisely three morphisms. Then for any category $\mathbb{C}$, the morphisms are in one-one correspondence with the functors (i.e., morphisms of $\mathbb{C}$at) from $[\longrightarrow]$ to $\mathbb{C}$. Our problem is to find a suitable definition for a binary operation on the morphisms $\mathbb{C}$at $([\longrightarrow], \mathbb{C})$ which will make the operation correspond to the composition rule of $\mathbb{C}$. Now, the category $\mathbb{C}$at has an automorphism which is not even equivalent to I the identity functor of $\mathbb{C}$at . Letting $D : \mathbb{C}at \longrightarrow \mathbb{C}at$ be the "opposition" functor that maps every category $\mathbb{C}$ on its dual $\mathbb{C}^{op}$, then D is an automorphism of $\mathbb{C}$at which is not equivalent to I . We have to show that there exist categories $\mathbb{C}$ such that $\mathbb{C}$ and $\mathbb{C}^{op}$ are not isomorphic in $\mathbb{C}$at . For example, every category $\mathbb{C}$ with initial objects but without terminal objects is not isomorphic to $\mathbb{C}^{op}$. For under isomorphism of categories initial objects are mapped on initial objects, and $\mathbb{C}^{op}$ has terminal objects only. The category $\mathbb{N}$, discussed in Section 7, is obviously such a category. Hence we must infer that we cannot reconstruct the composition rule of each category $\mathbb{C}$ in a unique manner from the class $\mathbb{C}$at $([\longrightarrow], \mathbb{C})$. We hope to be able to reconstruct $\mathbb{C}$ up to an "opposition".

Freyd suggests the following procedure for a reconstruction of $\mathbb{C}$ up to an "opposition" from the class $\mathbb{C}$at $([\longrightarrow], \mathbb{C})$. Let L and R be the objects of $[\longrightarrow]$ and $L \longrightarrow R$ be the single non-identity morphism of $[\longrightarrow]$. The sum category $[\longrightarrow] + [\longrightarrow]$ in $\mathbb{C}$at is the disjoint union of $[\longrightarrow]$ with itself. That is, it has four objects L_1 , R_1 , L_2 and R_2 , four identity morphisms and two non-identity morphisms $L_1 \longrightarrow R_1$ and $L_2 \longrightarrow R_2$. Let $[\longrightarrow\longrightarrow]$ be the category with three objects L , M and R , their identity morphisms, a single morphism $L \longrightarrow M$, a single morphism $M \longrightarrow R$ and their unique composition $L \longrightarrow R$ in $[\longrightarrow\longrightarrow]$. There are several functors from $[\longrightarrow]+[\longrightarrow]$ to $[\longrightarrow\longrightarrow]$ but only two of them are epic in $\mathbb{C}$at (i.e., surjective). Namely, we have an epic functor

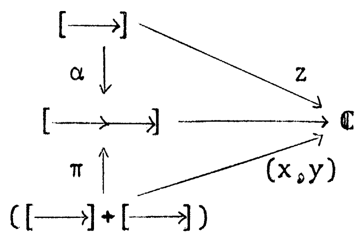
$$\pi_1 : ([\longrightarrow]+[\longrightarrow]) \longrightarrow [\longrightarrow\longrightarrow] ,$$

which maps $L_1 \longrightarrow R_1$ on $L \longrightarrow M$ and $L_2 \longrightarrow R_2$ on $M \longrightarrow R$, and an epic functor

$$\pi_2 : ([\longrightarrow]+[\longrightarrow]) \longrightarrow [\longrightarrow\longrightarrow] ,$$

which maps $L_1 \longrightarrow R_1$ on $M \longrightarrow R$ and $L_2 \longrightarrow R_2$ on $L \longrightarrow R$. Since the difference between π_1 and π_2 is a result of the opposition functor $D : \mathbb{C}at \longrightarrow \mathbb{C}at$, we cannot distinguish between π_1 and π_2 in $\mathbb{C}$at. Let $\alpha : [\longrightarrow] \longrightarrow [\longrightarrow\longrightarrow]$ be the functor which maps $L \longrightarrow R$ of $[\longrightarrow]$ on $L \longrightarrow R$ of $[\longrightarrow\longrightarrow]$.

Each one of the functors π_1, π_2, yields a reconstruction of $\mathbb{C}$ as follows. Let π be any epic functor from $[\longrightarrow]+[\longrightarrow]$ to $[\longrightarrow\longrightarrow]$, and let $x,y : [\longrightarrow] \longrightarrow \mathbb{C}$ any two functors. The functor $\pi(x,y) : [\longrightarrow] \longrightarrow \mathbb{C}$ is defined as the functor $z : [\longrightarrow] \longrightarrow \mathbb{C}$ for which there exists a functor $[\longrightarrow\longrightarrow] \longrightarrow \mathbb{C}$ such that



commutes. The functor $(x,y) : ([\longrightarrow]+[\longrightarrow]) \longrightarrow \mathbb{C}$ is the obvious functor determined by the sum diagram of $[\longrightarrow]+[\longrightarrow]$ in $\mathbb{C}$at and the functors $x,y : [\longrightarrow] \longrightarrow \mathbb{C}$. In this manner, the class $\mathbb{C}at\ ([\longrightarrow], \mathbb{C})$ with the composition rule

$$(x,y) \longrightarrow \pi(x,y)$$

becomes a category say, $R_\pi(\mathbb{C})$. It is clear that $R_{\pi_1}(\mathbb{C})$ is isomorphic to $\mathbb{C}$ and $R_{\pi_2}(\mathbb{C})$ is isomorphic to the dual of $\mathbb{C}$.

In order to verify that this construction (i.e., the predicate " $z = \pi_1(x,y)$ or $z = \pi_2(x,y)$ ") is categorical in $\mathbb{C}$at , we leave for the reader the verification of the following statements.

(i) The empty category is the initial object of $\mathbb{C}$at .

(ii) The category $\mathbb{1}$ with a unique morphism is isomorphic to any category $\mathbb{B}$ which is not empty and for which $\mathbb{C}at(\mathbb{B},\mathbb{B})$ contains a unique morphism.

(iii) The category $[\longrightarrow]$ is isomorphic to any category $\mathbb{B}$ such that $\mathbb{C}at(\mathbb{1},\mathbb{B})$ has two elements and $\mathbb{C}at(\mathbb{B},\mathbb{B})$ has three elements.

(iv) The category $[\longrightarrow\longrightarrow]$ is isomorphic to any category $\mathbb{B}$ such that $\mathbb{C}at(\mathbb{1},\mathbb{B})$ has three elements, $\mathbb{C}at([\longrightarrow],\mathbb{B})$ has six elements and thre exists an epic morphism $([\longrightarrow] + [\longrightarrow]) \longrightarrow \mathbb{B}$ in $\mathbb{C}$at .

(v) The functor $\alpha : [\longrightarrow] \longrightarrow [\longrightarrow\longrightarrow]$ is the only functor from $[\longrightarrow]$ to $[\longrightarrow\longrightarrow]$ which does not factor through any epic functor $([\longrightarrow]+[\longrightarrow]) \longrightarrow [\longrightarrow\longrightarrow]$.

From what we have said so far it follows that the typical classes in $\mathbb{C}$at are categorical iff they are closed under duality. That is, Ob($\mathbb{C}$at) is the cyclic group of order two. However, a further exmination of Freyd's construction implies that Aut($\mathbb{C}$at) is also the cyclic group of order two, and that any natural class in Cat is categorical iff it is closed under the "opposition" functor $D : \mathbb{C}at \longrightarrow \mathbb{C}at$. We proceed as follows. Based on the properties of $[\longrightarrow]$, we defined above two assignments R_{π_1} and R_{π_2} on the objects of $\mathbb{C}$at . It is evident, that these assignments can be extended into two functors

$$R_{\pi_1} : \mathbb{C}\mathrm{at} \longrightarrow \mathbb{C}\mathrm{at} \quad \text{and} \quad R_{\pi_2} : \mathbb{C}\mathrm{at} \longrightarrow \mathbb{C}\mathrm{at} .$$

This follows from the fact that the assignment

$$[\mathbb{C}\mathrm{at}([\longrightarrow], -)](T) : \mathbb{C}\mathrm{at}([\longrightarrow], \mathbb{C}_1) \longrightarrow \mathbb{C}\mathrm{at}([\longrightarrow], \mathbb{C}_2)$$

determines a functor

$$R_\pi(T) : R_\pi(\mathbb{C}_1) \longrightarrow R_\pi(\mathbb{C}_2)$$

for any functor (i.e., a morphism of $\mathbb{C}\mathrm{at}$) $T : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$. The proof of this assertion is straightforward but involved with a tedious diagram chasing as applied to the diagram which defines $\pi(x,y)$ above. Furthermore, for the automorphism D of $\mathbb{C}\mathrm{at}$ we have

$$R_{\pi_1} = D \circ R_{\pi_2} \quad \text{and} \quad R_{\pi_2} = D \circ R_{\pi_1} .$$

The procedure, by means of which we can prove that the previous observations on $\mathbb{C}\mathrm{at}$ imply that $\mathrm{Aut}(\mathbb{C}\mathrm{at})$ is the cyclic group of order two, and that every natural class in $\mathbb{C}\mathrm{at}$ is categorical iff it is closed under D , can be characterized by means of the following general definition.

DEFINITION 13.1. Let M be an object of a category $\mathbb{C}$. Denote by $\mathbb{C}[M]$ the class of all the objects $F(M)$ of $\mathbb{C}$ where F is an automorphism of $\mathbb{C}$ and $F(M)$ is isomorphic to M in $\mathbb{C}$. A (non-deterministic) reconstruction scheme $\langle M, R_\alpha : \alpha \in \mathcal{A} \rangle$ for the category $\mathbb{C}$ with the special object M is a family of functors

$$R_\alpha(M') : \mathbb{C} \longrightarrow \mathbb{C} ,$$

one for each $M' \in \mathbb{C}[M]$ and each α in a subgroup $\mathcal{A}$ of the total class group of all the automorphisms of $\mathbb{C}$ with the following properties.

(i) For any $\alpha \in \mathcal{A}$ and any $M_1, M_2 \in \mathbb{C}[M]$ the functors $R_\alpha(M_1)$ and $R_\alpha(M_2)$ are naturally equivalent.

(ii) For any $\alpha, \beta \in \mathcal{A}$ and any $M' \in \mathbb{C}[M]$, the functors $\alpha \circ R_\beta(M')$ and $R_{\beta\circ\alpha}(M')$ are naturally equivalent.

(iii) For the identity I of $\mathcal{A}$ (i.e., I is the identity functor of D), $R_I(M)$ is naturally equivalent to I.

(iv) For any automorphism F of C such that $F(M)$ is isomorphic to M in C (i.e., $F(M) \in \mathbb{C}[M]$) and for any $\alpha \in \mathcal{A}$, there exists an $\alpha_F \in \mathcal{A}$ such that $R_\alpha(M) \circ F$ is naturally equivalent to $R_{\alpha_F}(F(M))$.

As it follows from (ii) and (iii), and by the properties of $\mathbb{N}$at, for all $\alpha \in \mathcal{A}$ the functor $R_\alpha(M)$ is naturally equivalent to the automorphism α of C. We can prove a similar result for all the automorphisms F of C such that $F(M)$ is isomorphic to M in C (i.e., $F(M) \in \mathbb{C}[M]$).

LEMMA 13.2. Let $\langle M, R_\alpha : \alpha \in \mathcal{A} \rangle$ be a reconstruction scheme for a category $\mathbb{C}$ with the special object M of $\mathbb{C}$. Then for any automorphism F of $\mathbb{C}$ such $F(M)$ is isomorphic to M in $\mathbb{C}$, there exists an $\beta \in \mathcal{A}$ such that F is naturally equivalent to $R_\beta(M)$. In particular, by the notation of Definition 13.1, (iv), F is naturally equivalent to $R_{\alpha_F}(M)$ where $\alpha = I$.

PROOF. Let $\alpha_F \in \mathcal{A}$ be such that $R_I(M) \circ F$ is naturally equivalent to $R_{\alpha_F}(M)$. Since $F(M) \in \mathbb{C}[M]$, $R_{\alpha_F}(M)$ and $R_{\alpha_F}(F(M))$ are naturally equivalent. Since $R_I(M)$ is naturally equivalent to I, $R_I(M) \circ F$ is naturally equivalent to $F = I \circ F$. In conclusion, F is naturally equivalent to $R_{\alpha_F}(M)$.

Now, since α_F is naturally equivalent to $R_{\alpha_F}(M)$ we have the following result, which will guide most of our work in the rest of the paper.

THEOREM 13.3. Let $\langle M, R_\alpha : \alpha \in \mathcal{A} \rangle$ be a non-deterministic reconstruction scheme for a category $\mathbb{C}$ with the special object M, and let $\mathcal{A}^*$ be the quotient group of $\mathcal{A}$ under natural equivalence. If the isomorphism type of M in $\mathbb{C}$ is categorical, then $\mathrm{Aut}(\mathbb{C})$ is isomorphic to $\mathcal{A}^*$. In particular, a natural class of morphisms of $\mathbb{C}$ is categorical iff it is closed under $\mathcal{A}$.

Our examination of $\mathbb{C}$at shows that $\mathbb{C}$at has a non-deterministic reconstruction scheme indexed by the group of the two automorphisms I and D. The special object of this scheme is $[\longrightarrow]$, and since its isomorphism type in $\mathbb{C}$at is categorical, we get the desired results. Namely, a natural class of $\mathbb{C}$at is categorical iff it is closed under the "opposition" automorphism $D : \mathbb{C}at \longrightarrow \mathbb{C}at$, and $\mathrm{Aut}(\mathbb{C}at)$ is the cyclic group of order two.

Freyd [8; pp.29-32] provides us with sufficient information for the definition of non-deterministic reconstruction schemes for many well known categories. Although we differ slightly from Freyd's terminology and from his concepts, it is clear that the whole work presented in this chapter was motivated by the exercises of Chapter 1 in Freyd's book on abelian categories [8].

Freyd notes that since in the reconstruction scheme for $\mathbb{C}$at we make use of the categories $[\longrightarrow]$, $[\longrightarrow]+[\longrightarrow]$ and $[\longrightarrow\longrightarrow]$, one can reproduce the same scheme for any subcategory of $\mathbb{C}$at which contains these objects and all the functors among them. To name some important categories of this kind, we derive a result, similar to that about $\mathbb{C}$at, for each category in the following list: the category of small categories, the category of finite categories and the category of partially ordered sets (i.e., small categories $\mathbb{C}$ with $\mathbb{C}(A,B)$ containing at most a single morphism for any objects A, B, of $\mathbb{C}$).

In addition to these categories of categories, Freyd shows how to reconstruct the multiplication table of any group G from the properties of $\mathbb{G}(Z,G)$ in the category $\mathbb{G}$ of all groups. His construction determines in fact a non-deterministic reconstruction scheme for $\mathbb{G}$ with the special object Z and indexed by the cyclic group $\mathcal{A}$ of the two automorphisms of $\mathbb{G}$: the identity and $D : \mathbb{G} \longrightarrow \mathbb{G}$ which maps every group on its mirror image group. Since Z is distinguished in $\mathbb{G}$ by the same facts which distinguish it in $\mathbb{G}^{ab}$, the group Aut($\mathbb{G}$) is isomorphic to $\mathcal{A}^*$. But here $\mathcal{A}^*$ is trivial. The automorphism $D : \mathbb{G} \longrightarrow \mathbb{G}$ is naturally equivalent to $I_{\mathbb{G}}$ by the family of isomorphisms

$$\sigma(G) : D(G) \longrightarrow G : g \longrightarrow g^{-1} .$$

Hence $\mathbb{G}$ is auto-trivial. It is interesting to note that the same construction suggested by Freyd for $\mathbb{G}$ is applicable to $\mathbb{M}$, the category of all monoids. Here, however, the automorphism $D : \mathbb{M} \longrightarrow \mathbb{M}$ which carries each monoid into its mirror image is not even equivalent to the identity functor of $\mathbb{M}$. For example the three element monoid in which the product of any ordered pair of non-identity elements is equal to the first element in the pair, is obviously not isomorphic to its mirror image. Hence $\mathcal{A} = \mathcal{A}^*$ and both Aut($\mathbb{M}$) and Ob($\mathbb{M}$) (and therefore Tran($\mathbb{M}$)) are the cyclical group of order two. Furthermore, a natural class in $\mathbb{M}$ is categorical iff it is closed under $D : \mathbb{M} \longrightarrow \mathbb{M}$.

In particular, the class of surjective monoid homomorphisms, as a class of morphisms of $\mathbb{M}$, even though it is not identical with the class of the epic morphisms in $\mathbb{M}$, is nevertheless categorical in $\mathbb{M}$. Likewise, the class of (the identity morphisms of) the free monoids is categorical in $\mathbb{M}$, for obviously the mirror image of a free monoid is a free monoid. Note that a monoid F is free iff for any morphism $g : F \longrightarrow B$ and any <u>surjective</u> morphism $e : A \longrightarrow B$ of $\mathbb{M}$, there exists a morphism $f : F \longrightarrow A$ in $\mathbb{M}$ such that $ef = g$. Since epic morphisms in $\mathbb{M}$ need not be surjective, it is not surprising that the free monoids need not be projective in $\mathbb{M}$. For a definition of projective objects in categories see [8; Prop. 3.31] or in this paper, Part II, Ch. III, Section 10. In fact, the trivial monoid is the only projective object of $\mathbb{M}$.

PART II

THE CATEGORICAL PREDICATES IN CATEGORIES OF SEMIMODULES

INTRODUCTION TO PART II

Our main concern in this part is to determine the categorical predicates in certain categories of transition systems. Recent studies in automata theory indicate the importance of transition systems and their homomorphisms. Most types of automata which are studied have a free monoid as an input and a component which is defined as a transition system with a free input monoid. Frequently the transition system of an automaton is defined as a mapping from the cartesian product of a set (the set of states of the automaton) with a free monoid (the free monoid generated by the input alphabet of the automaton) back to the set of states of the automaton. This function is usually determined as a natural extension of an arbitrary function from the cartesian product of the set of states with the input alphabet back to the set of states (cf. [2], [9] and [16]). A homomorphism of automata of the same kind is usually determined by a homomorphism of their transition systems. Loosely speaking, a homomorphism from a transition system A to a transition system B (with the same input monoid) is a function from the states of A to the states of B which "preserves the functions of A and B". For a given fixed input monoid W, a category $\mathbf{S}^W$ of all the transition systems over W is determined by the homomorphisms of transition systems with the obvious composition rule.

Before we study the categorical predicates in $\mathbf{S}^W$, we define transition systems over an arbitrary input monoid W. Because of a certain analogy with modules, we will refer to these systems by the term "semimodules". We discover non-deterministic reconstruction schemes (cf. Part I, Ch. II, Section 13) for all the categories of semimodules which are defined analogously to $\mathbf{S}^W$ with a free monoid W. These schemes have a general form for all W, and in particular they are involved with groups of automorphisms of $\mathcal{S}^W$ which are induced by Aut(W) , the automorphism group of the input monoid W . Furthermore, for any W , the special object of the reconstruction scheme of $\mathcal{S}^W$ is W itself regarded as a semimodule. We denote this special object of $\mathcal{S}^W$ by M_W . Guided by Theorem 13.3 in Chapter II of Part I, we proceed and study several conditions under which the isomorphism type of M_W is categorical in $\mathcal{S}^W$. We find that for a very broad class of input monoids, the semimodule M_W is distinguished in $\mathcal{S}^W$ up to an isomorphism by means of a categorical predicate. This class of input monoids includes all the types of monoids encountered in automata theory (e.g., the free monoids, groups, abelian monoids and finite monoids). In order to establish this result, we study the category $\mathcal{S}^W$, for an arbitrary input monoid W , and yield the characterization of several categorically interesting classes of semimodules. Unfortunately, we do not have the desired result that for all input monoids W , the isomorphism type of M_W is categorical in $\mathcal{S}^W$.

As for the categorical predicates in $\mathcal{S}^W$ and the class group Aut($\mathcal{S}^W$) , we find the following results:

(i) for any monoid W , the class group Aut($\mathcal{S}^W$) has a subgroup isomorphic to Aut(W) ;

(ii) if the isomorphism type of M_W is categorical in $\mathcal{S}^W$, then Aut($\mathcal{S}^W$) is isomorphic to Aut(W) ;

(iii) for any monoid W , a categorical class in $\mathcal{S}^W$ must be closed under the automorphisms of $\mathcal{S}^W$ induced by Aut(W) ;

(iv) if the isomorphism type of M_W is categorical in $\mathcal{S}^W$, then a natural class in $\mathcal{S}^W$ is categorical iff it is closed under the automorphisms of $\mathcal{S}^W$ induced by Aut(W) .

The automorphisms of $\mathcal{S}^W$ which are induced by Aut(W) amount to the "relabeling" of W , in any semimodule over W , under an automorphism of W. In the case of semimodules over a free monoid W, the automorphisms induced by $Aut(W)$ amount to the permutations of the input alphabet labels. Thus, the categorical properties of transition systems with a free input monoid are precisely those properties which do not depend on the particular labeling of the alphabet letters. In Part III we will examine these properties in detail.

CHAPTER I: A PRELIMINARY STUDY OF CATEGORIES OF SEMIMODULES

1. The definition of $\mathbb{S}^W$

DEFINITION 1.1. A transition system A over a monoid W is a system of the form

$$A = (\lambda_A : S(A)xW \longrightarrow S(A))$$

where:

(i) $S(A)$ is an arbitrary set, the set of states of A ;

(ii) $\lambda_A : S(A)xW \longrightarrow S(A)$ is a function, the transition function of A, which satisfy the following requirements: for any $s \in S(A)$, if $w_1, w_2 \in W$ then

$$\lambda_A(\lambda_A(s,w_1),w_2) = \lambda_A(s,w_1w_2) ,$$

and if 1 is the identity element of W, then

$$\lambda_A(s,1) = s .$$

For example, Rabin and Scott, in their paper on finite automata [16], show that for any function $M : SxV \longrightarrow S$, where S and V are sets, there exists a function $M^* : SxV^* \longrightarrow S$, where V^* is the free monoid generated by V, such that $M^* : SxV^* \longrightarrow S$ is an extension of M and it is a transition system over V^*. The definition of M^* is the obvious one. Other examples of transition systems over an arbitrary monoid W are the following. We denote by $\emptyset_W$ the unique transition system over W with an empty set of states. The transition function of $\emptyset_W$ is the empty function $\emptyset xW \longrightarrow \emptyset$. For any singleton set U, we denote by U_W the transition system over W with U as its set of states. If $U = \{z\}$, then the transition function of U_W is defined by

$\lambda_{U_W}(z,w) = z$ for all $w \varepsilon W$. Finally, every monoid W can be regarded as a transition system over W which will be denoted by M_W. The transition function of M_W is defined by the multiplication in W. That is, the set of states of M_W is W and for any $w_1, w_2 \varepsilon W$ we define $\lambda_{M_W}(w_1, w_2) = w_1 w_2$. Note that M_W is a transition system over W precisely because W is a monoid.

We notice that transition systems over monoids can appear in mathematics in several contexts. Before we illustrate the ubiquity of transition systems, we introduce the notion of left transition systems as "mirror image" of transition systems. Thus, a left transition system over W has the form $B = (\lambda_B^* : W x S(B) \longrightarrow S(B))$ where λ_B^* satisfies now the equations

$$\lambda_B^*(w_1, \lambda_B^*(w_2, s)) = \lambda_B^*(w_1 w_2, s) ,$$

and

$$\lambda_B^*(1, s) = s .$$

It is obvious that the theory of left transition systems is the theory of transition systems as defined in Definition 1.1, but "through the looking glass".

Consider for example monadic algebras ([1], [5]). A monadic algebra A is a set $S(A)$ (the carrier of A) together with a set Ω (the operator domain of A) provided with a mapping τ from Ω to the set of all functions $S(A) \longrightarrow S(A)$. For any $s \varepsilon S(A)$ we will write ωs instead of $[\tau(\omega)](s)$, where ω is any arbitrary operator of A (i.e., $\omega \varepsilon \Omega$). Clearly, A can be regarded as a left transition system over Ω^*, the free monoid generated by Ω. Define $\lambda_A^*(\omega, s) = \omega s$. The relationship between monadic algebras and finite automata were discovered first by Büchi (cf. [2] and [17]).

Closely related to monadic algebras and its relationships to transition systems is the notion of groups operating on sets [3] and its generalization, the notion of monoids operating on sets. A homomorphism H of a group G into the group of all permutations on a set S(A) is called an <u>operation of G on S(A)</u>. If for $g \in G$ and $s \in S(A)$ we denote by gs the element [H(g)](s), then the operation H of G on S(A) can be regarded as both a monadic algebra with carrier S(A) and operator domain G (and therefore as a left transition system over G*, the free monoid generated by the set of elements of G) and a left transition system over G itself. A homomorphism H of a monoid W into the monoid of all the functions $S(A) \longrightarrow S(A)$ is called an <u>operation of W on S(A)</u>. Similar to operations of groups on sets, an operation of a monoid W on a set S(A) yields both a monadic algebra (with carrier S(A) and operator domain W) and a left transition system over W (with S(A) as its set of states).

Finally, modules over rings with identity ([3], [5]) yield also transition systems when one "forgets" their additive structure. A <u>(left) module</u> over a ring R with identity is a homomorphism M from R into the ring End(G) of the endomorphisms of an abelian group G. If in a given module $M : R \longrightarrow End(G)$ we ignore both the addition in R and the addition in G, we are left with a homomorphism of the multiplicative monoid of R into the monoid of the functions from the set of elements of G into itself. Hence by "forgetting" the additive structure of modules we get operations of monoids on sets, and therefore left-transition systems. Due to this fact, Cohn [5; pp.53, 55] calls an operation of a group on a set by the term <u>G-module</u>, and an operation of a monoid S on a set by the term <u>S-module</u>.

Since the structure of transition systems lacks the essential additive structure of modules, we are reluctant to use the term "W-module" for a transition system over W. However, because of the similarity that exists between transition systems and (right) modules, we suggest the compromising term W-semimodule (or simply semimodule) for referring to a transition system over W. We will reserve the term "transition system" for the contexts in which the relevance to automata theory is apparent. However, the arbitrary decision to discuss right transition systems instead of left transition systems, is motivated by the customary notation used in automata theory (cf. [9], [16] and [18]). On the other hand, we borrow from module theory the following notation. Instead of writing "$\lambda_A(s,w)$" we will use the abbreviation "$s \circ w$", especially whenever confusion with respect to the identification of the semimodule under consideration is not possible. For example, with this notation, transition functions (cf. Definition 1.1) satisfy precisely the multiplicative axioms of right modules:

$$s \circ 1 = s \quad \text{and} \quad s \circ (w_1 w_2) = (s \circ w_1) \circ w_2 .$$

The homomorphisms of transition systems over a fixed input monoid W also resemble the homomorphisms of modules over a fixed ring.

DEFINITION 1.2. Given two W-semimodules A and B, a function $f : S(A) \longrightarrow S(B)$ is said to determine the homomorphism $f : A \longrightarrow B$ of W-semimodules iff

$$f(s \circ w) = f(s) \circ w \quad \text{for all} \quad s \in S(A) \quad \text{and} \quad w \in W .$$

Thus, with any function $f : S(A) \longrightarrow S(B)$ which satisfies the above mentioned condition, we have a unique homomorphism $f : A \longrightarrow B$ of W-semimodules. Obviously a composition of functions which determine homomorphisms of W-semimodules, also determines a homomorphism of W-semimodules. Hence the class of all W-semimodules together with their homomorphisms forms a category whose composition rule is given by (and only by):

$$(f : A \longrightarrow B)(g : C \longrightarrow A) = (fg : C \longrightarrow B) .$$

We denote this category by $\mathcal{S}^W$.

2. Products and sums in $\mathcal{S}^W$

As we show in this section, the category $\mathcal{S}^W$ admits products and sums in an analogous manner to $\mathcal{S}$, the category of sets. Recall that in $\mathcal{S}$ the products are determined by cartesian products and the sums by disjoint union of sets (cf. Part I, Chapter I, Section 5).

Let $\{T_i\}$ be any family of sets indexed by a set I . The cartesian product $\Pi\{T_i\}$ is a product object of $\{T_i\}$ in $\mathcal{S}$ with the obvious projections $\{p_i : \Pi\{T_i\} \longrightarrow T_i\}$ as a product diagram. If $\{T_i'\}$ is another family of sets and $\{f_i : T_i \longrightarrow T_i'\}$ is a family of functions both indexed by I , then

$$\Pi\{f_i : T_i \longrightarrow T_i'\} : \Pi\{T_i\} \longrightarrow \Pi\{T_i'\}$$

is well defined as a product object of the family of the subsets $T_i x T_i'$ which represent the functions f_i as sets.

Similarly, the disjoint union $\Sigma\{T_i\}$ (e.g., $\Sigma\{T_i\} = \cup\{T_i x\{i\}\}$) is a sum object of $\{T_i\}$ in $\mathcal{S}$ with the obvious sum diagram $\{g_i : T_i \longrightarrow \Sigma\{T_i\}\}$ of injections. As for the family $\{f_i : T_i \longrightarrow T_i'\}$, it yields now a sum function

$$\Sigma\{f_i : T_i \longrightarrow T_i'\} : \Sigma\{T_i\} \longrightarrow \Sigma\{T_i'\} .$$

We leave it for the reader to verify the following lemma which describes the products and sums in $\mathbf{S}^W$.

LEMMA 2.1. Let $\{A_i\}$ be a family of W-semimodules indexed by some index set I . The object $\Pi\{A_i\}$ which is defined by

$$S(\Pi\{A_i\}) = \Pi\{S(A_i)\} \text{ and } \lambda_{\Pi\{A_i\}} = \Pi\{\lambda_{A_i}\}$$

is a product object of $\{A_i\}$ in $\mathbf{S}^W$. Furthermore, the family $\{p_i : \Pi\{A_i\} \longrightarrow A_i\}$ of the homomorphisms determined by the product diagram $\{p_i : \Pi\{S(A_i)\} \longrightarrow S(A_i)\}$ of $\{S(A_i)\}$ in $\mathbf{S}$, is a product diagram of $\Pi\{A_i\}$ in $\mathbf{S}^W$.

The object $\Sigma(A_i)$ which is defined by

$$S(\Sigma\{A_i\}) = \Sigma\{S(A_i)\} \text{ and } \lambda_{\Sigma\{A_i\}} = \Sigma\{\lambda_{A_i}\}$$

is a sum object of $\{A_i\}$ in $\mathbf{S}^W$. Furthermore, the family $\{q_i : A_i \longrightarrow \Sigma\{A_i\}\}$ of the homomorphisms determined by the sum diagram $\{q_i : S(A_i) \longrightarrow \Sigma\{S(A_i)\}\}$ of $\{S(A_i)\}$ in $\mathbf{S}$ is a sum diagram of $\Sigma\{A_i\}$ in $\mathbf{S}^W$.

Hence we have

THEOREM 2.2. The category $\mathbf{S}^W$ admits products and sums of any set indexed family of objects.

The special case where all the objects in the family $\{A_i\}$ are identical and equal, say, to A, deserves special attention. For a change, let us denote by T the index set of $\{A_t\}$ where for all $t \in T$, $A_t = A$. In this case, a product object of $\{A_t\}$ can be A^T , where A^T is defined by

$$S(A^T) = S(A)^T$$

the set of all functions $\phi : T \longrightarrow S(A)$, and

$$\phi \circ w : T \longrightarrow S(A)$$

is the function which maps any $t \in T$ on $\phi(t) \circ w$. Similarly, a sum object of $\{A_i\}$ can be $T \cdot A$, where $T \cdot A$ is defined by

$$S(T \cdot A) = T \times S(A) \quad \text{and} \quad (t,s) \circ w = (t, s \cdot w) \text{ .}$$

Note that like in any category with products and sums, the products and sums in $\mathbb{S}^W$ yield functors from products of $\mathbb{S}^W$ to $\mathbb{S}^W$ as follows. Let T be any index set and denote by $(\mathbb{S}^W)^T$ the product category of the families of objects of $\mathbb{S}^W$ indexed by T (i.e., $(\mathbb{S}^W)^T$ is a product object in $\mathbb{C}$at , the category of all categories). The product in $\mathbb{S}^W$ yields therefore a functor

$$\Pi^T : (\mathbb{S}^W)^T \longrightarrow \mathbb{S}^W$$

which maps any family of objects of $\mathbb{S}^W$ indexed by T , on its product object. Similarly the sum in $\mathbb{S}^W$ yields a functor

$$\Sigma^T : (\mathbb{S}^W)^T \longrightarrow \mathbb{S}^W$$

which maps any object of $(\mathbb{S}^W)^T$ on its sum object in $\mathbb{S}^W$.

3. The basic types of the morphisms of $\mathbb{S}^W$ and the forgetful functor $S : \mathbb{S}^W \longrightarrow \mathbb{S}$

The W-semimodules are defined as (right) operations of W on sets. The homomorphisms of W-semimodules are determined by suitable functions on sets. As morphisms of $\mathbb{S}^W$ they may belong to some of the basic types of morphisms in categories (i.e., epic, monic or invertible morphisms, that is isomorphisms). As functions they may be surjective, injective or bijective (or what have you). As we noted in Part I, Chapter I, this situation is analogous to that of many familiar categories.

First, we note that the assignment S from semimodules to their sets of states and from homomorphisms of semimodules to their underlying functions determines a functor $S : \mathbb{S}^W \longrightarrow \mathbb{S}$. Observe that the very definition of $\mathbb{S}^W$ implies that $S : \mathbb{S}^W \longrightarrow \mathbb{S}$, the forgetful functor of $\mathbb{S}^W$, is indeed a functor. Since for any fixed pair of W-semimodules A and B, by definition (Definition 1.2), the underlying functions of the homomorphisms $A \longrightarrow B$ determine the homomorphisms uniquely, the forgetful functor $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ is an embedding. We prove now that $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ is surjective but not injective, unless W is the trivial monoid.

We define the functor $\mathrm{Triv} : \mathbb{S} \longrightarrow \mathbb{S}^W$, where for any set T, $\mathrm{Triv}(T) = T_W$ is the trivial operation of W on T. That is, T_W is defined by

$$S(T_W) = T \quad \text{and} \quad t \circ w = t .$$

Any function $f : T_1 \longrightarrow T_2$ determines a homomorphism $f : (T_1)_W \longrightarrow (T_2)_W$ and we define

$$\mathrm{Triv}(f : T_1 \longrightarrow T_2) = (f : (T_1)_W \longrightarrow (T_2)_W) .$$

Clearly $S \circ \mathrm{Triv}$ is the identity functor of $\mathbb{S}$, and therefore S is surjective. Furthermore, if W is trivial, then every W-semimodule A is $S(A)_W$, and therefore $S(A) = S(B)$ implies $A = B$. That is, S is injective and therefore, in this case, $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ is the inverse of $\mathrm{Triv} : \mathbb{S} \longrightarrow \mathbb{S}^W$ and it is an isomorphism of categories. If W is not trivial, then $W_W \neq M_W$, and yet

$$S(M_W) = S(W_W) = W .$$

From the definition of the morphisms of $\mathbb{S}^W$ (i.e., Definition 1.2) follows also that a morphism of $\mathbb{S}^W$ is an isomorphism iff its underlying function is bijective. In one direction, this statement is trivial because every functor maps isomorphisms on isomorphisms. On the other hand, if the underlying function of a morphism f of $\mathbb{S}^W$ is bijective, then the inverse function f^{-1} of f in $\mathbb{S}$ satisfies also the condition stated in Definition 1.2 for the morphisms in $\mathbb{S}^W$. Hence f must be invertible in $\mathbb{S}^W$; i.e., an isomorphism.

From the fact that $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ is an embedding functor follows that it <u>reflects</u> epic and monic morphisms from $\mathbb{S}$ to $\mathbb{S}^W$. That is, if for a morphism $f : A \longrightarrow B$, $f : S(A) \longrightarrow S(B)$ is surjective (or respectively, injective), then $f : A \longrightarrow B$ is epic (or respectively, monic) in $\mathbb{S}^W$. We prove this statement for the surjective case. The proof for the dual case follows dually. Assume that the underlying function $f : S(A) \longrightarrow S(B)$ of a morphism $f : A \longrightarrow B$ of $\mathbb{S}^W$ is surjective, and that $g_1, g_2 : B \longrightarrow C$ are morphisms with $g_1 f = g_2 f$. Since $f : S(A) \longrightarrow S(B)$ is epic in $\mathbb{S}$, the underlying functions of $g_1 : B \longrightarrow C$ and $g_2 : B \longrightarrow C$ are identical. But $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ is an embedding, and therefore we have $g_1 = g_2$ in $\mathbb{S}^W$, which proves that $f : A \longrightarrow B$ is epic.

In order to prove the converse (i.e., that S preserves epic and monic morphisms from $\mathbb{S}^W$ to $\mathbb{S}$), we need some preliminaries. Note also that the arguments used in the forthcoming proof are analogous to a proof to the effect that the forgetful functor of the category $\mathbb{G}$ of groups preserves epic and monic morphisms.

DEFINITION 3.1. Let A be a W-semimodule. A subset T of $S(A)$ is said to <u>determine</u> the <u>subsystem</u> $A(T)$ of A iff

(i) for all $t \in T$ and $w \in W$, $t \circ w \in T$, and

(ii) $A(T)$ is the W-semimodule defined by $S(A(T)) = T$ and $\lambda_{A(T)}$ is the restriction of λ_A on T .

LEMMA 3.2. If $f : A \longrightarrow B$ is a homomorphism of W-semimodules, then $f(S(A))$ determines a subsystem of B .

We denote by $f(A)$ the subsystem of B determined by $f(S(A))$. Subsystems of semimodules determine <u>quotient</u> systems in a natural manner.

DEFINITION 3.3. Let C be a subsystem of the W-semimodule B . We denote by B/C the object defined by

$$S(B/C) = (S(B)-S(C)) \cup \{s_*\} \quad \text{where} \quad s_* \notin S(B) ,$$

and the transition function of B/C is identical with λ_B for any pair $s \in S(B)$ and $w \in W$ where both s and $\lambda_B(s,w)$ are in $S(B)-S(C)$, otherwise we have $\lambda_{B/C}(s,w) = s_*$.

We define also $q_C : S(B) \longrightarrow S(B/C)$ as the function which maps all of $S(C)$ on s_* , otherwise it is the identity on $S(B)-S(C)$.

It is easy to verify that B/C is a W-semimodule and that $q_C : S(B) \longrightarrow S(B/C)$ determines a homomorphism $q_C : B \longrightarrow B/C$ of W-semimodules. We call it the <u>canonical</u> morphism from B to B/C . It is by no means the only morphism from B to B/C. For example, we denote by $z : B \longrightarrow B/C$ the morphism whose underlying function maps all of $S(B)$ on s_* . Clearly $z = q_C$ iff $C = B$, and then $B/C = \{s_*\}_W$. We call $z : B \longrightarrow B/C$ the <u>trivial</u> morphism from B to B/C .

We prove now

LEMMA 3.4. If $f : A \longrightarrow B$ is an epic morphism of $\mathbf{S}^W$, then $f : S(A) \longrightarrow S(B)$ is surjective.

PROOF. Consider the canonical morphism $q_{f(A)} : B \longrightarrow B/f(A)$ and the trivial morphism $z : B \longrightarrow B/f(A)$. Obviously $f : S(A) \longrightarrow S(B)$ is surjective iff $q_{f(A)} = z$. Both $zf : A \longrightarrow B/f(A)$ and $q_{f(A)}f : A \longrightarrow B/f(A)$ are determined by the same underlying function which maps all of $S(A)$ on s_*. Hence $zf = q_{f(A)}f$ in $\mathbb{S}^W$. But f is epic in $\mathbb{S}^W$, and therefore $z = q_{f(A)}$ and $f : S(A) \longrightarrow S(B)$ is surjective.

A different argument is needed for the proof of the dual of 3.4.

LEMMA 3.5. If $f : A \longrightarrow B$ is a monic morphism of $\mathbb{S}^W$, then $f : S(A) \longrightarrow S(B)$ is injective.

PROOF. For any $s \in S(A)$ denote by $f_s : M_W \longrightarrow A$ the morphism for which $f_s(1) = s$ (and therefore, $f_s(w) = s\circ w$ for all $w \in W$). Let $s_1, s_2 \in S(A)$ be such that $f(s_1) = f(s_2)$. Then, for $f_{s_1}, f_{s_2} : M_W \longrightarrow A$ we have $ff_{s_1} = ff_{s_2}$, since both morphisms yield the same state of B when evaluated at $1 \in W$. But $f : A \longrightarrow B$ is monic in $\mathbb{S}^W$, and therefore $f_{s_1} = f_{s_2}$. Now, by evaluating at $1 \in W$ we get $s_1 = s_2$, and therefore $f : S(A) \longrightarrow S(B)$ is injective.

In conclusion we have

THEOREM 3.6. The forgetful functor $S : \mathbb{S}^W \longrightarrow \mathbb{S}$ relates the basic types of the morphisms of $\mathbb{S}^W$ and $\mathbb{S}$. In particular, a morphism of $\mathbb{S}^W$ is an isomorphism iff it is both monid and epic.

We remark, in passing, that for any W-semimodule A, the set of all the subsystems of A forms a modular lattice with respect to set unions and set intersections. Hence all the composition series theorems hold for W-semimodules. In particular, if C is a subsystem of B, then B/C can be taken as a measure of the interval between B and C for which Jordan-Hölder theorem holds (cf. [5]).

CHAPTER II: A NON-DETERMINISTIC RECONSTRUCTION SCHEME FOR $\mathcal{S}^W$

In this chapter we direct our study of $\mathcal{S}^W$ towards the discovery of a non-deterministic reconstruction scheme for $\mathcal{S}^W$ with M_W as its special object (cf. Part I, Chapter II, Section 13). We begin with a "representation theorem" for $\mathcal{S}^W$ stated in Proposition 4.3. This proposition states that every W-semimodule A is isomorphic to a W-semimodule with $\mathcal{S}^W(M_W,A)$ as its set of states. The transition function of this representation of A is derived from the composition rule of $\mathcal{S}^W$ itself. Unfortunately, unless W satisfies certain strong conditions (e.g., that W has only a trivial automorphism) this representation of A is not defined categorically from $\mathcal{S}^W(M_W,A)$.

We derive this representation from a "representation" functor

$$r^* : \mathcal{S}^W \longrightarrow \mathcal{S}^{W^*} ,$$

where W^* is the endomorphism monoid of M_W in $\mathcal{S}^W$. Since W^* is an isomorphic copy of W, then for any W-semimodule A, every isomorphism $W \longrightarrow W^*$ yields a W-semimodule out of the W^*-semimodule $r^*(A)$. In particular, for the isomorphism $\psi_o : W \longrightarrow W^*$ which is determined by

$$\psi_o = \rho(M_W) : W \longrightarrow \mathcal{S}^W(M_W,M_W) : w \longrightarrow f_w$$

where $f_w : M_W \longrightarrow M_W$ is determined by $f_w(1) = w$, we get a W-semimodule $r_{\psi_o}(A)$ which is isomorphic to A and $S(r_{\psi_o}(A)) = \mathcal{S}^W(M_W,A)$.

Each isomorphism $\psi : W \longrightarrow W^*$ determines a functor $r_\psi : \mathcal{S}^W \longrightarrow \mathcal{S}^W$ with $S \circ r_\psi = \mathcal{S}^W(M_W,-)$. And for $\psi_o = \rho(M_W)$ defined above we get a functor

$$r_{\psi_o} : \mathcal{S}^W \longrightarrow \mathcal{S}^W$$

which is naturally equivalent to the identity functor of $\mathcal{S}^W$.

Our study of $\mathrm{Iso}(W,W^*)$, the set of all monoid isomorphisms from W to W^*, leads us finally to the desired reconstruction scheme for $\$^W$. In particular, we turn to study the group $\mathcal{A}$ of the automorphisms of $\$^W$ which are induced in a natural manner by the automorphisms of the monoid W. A further study of $\mathcal{A}$ and the functors $r_\psi : \$^W \longrightarrow \$^W$, yields a reconstruction scheme for $\$^W$ with M_W as its special object, and which is indexed by $\mathcal{A}$.

4. The representation functor of $\$^W$

We elaborate here on the key idea of the proof of Lemma 3.5.

LEMMA 4.1. The family of mappings

$$\rho(A) : S(A) \longrightarrow \$^W(M_W,A) : s \longrightarrow f_s$$

(where $f_s : M_W \longrightarrow A$ is determined by $f_s(1) = s$), one for each object A of $\$^W$, determines a natural equivalence $\rho : S \longrightarrow \$^W(M_W,-)$. In particular, for any morphism $g : A \longrightarrow B$ of S^W and for any $s \in S(A)$ we have

$$gf_s = f_{g(s)} .$$

PROOF. We prove first that $\rho(A)$ is a bijection. (Note that we used this property of $\rho(A)$ in the proof of Lemma 3.5.) Since

$$f_s(w_1 \circ w_w) = f_s(w_1w_2) = s \circ w_1w_2 = (s \circ w_1) \circ w_2 = f_s(w_1) \circ w_2 ,$$

it follows that for all $s \in S(A)$, $f_s : M_W \longrightarrow A$ is indeed a morphism of $\$^W$. Now if $h : M_W \longrightarrow A$ is any morphism of $\$^W$, then

$$f_{h(1)}(w) = h(1) \circ w = h(w) , \text{ for all } w \in W .$$

Hence $f_{h(1)} = h$, which shows that $\rho(A)$ is surjective. On the other hand, if $f_{s_1} = f_{s_2}$, then by evaluating at 1 we get $s_1 = s_2$. Thus

$\rho(A) : S(A) \longrightarrow \mathbb{S}^W(M_W,A)$ is a bijection.

Finally, for any morphism $g : A \longrightarrow B$ of $\mathbb{S}^W$ we have the equality $gf_s = f_{g(s)}$, because for any $s \in S(A)$ we have $(gf_s)(1) = g(s)$.

Now, since for any morphism $g : A \longrightarrow B$ and any $s \in S(A)$ we have

$$[\rho(B)g](s) = f_{g(s)} \quad \text{and} \quad [\mathbb{S}^W(M_W,-)(g)][\rho(A)](s) = gf_s ,$$

the equality $gf_s = f_{g(s)}$ expresses precisely the requirement that the family of bijections $\rho(A) : S(A) \longrightarrow \mathbb{S}^W(M_W,-)$ determines a natural equivalence $\rho : S \longrightarrow \mathbb{S}^W(M_W,-)$.

If we apply now Lemma 4.1 to $A = M_W$, we get for all $w_1,w_2 \in W$,

$$f_{w_1}f_{w_2} = f_{f_{w_1}(w_2)} = f_{w_1w_2} .$$

This equality now expresses precisely the analog of the Cayley Theorem for monids, because the morphisms $M_W \longrightarrow M_W$ are determined by the left translations of W. Thus we have

COROLLARY 4.2. Let W^* denote the endomorphism monoid of M_W in $\mathbb{S}^W$. The bijection $\rho(M_W) : W \longrightarrow \mathbb{S}^W(M_W,M_W)$ determines a monoid isomorphism $W \longrightarrow W^*$.

So far we have the following features of the functor $\mathbb{S}^W(M_W,-)$. First, for any object A of $\mathbb{S}^W$, the set $\mathbb{S}^W(M_W,A)$ is isomorphic to the set $S(A)$ of states of A. And, the set isomorphism from the carrier of W (which is the set of states of M_W) to $\mathbb{S}^W(M_W,M_W)$ determines a monoid isomorphism from W to W^*. We need now a suitable definition for a function

$$\mathbb{S}^W(M_W,A) \times \mathbb{S}^W(M_W,M_W) \longrightarrow \mathbb{S}^W(M_W,A) ,$$

that will yield a transition function of a semimodule over W^* with $\mathbb{S}^W(M_W,A)$ as its set of states. The composition rule in $\mathbb{S}^W$ seems to be a good suggestion.

We define an assignment r^* from the objects of $\mathcal{S}^W$ to the objects of $\mathcal{S}^{W^*}$. For any W-semimodule A, the object $r^*(A)$ is the W^*-semimodule defined by

$$S(r^*(A)) = \mathcal{S}^W(M_W, A) \quad \text{and} \quad f \circ h = fh$$

for all $f \in \mathcal{S}^W(M_W, A)$ and $h \in \mathcal{S}^W(M_W, M_W)$. It is easy to verify that $r^*(A)$ is indeed a semimodule over W^*. Furthermore, for any morphism $f : A \longrightarrow B$ of $\mathcal{S}^W$, the function

$$[\mathcal{S}^W(M_W, -)](f) : \mathcal{S}^W(M_W, A) \longrightarrow \mathcal{S}^W(M_W, B) : f_s \longrightarrow ff_s ,$$

determines a morphism of $\mathcal{S}^{W^*}$. This follows simply from the associativity of the composition rule of $\mathcal{S}^W$: $f(f_s h) = (ff_s)h$. We denote the morphism of $\mathcal{S}^{W^*}$ determined by $f : A \longrightarrow B$ by

$$r^*(f) : r^*(A) \longrightarrow r^*(B) .$$

In conclusion, we have a functor

$$r^* : \mathcal{S}^W \longrightarrow \mathcal{S}^{W^*}$$

for which $S \circ r^* = \mathcal{S}^W(M_W, -)$. We call this functor, the <u>representation</u> functor of $\mathcal{S}^W$. Now Lemma 4.1 implies directly the equality

$$f_{s \circ w} = f_{f_s(w)} = f_s f_w ,$$

valid for all $s \in S(A)$ and $w \in W$. Hence we have that

$$\begin{array}{ccccc}
S(A) & \times & W & \xrightarrow{\lambda_A} & S(A) \\
\downarrow{\scriptstyle \rho(A)} & & \downarrow{\scriptstyle \rho(M_W)} & & \downarrow{\scriptstyle \rho(A)} \\
\mathcal{S}^W(M_W, A) & \times & \mathcal{S}^W(M_W, M_W) & \xrightarrow[\lambda_{r^*(A)}]{} & \mathcal{S}^W(M_W, M_W)
\end{array}$$

commutes for any W-semimodule A . Since $\rho(A)$ is a bijection and $\rho(M_W)$ is an isomorphism of monoids, we can say that $r^*(A)$ represents A in terms of the morphisms and their composition rule in $\mathbb{S}^W$.

Unfortunately, W^* is not identical with W , and therefore $r^*(A)$ is not an object of $\mathbb{S}^W$. Still we may interpret W^* as W , according to various possible isomorphisms from W onto W^* . For example, let

$$\psi_o : W \longrightarrow W^*$$

be the isomorphism determined by $\rho(M_W) : W \longrightarrow \mathbb{S}^W(M_W, M_W)$ (cf. Corollary 4.2). Furthermore, let $r_{\psi_o}(A)$ be the W-semimodule defined by

$$S(r_{\psi_o}(A)) = \mathbb{S}^W(M_W, A) \text{ , and}$$

$$\lambda_{r_{\psi_o}(A)}(f_s, w) = f_s f_w = \lambda_{r^*(A)}(f_s, \psi_o(w)) \text{ .}$$

Lemma 4.1 implies directly the "representation theorem" stated in the following proposition.

PROPOSITION 4.3. With the notation introduced above, for any W-semimodule A , the function

$$\rho(A) : S(A) \longrightarrow S(r_{\psi_o}(A)) : s \longrightarrow f_s$$

determines an isomorphism $\rho_{\psi_o}(A) : A \longrightarrow r_{\psi_o}(A)$ of $\mathbb{S}^W$.

It is clear that if there were only a single isomorphism $W \longrightarrow W^*$, then $r_{\psi_o}(A)$ is defined categorically from $\mathbb{S}^W(M_W, A)$. In this case, $r_{\psi_o}(A)$ tells us everything that can be said about A in terms of M_W and additional categorical predicates in $\mathbb{S}^W$. Such a case happens when W has a trivial automorphism group. For note that if ψ_1 and ψ_2 are two different isomorphisms from W onto W^* , then $\psi_1\psi_2^{-1}$ and $\psi_2\psi_1^{-1}$ are two non-trivial automorphisms of W . Before we elaborate on the effect of Aut(W) on the possible interpretations of $r^*(A)$ as a W-semimodule, we define these interpretations as certain functors $r_\psi : \mathbb{S}^W \longrightarrow \mathbb{S}^W$ induced by $r^* : \mathbb{S}^W \longrightarrow \mathbb{S}^{W^*}$ and the isomorphisms $\psi : W \longrightarrow W^*$.

5. The reconstruction functors $r_\psi : \mathbb{S}^W \longrightarrow \mathbb{S}^W$

Here and in the rest of this chapter we will define several families of functors by means of the usual schemes of definition for semimodules and their homomorphisms that we have employed so far. We will leave for the reader the routine task of verifying that the defined entities are, in fact, W-semimodules, homomorphisms of seminodules, or functors, as the case will be. For the simplicity of our notation we will denote the functor $\mathbb{S}^W(M_W,-) : \mathbb{S}^W \longrightarrow \mathbb{S}$ by $r : \mathbb{S}^W \text{——} \mathbb{S}$.

DEFINITION 15.1. Let $\psi : W \longrightarrow W^*$ be an isomorphism of monoids. The functor

$$r_\psi : \mathbb{S}^W \longrightarrow \mathbb{S}^W$$

is defined as follows. For any W-semimodule A , $r_\psi(A)$ is defined by

$$S(r_\psi(A)) = r(A) \quad \text{and} \quad \lambda_{r_\psi(A)}(f_s,w) = f_s\psi(w) \ .$$

For any morphism $f : A \longrightarrow B$ of $\mathbb{S}^W$, the morphism $r_\psi(f) : r_\psi(A) \longrightarrow r_\psi(B)$ is defined to be the homomorphism whose underlying function is $r(f) : r(A) \longrightarrow r(B)$.

Note that for any isomorphism $\psi : W \longrightarrow W^*$ we have $S \circ r_\psi = r = \mathbb{S}^W(M_W,-)$. Furthermore the transition function of $r_\psi(A)$ is related to the transition function of $r^*(A)$ by the equality

$$\lambda_{r_\psi(A)}(f_s,w) = \lambda_{r^*(A)}(f_s,\psi(w))$$

valid for any W-semimodule A, $s \varepsilon S(A)$ and $w \varepsilon W$. The relationship between the isomorphisms $W \longrightarrow W^*$ and the automorphisms of W itself is implied from the following lemma. Since the proof of this lemma is straight forward, we spell it out only for one subcase, while the rest of the proof is left for the reader.

LEMMA 5.2. Let $x : W \longrightarrow W$ be any function (i.e., from the carrier of W to itself). We define the mapping

$$h_x : W \longrightarrow S^W(M_W, M_W)$$

by

$$[h_x(w)](1) = x(w) \text{ ; i.e., } h_x(w) = f_{x(w)} \text{ .}$$

Then we have:

(i) The correspondence $x \longrightarrow h_x$ is a bijection from the set of <u>functions</u> $W \longrightarrow W$ to the set of <u>functions</u> $W \longrightarrow S^W(W,W)$.

(ii) For any $x_1, x_2 : W \longrightarrow W$ we have

$$h_{x_1} \circ x_2 = h_{x_1 \circ x_2} \text{ .}$$

(iii) The function $h_x : W \longrightarrow S^W(M_W, M_W)$ determines a homomorphism of monoids $h_x : W \longrightarrow W^*$ iff $x : W \longrightarrow W$ itself determines a homomorphism of monoids.

(iv) The function h_x is surjective (respectively, injective or bijective) iff x itself is surjective (respectively, injective or bijective).

PROOF. We prove that $h_x : W \longrightarrow S^W(M_W, M_W)$ is surjective iff $x : W \longrightarrow W$ is. Assume that h_x is surjective. Then for any morphism $f_{w_1} : M_W \longrightarrow M_W$ there exists a $w_2 \varepsilon W$ such that $h_x(w_2) = f_{w_1}$. By the evaluation at 1 we get that for any $w_1 \varepsilon W$ there exists a $w_2 \varepsilon W$

such that

$$x(w_2) = [h_x(w_2)](1) = f_{w_1}(1) = w_1 .$$

Hence $x : W \longrightarrow W$ is surjective. On the other hand, if $x : W \longrightarrow W$ is surjective, then for any $w_1 \in W$ there exists a $w_2 \in W$ with $x(w_2) = w_1$. Hence for any $w_1 \in W$ there exists a $w_2 \in W$ such that

$$h_x(w_2) = f_{x(w_2)} = f_{w_1} ,$$

which implies that $h_x : W \longrightarrow W^*$ is surjective.

The rest of the assertions of this lemma are proven by similar arguments. Note that Lemma 4.1 implies assertion (iii). By combining assertions (iii) and (iv) of Lemma 5.2, we derive the following corollary.

COROLLARY 5.3. Let $\mathrm{Aut}(W)$ be the automorphism group of W and let $\mathrm{Iso}(W,W^*)$ be the set of isomorphisms from W onto W^*. The correspondence $\alpha \longrightarrow h_\alpha$ determines a bijection from $\mathrm{Aut}(W)$ onto $\mathrm{Iso}(W,W^*)$.

The relationship between this bijection and the functors $r_\psi : \$^W \longrightarrow \$^W$ is expressed by the following immediate lemma.

LEMMA 5.4. For any $\alpha \in \mathrm{Aut}(W)$, W-semimodule A, $s \in S(A)$ and $w \in W$, we have

$$\lambda_{r_{h_\alpha}(A)}(f_s,w) = f_s h_\alpha(w) = f_s f_{\alpha(w)} = f_{s\circ\alpha(w)} .$$

If, in addition, $\beta \in \mathrm{Aut}(W)$, then

$$\lambda_{r_{h_\alpha}(A)}(f_s,w) = \lambda_{r_{h_\beta}(A)}(f_s,\beta^{-1}\alpha(w)) .$$

Hence, in particular,

$$\lambda_{r_{h_\alpha}(A)}(f_s,w) = \lambda_{r_{\psi_o}(A)}(f_s,\alpha(w)) \ .$$

Thus, every "reconstruction" functor $r_\psi : \$^W \longrightarrow \$^W$ can be derived from r_{ψ_o} by "modifying the input by an automorphism α of W" (i.e., by $\alpha \in \mathrm{Aut}(W)$ for which $h_\alpha = \psi_o$.) This relation of the functors $r_\psi : \$^W \longrightarrow \$^W$ is especially important because of the following property of r_{ψ_o} which increases the import of Proposition 4.3. Lemma 4.1 implies not only that $\rho(A) : S(A) \longrightarrow S(r_{\psi_o}(A))$ determines an isomorphism $\rho_{\psi_o}(A) : A \longrightarrow r_{\psi_o}(A)$, for all W-semimodules A, but also that this family of isomorphisms is natural. Namely, the commutative diagram for the natural equivalence $\rho : S \longrightarrow \$^W(M_W,-)$, that is, the equality $gf_s = f_{g(s)}$, implies directly the proof of the following lemma (cf. Lemma 4.1).

LEMMA 5.5. The family of isomorphisms $\rho_{\psi_o}(A) : A \longrightarrow r_{\psi_o}(A)$, for all objects A of $\$^W$, determines a natural equivalence

$$\rho_{\psi_o} : I \longrightarrow r_{\psi_o}$$

from the identity functor I of $\$^W$ to $r_{\psi_o} : \$^W \longrightarrow \$^W$. Furthermore, we have $S \circ \rho_{\psi_o} = \rho$ (i.e., $S(\rho_{\psi_o}(A)) = \rho(A)$ for all objects A of $\$^W$).

6. The automorphisms of $\$^W$ induced by Aut(W)

The relationships among the reconstruction functors of $\$^W$, as expressed in Lemma 5.4, lead us to consider transformations of W-semimodules which are induced by functions $x : W \longrightarrow W$. In particular, we will be interested in those translations which are induced by $\mathrm{Aut}(W)$. So let A be an arbitrary W-semimodule and $x : W \longrightarrow W$ be any function. We define a system $A^x = (\lambda_A^x : S(A) \mathrm{x} W \longrightarrow S(A))$ by

$$\lambda_A^x(s,w) = \lambda_A(s,x(w)) \ .$$

We want to find out under what conditions on $x : W \longrightarrow W$, A^x is a W-semimodule for any W-semimodule A. In order to do so, we examine $\lambda^x_{M_W}$. In particular, we examine $\lambda^x_{M_W}(s,w)$ for $s = 1$, the identity element of W. By the definition of $\lambda^x_{M_W}$, we have

$$\lambda^x_{M_W}(1,w) = x(w) .$$

Hence M^x_W is a W-semimodule iff $x : W \longrightarrow W$ is a homomorphism of monoids. We denote by $\mathrm{End}(W)$ the endomorphism monoid of W.

DEFINITION 6.1. For any $\psi \in \mathrm{End}(W)$ we define the functor $T_\psi : \mathcal{S}^W \longrightarrow \mathcal{S}^W$ as follows. For any object A of $\mathcal{S}^W$, $T_\psi(A)$ is defined by

$$S(T_\psi(A)) = S(A) \quad \text{and} \quad \lambda_{T_\psi(A)}(s,w) = \lambda_A(s,\psi(w)) .$$

For any morphism $f : A \longrightarrow B$, the morphism $T_\psi(f) : T_\psi(A) \longrightarrow T_\psi(B)$ is the homomorphism from $T_\psi(A)$ to $T_\psi(B)$ with the <u>same</u> underlying function as $f : A \longrightarrow B$.

It follows immediately from the definition of $T_\psi : \mathcal{S}^W \longrightarrow \mathcal{S}^W$ that it is an embedding functor and $S \circ T_\psi = S$. Note that in our discussion of the properties of the forgetful functor of $\mathcal{S}^W$ (cf. Section 3), we noted the fact that $T_\psi(M_W) \neq M_W$ whenever W is not the trivial monoid and where $\psi : W \longrightarrow W$ is the trivial endomorphism of W. The following lemma establishes the fact that the set of the functors $T_\psi : \mathcal{S}^W \longrightarrow \mathcal{S}^W$ induced by $\mathrm{End}(W)$ is closed under the composition of functors, and therefore, with this composition, it is a monoid.

LEMMA 6.2. If $\psi_1, \psi_2 \in \mathrm{End}(W)$, then

$$T_{\psi_1} \circ T_{\psi_2} = T_{\psi_2 \circ \psi_1} .$$

PROOF. All we have to show is that for any W-semimodule A we have the equality

$$\lambda_{[T_{\psi_1}\circ T_{\psi_2}](A)} = \lambda_{T_{\psi_2\circ\psi_1}(A)} .$$

So let $s \varepsilon S(A)$ and $w \varepsilon W$, then we have

$$\begin{aligned}\lambda_{[T_{\psi_1}\circ T_{\psi_2}](A)}(s,w) &= \lambda_{T_{\psi_1}(T_{\psi_2}(A))}(s,w) \\ &= \lambda_{T_{\psi_2}(A)}(s,\psi_1(w)) \\ &= \lambda_A(s,[\psi_2\circ\psi_1](w)) \\ &= \lambda_{T_{\psi_2\circ\psi_1}(A)}(s,w) .\end{aligned}$$

By evaluating at M_W, and in particular for the state 1 of M_W, we get that $T_{\psi_1} = T_{\psi_2}$ implies $\psi_1 = \psi_2$. Hence, the mapping $\psi \longrightarrow T_\psi$ determines an anti-isomorphism of End(W) with the monoid of the functors $T_\psi : \mathcal{S}^W \longrightarrow \mathcal{S}^W$ induced by End(W). In particular we have the following corollary.

COROLLARY 6.3. If $\alpha \varepsilon \mathrm{Aut}(W)$, then T_α is an automorphism of $\mathcal{S}^W$ with $T_\alpha\circ T_{\alpha^{-1}} = T_{\alpha^{-1}}\circ T_\alpha = 1$.

We can rephrase now the essential part of Lemma 5.4 by the following lemma which also explicates the meaning of "modifying the input by an automorphism α of W".

LEMMA 6.4. For any $\alpha,\beta \varepsilon \mathrm{Aut}(W)$ we have

$$T_\alpha\circ r_{h_\beta} = r_{h_{\beta\circ\alpha}} ,$$

and, in particular, since h_{ψ_o} is the identity automorphism of W, we have

$$T_\alpha \circ r_{\psi_o} = r_{h_\alpha} .$$

From Lemma 5.5 we know that $\rho_{\psi_o} : I \longrightarrow r_{\psi_o}$ is a natural equivalence with $S \circ \rho_{\psi_o} = \rho$. By the properties of $\mathbb{N}$at (cf. Part I, Chapter II, Section 6), we derive the following corollary of Lemma 5.5 and Lemma 6.4.

COROLLARY 6.5. For any $\alpha \in \mathrm{Aut}(W)$ we have a natural equivalence

$$T_\alpha \circ \rho_{\psi_o} : T_\alpha \longrightarrow r_{h_\alpha} ,$$

where $[T_\alpha \circ \rho_{\psi_o}](A) = T_\alpha(\rho_{\psi_o}(A))$ for all objects A of $\mathbb{S}^W$, and $S \circ [T_\alpha \circ \rho_{\psi_o}] = \rho$.

Our goal in the next section is to show that the reconstruction functors $r_\psi : \mathbb{S}^W \longrightarrow \mathbb{S}^W$ can be extended into a non-deterministic reconstruction scheme $\langle M_W, R_\alpha : \alpha \in \mathcal{A} \rangle$ for $\mathbb{S}^W$ indexed by the group $\mathcal{A}$ of the automorphisms of $\mathbb{S}^W$ induced by the automorphisms of W. In particular we will have

$$R_\alpha(M_W) = r_{h_\alpha} .$$

Before we do so, we examine $\mathcal{A}^*$, the quotient group of $\mathcal{A}$ under natural equivalence.

LEMMA 6.6. For any $\psi \in \mathrm{End}(W)$, the underlying function of the endomorphism $\psi : W \longrightarrow W$ determines a homomorphism

$$\psi : M_W \longrightarrow T_\psi(M_W)$$

of W-semimodules.

PROOF. From $\psi(w_1 w_2) = \psi(w_1)\psi(w_2)$ follows that

$$\psi(\lambda_{M_W}(w_1, w_2)) = \lambda_{T_\psi(M_W)}(\psi(w_1), w_2) .$$

COROLLARY 6.7. If $\alpha \in \mathrm{Aut}(W)$, then $\alpha : M_W \longrightarrow T_\alpha(M_W)$ is an isomorphism in $\mathcal{S}^W$. Hence if $F : M_W \longrightarrow M_W$ is any functor equivalent to T_α, for some $\alpha \in \mathrm{Aut}(W)$, then M_W is isomorphic to $F(M_W)$ in $\mathcal{S}^W$.

Assume now that $\tau : I \longrightarrow T_\alpha$ is a natural equivalence of the identity functor of $\mathcal{S}^W$ with T_α, for some $\alpha \in \mathrm{Aut}(W)$. We evaluate the commutative diagram for $\tau : I \longrightarrow T_\alpha$ at M_W and in particular at $1 \in S(M_W)$ for an arbitrary morphism $f_w : M_W \longrightarrow M_W$. Denote by τ_α the isomorphism $\tau(M_W) : M_W \longrightarrow T_\alpha(M_W)$ given by the transformation $\tau : I \longrightarrow T$. Note that although we do not necessarily have $\tau_\alpha = \alpha$, we must have

$$\tau_\alpha(w_1 w_2) = \tau_\alpha(w_1)\alpha(w_2) \quad \text{and} \quad \tau_\alpha(w) = \tau_\alpha(1)\alpha(w) .$$

Furthermore, $\tau_\alpha : W \longrightarrow W$, the underlying function of the isomorphism $\tau(M_W) : M_W \longrightarrow T_\alpha(M_W)$, is of course injective.

Since $\tau : I \longrightarrow \tau_\alpha$ is natural, we have

$$\tau_\alpha f_w = f_w \tau_\alpha$$

for any morphism $f_w : M_W \longrightarrow M_W$. Thus, in particular, we have

$$\tau_\alpha(w) = [\tau_\alpha f_w](1) = [f_w \tau_\alpha](1) = \tau_\alpha(1)w = \tau_\alpha(\alpha^{-1}(w)) .$$

But τ_α is injective, and therefore, $w = \alpha^{-1}(w)$, which implies that α is the identity automorphism of W. Since T_α is naturally equivalent to T_β (for $\alpha, \beta \in \mathrm{Aut}(W)$) iff $T_{\alpha^{-1}} \circ T_\beta$ is naturally equivalent to the identity functor of $\mathcal{S}^W$, and $T_{\alpha^{-1}} \circ T_\beta = T_{\beta \circ \alpha^{-1}}$, we have concluded the proof of the following proposition.

PROPOSITION 6.8. Let $\mathcal{A}$ be the group of the automorphisms T_α of $\mathcal{S}^W$ induced by $\alpha \in \mathrm{Aut}(W)$, and let $\mathcal{A}^*$ be the quotient group of $\mathcal{A}$

under natural equivalence. Then $\mathcal{A}^*$ is isomorphic to $\mathcal{A}$; that is, for any $\alpha, \beta \in \mathrm{Aut}(W)$, T_α is naturally equivalent to T_β iff $\alpha = \beta$. Furthermore, if we denote by $[T_\alpha]$ the natural equivalence class of T_α in the total class group of the automorphisms of $\mathbf{S}^W$, then the mapping $\alpha \longrightarrow [T_{\alpha^{-1}}]$ determines a monomorphism (i.e., monic morphism) of $\mathrm{Aut}(W)$ into $\mathrm{Aut}(\mathbf{S}^W)$.

Thus we know now that every categorical class in $\mathbf{S}^W$ is closed under the automorphisms in $\mathcal{A}$. The derivation of a reconstruction scheme for $\mathbf{S}^W$ with the special object M_W and indexed by $\mathcal{A}$, will reduce the isomorphism of $\mathcal{A}$ with $\mathrm{Aut}(\mathbf{S}^W)$ to the categoricity of the isomorphism type of M_W in $\mathbf{S}^W$.

7. The reconstruction scheme for $\mathbf{S}^W$

A functor $F : \mathbf{S}^W \longrightarrow \mathbf{S}^W$ is said to be M_W-invariant iff M_W is isomorphic to $F(M_W)$ in $\mathbf{S}^W$. The semimodule M_W is said to be distinguishable in $\mathbf{S}^W$ iff every automorphism of $\mathbf{S}^W$ is M_W-invariant; that is, iff the isomorphism type of M_W is categorical in $\mathbf{S}^W$. Before we discuss the reconstruction scheme that we have in mind for $\mathbf{S}^W$, let us make an abuse of our notation, and denote by α both the automorphism of W and the automorphism T_α of $\mathbf{S}^W$ induced by α. According to our notation as introduced in Part I, Chapter II, Section 13, we denote by $\mathbf{S}^W[M_W]$ the class of all semimodules $F(M_W)$ where F is any M_W-invariant automorphism of $\mathbf{S}^W$. We have to define now, for each $\alpha \in \mathcal{A}$ and each $M \in \mathbf{S}^W[M_W]$ a functor

$$R_\alpha(M) : \mathbf{S}^W \longrightarrow \mathbf{S}^W .$$

In fact, we first define a slightly different class of functors.

DEFINITION 7.1. For any M_W-invariant automorphism F of $\mathcal{S}^W$, and any isomorphism $\phi : M_W \longrightarrow F(M_W)$ of $\mathcal{S}^W$, we define the functor

$$r(\phi,F) : \mathcal{S}^W \longrightarrow \mathcal{S}^W$$

as follows. For any W-semimodule A, the semimodule $r(\phi,F)(A)$ is defined by

$$S(r(\phi,F)(A)) = \mathcal{S}^W(F(M_W),F(A)) \text{ , and}$$

$$\lambda(F(f_s),w) = F(f_s)(\phi\psi_o(w)\phi^{-1}) \text{ ,}$$

where λ denotes the transition function of $r(\phi,F)(A)$. For any morphism $g : A \longrightarrow B$ of $\mathcal{S}^W$ we define

$$r(\phi,F)(g) : r(\phi,F)(A) \longrightarrow r(\phi,F)(B)$$

to be the homomorphism whose underlying function is

$$\mathcal{S}^W(F(M_W),F(A)) \longrightarrow \mathcal{S}^W(F(M_W),F(B)) : F(f_s) \longrightarrow F(g)F(f_s) \text{ ;}$$

i.e., $r(\phi,F)(g) = r_{\psi_o}(F(g))$.

Recall that $\psi_o : W \longrightarrow W^*$ is the isomorphism determined by $\rho(M_W) : W \longrightarrow \mathcal{S}^W(M_W,M_W)$; $w \longrightarrow f_w$. We prove now the following lemma.

LEMMA 7.2. The transformation

$$\tau(\phi,F) : r_{\psi_o} \circ F \longrightarrow r(\phi,F)$$

defined by

$$\tau(\phi,F)(A) : \mathcal{S}^W(M_W,F(A)) \longrightarrow \mathcal{S}^W(F(M_W),F(A)) : g \longrightarrow g\phi^{-1} \text{ ,}$$

is a natural equivalence.

PROOF. Obviously $\tau(\phi,F)(A)$ is bijective. Let us verify that it determines a homomorphism of W-semimodules.

If $f : M_W \longrightarrow F(A)$ is morphism of $\mathbb{S}^W$ and $w \in W$, then we have

$$[\tau(\phi,F)(A)](\lambda_{r_{\psi_o}(F(A))}(f,w)) = [\tau(\phi,F)(A)](f\psi_o(w))$$

$$= f\psi_o(w)\phi^{-1} ;$$

and

$$\lambda([\tau(\phi,F)(A)](f),w) = \lambda(f\phi^{-1},w)$$

$$= (f\phi^{-1})(\phi^{-1}\psi_o(w)\phi^{-1})$$

$$= f\psi_o(w)\phi^{-1} .$$

Thus, for any W-semimodule A , $\tau(\phi,F)(A)$ is an isomorphism of W-semimodules. We prove now that it is natural.

If $g : A \longrightarrow B$ is a morphism of $\mathbb{S}^W$, then for any morphism $f : M_W \longrightarrow F(A)$ we have

$$[\tau(\phi,F)(B)\circ(r_{\psi_o}\circ F)(g)](f) = [\tau(\phi,F)(B)](F(g)f)$$

$$= F(g)f\phi^{-1} ;$$

and

$$[r(\phi,F)(g)\circ\tau(\phi,F))A)](f) = [r(\phi,F)(g)](f\phi^{-1})$$

$$= [(r_{\psi_o}\circ F)(g)](f\phi^{-1})$$

$$= F(g)f\phi^{-1} .$$

Hence $\tau(\phi,F) : r_{\psi_o}\circ F \longrightarrow r(\phi,F)$ is a natural equivalence.

Note that if F is the identity functor I of $\mathbb{S}^W$ and ϕ is the identity morphism of M_W, then $r(\phi,F)$ is precisely r_{ψ_o} and $\tau(\phi,F)$ is the identity natural equivalence. If instead of $\psi_o : W \longrightarrow W^*$ we employ any arbitrary isomorphism $\psi \in \mathrm{Iso}(W,W^*)$ in the definition of $r(\phi,F)$, we get a family of functors

$$r_\psi(\phi,F) : \mathbb{S}^W \longrightarrow \mathbb{S}^W .$$

Namely the transition function λ_ψ of $r_\psi(\phi,F)(A)$ is defined by

$$\lambda_\psi(F(f_s),w) = F(f_s)(\phi\psi(w)\phi^{-1}) .$$

It is easy to verify that now we have a natural equivalence

$$\tau_\psi(\phi,F) : r_\psi \circ F \longrightarrow r_\psi(\phi,F) ,$$

defined in a complete analogy with $\tau(\phi,F)$. Furthermore, it is clear that we have

$$r_{h_\alpha}(\phi,F) = \alpha \circ r(\phi,F) ,$$

for any $\alpha \in \mathcal{A}$. Equivalently, we can take the last equality as the definition of $r_\psi(\phi,F)$, and then by the properties of $\mathbb{N}\mathrm{at}$, we derive that

$$\alpha \circ \tau(\phi,F) : \alpha \circ r_{\psi_o} \circ F \longrightarrow \alpha \circ r(\phi,F)$$

is a natural equivalence. But $\alpha \circ r_{\psi_o} = r_{h_\alpha}$ and $\alpha \circ r(\phi,F) = r_{h_\alpha}(\phi,F)$, and so we have a natural equivalence

$$\alpha \circ \tau(\phi,F) : r_{h_\alpha} \circ F \longrightarrow r_{h_\alpha}(\phi,F) .$$

Finally, we prove the following lemma that is needed in order that we can choose from the family of functors $r_\psi(\phi,F) : \mathbb{S}^W \longrightarrow \mathbb{S}^W$ a reconstruction scheme for $\mathbb{S}^W$.

LEMMA 7.3. Let F be an M_W-invariant automorphism of $\mathbb{S}^W$, $\phi : M_W \longrightarrow F(M_W)$ an isomorphism in $\mathbb{S}^W$ and $\psi \in \mathrm{Iso}(W,W^*)$. The functor F restricted to $\mathbb{S}^W(M_W,-)$, determines a natural equivalence

$$\tau_F : r_{\psi(\phi,F)} \longrightarrow r_\psi(\phi,F) ,$$

where $\psi(\phi,F) \in \mathrm{Iso}(W,W^*)$ is determined by

$$[\psi(\phi,F)] = F^{-1}(\phi\psi(w)\phi^{-1}) .$$

PROOF. By Lemma 5.2 and the fact that functors preserve isomorphisms, it follows that $\psi(\phi,F)$ is an isomorphism of W onto W^*. We prove next that for any W-semimodule A, the function

$$\tau_F(A) : \mathbb{S}^W(M_W,A) \longrightarrow \mathbb{S}^W(F(M_W),F(A)) : f_s \longrightarrow F(f_s) ,$$

which is obviously a bijection, determines a homomorphism of W-semimodules. So let $f_s : M_W \longrightarrow A$ be a morphism and $w \in W$. Then

$$\begin{aligned} [\tau_F(A)](\lambda_{r_{\psi(\phi,F)}(A)}(f_s,w)) &= [\tau_F(A)](f_s\psi(\phi,F)) \\ &= F(f_s\psi(\phi,F)) \\ &= F(f_s)\phi\psi(w)\phi^{-1} ; \end{aligned}$$

and

$$\begin{aligned} \lambda_{r_\psi(\phi,F)(A)}([\tau_F(A)](f_s),w) &= \lambda_\psi(F(f_s),w) \\ &= F(f_s)\phi\psi(w)\phi^{-1}. \end{aligned}$$

Hence $\tau_F(A) : r_{\psi(\phi,F)}(A) \longrightarrow r_\psi(\phi,F)(A)$ is an isomorphism in $\mathbb{S}^W$. Let now $g : A \longrightarrow B$ be any morphism of $\mathbb{S}^W$ and let $f_s : M_W \longrightarrow A$ be any state of $r_{\psi(\phi,F)}(A)$. Then we have

$$[r_\psi(\phi,F)(g)\circ\tau_F(A)](f_s) = [r_\psi(\phi,F)(g)](F(f_s))$$

$$= [r_{\psi_o}(F(g))](F(f_s))$$

$$= F(g)F(f_s) \ ;$$

and

$$[\tau_F(B)\circ r_{\psi(\phi,F)}(g)](f_s) = [\tau_F(B)](gf_s)$$

$$= F(gf_s)$$

$$= F(g)F(f_s) \ .$$

Hence τ_F is a natural equivalence.

The families of natural equivalences

$$\alpha\circ\tau(\phi,F) : r_{h_\alpha}\circ F \longrightarrow r_{h_\alpha}(\phi,F) \ ,$$

and

$$\tau_F : r_{\psi(\phi,F)} \longrightarrow r_\psi(\phi,F)$$

leave us much freedom in choosing a reconstruction scheme out of the family of functors

$$r_\psi(\phi,F) : \mathbb{S}^W \longrightarrow \mathbb{S}^W \ .$$

For example, for any $\alpha \in \mathcal{A}$ and any $M \in S^W[M_W]$ we define

$$R_\alpha(M) = r_{h_\alpha}(\phi,F)$$

where $M = F(M_W)$ and $\phi : M_W \longrightarrow F(M_W)$ is any isomorphism, with the provision that if $M = M_W$ then both F and ϕ are identities. Lemma 7.2 and its implications and Lemma 7.3 imply that $\langle M, R_\alpha : \alpha \in \mathcal{A}\rangle$ thus defined is a non-deterministic reconstruction scheme for $\mathbb{S}^W$ with the special object M_W (cf. Part I, Chapter II, Section 13). In particular, we can infer, either from the properties of reconstruction schemes (ibid, Lemma 13.2) or directly from Lemmata 7.2 and 7.3, that an automorphism F of $\mathbb{S}^W$ is naturally equivalent to T_α, for some $\alpha \in \mathrm{Aut}(W)$, iff F is M_W-invariant.

In conclusion, we have proved the following reconstruction theorem for $\mathbb{S}^W$.

THEOREM 7.4. Let W be any monoid. The mapping $\alpha \longrightarrow [T_{\alpha^{-1}}]$, where $[T_{\alpha^{-1}}]$ **is the natural equivalence class of** $T_{\alpha^{-1}}$ **in the total class** group of the automorphisms of $\mathbb{S}^W$, is a group isomorphism of $\mathrm{Aut}(W)$ onto $\mathrm{Aut}(\mathbb{S}^W)$ iff M_W is distinguishable in $\mathbb{S}^W$. In particular, if M_W is distinguishable in $\mathbb{S}^W$, then a natural class of morphisms in $\mathbb{S}^W$ is categorical iff it is closed under $\mathcal{A}$, the group of the automorphisms of $\mathbb{S}^W$ induced by $\mathrm{Aut}(W)$.

Thus, we know how to compute $\mathrm{Aut}(\mathbb{S}^W)$ provided that we know that M_W is distinguishable in $\mathbb{S}^W$. As for the group $\mathrm{Ob}(\mathbb{S}^W)$, we know that if W is free then $\mathrm{Ob}(\mathbb{S}^W)$ is isomorphic to $\mathrm{Aut}(\mathbb{S}^W)$. We indicate the proof for the case where W is the free monoid generated by two elements, say a and b. Let A be the W-semimodule determined by

$$S(A) = \{s_a, s_b\},$$

$$\lambda_A(s_a, a) = s_a, \quad \lambda_A(s_a, b) = s_b,$$

and

$$\lambda_A(s_b, a) = \lambda_A(s_b, b) = s_b.$$

The monoid W has two automorphisms, the identity and the one determined by the interchanging of a with b . For any automorphism α of W , $T_\alpha(A)$ is isomorphic to A iff α is the identity automorphism of W . If W is any free monoid, then for any $\alpha \in \mathrm{Aut}(W)$ there exists a W-semimodule A_α such that A_α and $T_\alpha(A_\alpha)$ are not isomorphic. Hence, in this case, $\mathrm{Aut}(\mathbf{S}^W)$ and $\mathrm{Ob}(\mathbf{S}^W)$, and therefore $\mathrm{Tran}(\mathbf{S}^W)$ are isomorphic. The construction of A_α is analogous to the construction of A above, and will not be given here. If W is an arbitrary monoid, then I do not know whether T_α is equivalent to I iff α is the identity automorphism of W . As we will prove in the next chapter, the class of monoids for which M_W is distinguishable in $\mathbf{S}^W$, includes the free monoids.

CHAPTER III: TOWARDS THE DISTINGUISHABILITY OF M_W IN $\mathbf{S}^W$

Our main result in Chapter II, as expressed by Theorem 7.4, reduces a certain characterization of the categorical predicates in $\mathbf{S}^W$ to the distinguishability of M_W in $\mathbf{S}^W$. To be specific, we can be sure that the categorical classes in $\mathbf{S}^W$ are the natural classes which are closed under $\mathcal{A}$ (the group of the automorphisms of $\mathbf{S}^W$ induced by Aut(W)) only if M_W is distinguishable in $\mathbf{S}^W$. Our goal in this chapter is to establish conditions on W which are sufficient for the distinguishability of M_W in $\mathbf{S}^W$. Unfortunately, we can prove that M_W is distinguishable only for a limited class of monoids. The problem whether M_W is distinguishable in $\mathbf{S}^W$ for any input monoid is still open. However, since our study of $\mathbf{S}^W$ is motivated by automata theory, we note that our results guarantee the distinguishability of M_W for any type of input monoid which is encountered in automata theory.

Our search for the conditions for the distinguishability of M_W leads us to a further study of the category $\mathbf{S}^W$. In this study we employ several methods of the theory of abelian categories [8]. In spite of the poor structure of semimodules (e.g., in comparison with modules), we derive some properties of $\mathbf{S}^W$ which may motivate a thorough study of $\mathbf{S}^W$ with applications to the theory of monoids. For example, our characterization of the projective objects in $\mathbf{S}^W$ indicates a significant relationship between the structure of $\mathbf{S}^W$ and the structure of W.

Our strategy in the **pursuit** of the distinguishability of M_W in $\mathbf{S}^W$ is not systematic. While establishing the reconstruction scheme for $\mathbf{S}^W$, we have discovered several outstanding properties of M_W. We check each one of these properties to see whether they yield a characterization of M_W. Unfortunately only some of the known properties of M_W imply a characterization of M_W , and even this is true only for some monoids W . In particular we do not know how to make use of the following two remarkable properties of M_W . The first property is that M_W has an endomorphism monoid isomorphic to W . We made use of this property in the discussion of the reconstruction scheme for $\mathbb{S}^W$, but it is hard to see what can be said on the structure of a W-semimodule A with endomorphism monoid isomorphic to W . Every obvious attempt to relate this property to the distinguishability of M_W fails to extend the class of monoids W for which we know that M_W is distinguishable in $\mathbb{S}^W$.

The second property of M_W is related to the previous one but it deserves special attention. The definition of the representation functor

$$r^* : \mathbb{S}^W \longrightarrow \mathbb{S}^{W^*}$$

in Chapter II, Section 4, suggests the following construction. Let $\mathbb{C}$ be any category which satisfies the set-theoretic axiom: for any objects A and B of $\mathbb{C}$, $\mathbb{C}(A,B)$ is a set. For any object M of $\mathbb{C}$ we define a functor

$$r^*_M : \mathbb{C} \longrightarrow \mathbb{S}^{E(M)} ,$$

where $E(M)$ is the endomorphism monoid of M in $\mathbb{C}$. For any object A of $\mathbb{C}$, the $E(M)$-semimodule $r^*_M(A)$ is defined by

$$s(r^*_M(A)) = \mathbb{C}(M,A) \quad \text{and} \quad \lambda_{r^*_M(A)}(f,h) = fh$$

(for all $f \in \mathbb{C}(M,A)$ and $h \in E(M)$). Because of the associativity of the composition rule in $\mathbb{C}$, the functor $\mathbb{C}(M,-)$ determines homomorphisms of $E(M)$-modules. Thus for any morphism $g : A \longrightarrow B$, we define $r^*_M(g) : r^*_M(A) \longrightarrow r^*_M(B)$ to be the homomorphism determined by the function

$$[\mathbb{C}(M,-)](g) : \mathbb{C}(M,A) \longrightarrow \mathbb{C}(M,B) : f \longrightarrow gf .$$

Clearly, we have

$$S \circ r_M^* = \mathbb{C}(M,-) .$$

Now, the functor $r_M^* : \mathbb{C} \longrightarrow S^{E(M)}$ is an embedding iff M is a <u>generator</u> of $\mathbb{C}$; i.e., iff $\mathbb{C}(M,-)$ is an embedding. (To recall, note that a functor $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ is an embedding iff for any pair of objects A and B of $\mathbb{C}_1$, the function

$$F : \mathbb{C}_1(A,B) \longrightarrow \mathbb{C}_2(F(A),F(B)) : f \longrightarrow F(f)$$

is injective.) Thus we have that every category with a generator, and which satisfies the set-theoretic axiom, has an embedding functor into a category of semimodules. A natural question arises: when is $r_M^* : \mathbb{C} \longrightarrow \mathbb{S}^{E(M)}$ a full embedding? We recall that a functor $F : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ is full iff for any pair of objects A and B of $\mathbb{C}_1$, the function $F: \mathbb{C}_1(A,B) \longrightarrow \mathbb{C}_2(F(A),F(B)) : f \longrightarrow F(f)$ surjective. It is not difficult to verify that M_W has the property expressed by the fact that $r_M^* : \mathbb{S}^W \longrightarrow \mathbb{S}^{W^*}$, which is precisely $r_{M_W}^* : \mathbb{S}^W \longrightarrow \mathbb{S}^{E(M_W)}$ according to our recent notation, is a full embedding functor. That is, for every function

$$\phi : \mathbb{S}^W(M_W,A) \longrightarrow \mathbb{S}^W(M_W,B)$$

which has the property that for every $f_w : M_W \longrightarrow M_W$ and every $f_s : M_W \longrightarrow A$

$$\phi(f_s f_w) = \phi(f_s) f_w ,$$

there exists a unique morphism $f_\phi : A \longrightarrow B$ such that ϕ is the underlying function of $r^*(f_\phi) : r^*(A) \longrightarrow r^*(B)$. Simply define $f_\phi : S(A) \longrightarrow S(B)$

by $f_\phi(s) = [\phi(f_s)](1)$. Furthermore, the natural equivalence of the forgetful functor of $\mathbb{S}^W$ with $\mathbb{S}^W(M_W,-)$ implies that M_W is a generator. It is interesting to note that we can make use of the fact that M_W is a generator in order to derive simple conditions which are sufficient for the distinguishability of M_W in $\mathbb{S}^W$. But we do not know what to infer from the fullness of the functor $r^* : \mathbb{S}^W \longrightarrow \mathbb{S}^{W^*}$. Nor do we know how to characterize those categories $\mathbb{C}$ with a generator M for which the embedding $r^*_M : \mathbb{C} \longrightarrow \mathbb{S}^{E(M)}$ is full.

In this chapter we will discuss mainly three properties of M_W and show how they yield sufficient conditions for the distinguishability of M_W . In doing so we also provide characterizations of the free semimodules, the projective objects and the generators of $\mathbb{S}^W$. We also discuss briefly the structure of the endomorphism monoids of some projective objects in $\mathbb{S}^W$.

8. The monogenic semimodules

Let A be a W-semimodule. A subset T of $S(A)$ is said to generate A iff

$$S(A) = \{t \circ w : t \in T \text{ and } w \in W\} .$$

In particular, A is said to be monogenic iff there exists a singleton set which generates A . Thus A is monogenic iff there exists an $s_o \in S(A)$ such that

$$S(A) = \{s_o \circ w : w \in W\} .$$

It is easy to see that M_W is monogenic, for it is generated by $\{1\}$. In fact M_W is generated by $\{u\}$ iff there exists a $v \in W$ such that $uv = 1$. In order to realize that the class of monogenic W-semimodules is categorical in $\mathbb{S}^W$, we prove the following lemma.

LEMMA 8.1. A W-semimodule A is monogenic iff for any set $\{T_i\}$ of sets which determine subsystems of A, $\cup\{T_i\} = S(A)$ implies that $S(A) = T_j$; for some $T_j \in \{T_i\}$.

PROOF. If A is monogenic, say by $s_o \in S(A)$, and $\cup\{T_i\} = S(A)$, then $s_o \in T_j$; for some $T_j \in \{T_i\}$, and therefore $T_j = S(A)$. On the other hand, let $T_s = s\circ W = \{s\circ w : w \in W\}$, for all $s \in S(A)$. Then clearly $\{T_s\}$ is a set of sets which determine subsystems of A, and $\cup\{T_s\} = S(A)$. If for some $T_{s_o} \in \{T_s\}$ we have $T_{s_o} = S(A)$, then A is generated by $\{s_o\}$.

We can translate now the statements of Lemma 8.1 and its proof and get a categorical characterization of the class of monogenic W-semimodules in $\mathbb{S}^W$. Recall that a categorical class in $\mathbb{S}^W$ is any class which is closed under the automorphisms of $\mathbb{S}^W$.

PROPOSITION 8.2. The class of the monogenic W-semimodules is categorical in $\mathbb{S}^W$.

PROOF. We make use of the fact that the subsets of S(A) which determine subsystems of A are represented by monic morphisms into A. Furthermore, the union of a given set of such subsets represented by monic morphisms, is represented by a monic morphism with a universality property. That is, the union is represented by an initial object in an especially constructed category of the monic morphisms into A.

Given a set J of monic morphisms $j : A_j \longrightarrow A$ of $\mathbb{S}^W$, with a common range A, we define a category $\mathbb{J}$ whose objects are all the monic morphisms $b : B \longrightarrow A$ such that for any $j \in J$ there exists a morphism $b_j : A_j \longrightarrow B$ with $bb_j = j$ (and therefore b_j must be monic). There exists a morphism in $\mathbb{J}$ from $(b : B \longrightarrow A)$ to $(b' : B' \longrightarrow A)$ for each morphism $f : B \longrightarrow B'$ with $b'f = b$ (and hence f is monic). For any set J of monic morphisms of $\mathbb{S}^W$ with range A , the category $\mathcal{J}$ has an initial object U(J) . In fact, a monic morphism $b : B \longrightarrow A$ is initial in $\mathcal{J}$ iff the image semimodule b(B) is determined by the union of the images of the morphisms in $\mathcal{J}$.

Hence, an object A of $\mathbb{S}^W$ is monogenic iff for any set J of monic morphisms of $\mathbb{S}^W$ with range A , if U(J) is epic, then J contains an epic morphism. Since obviously monic and epic morphisms are preserved under automorphisms, if A is monogenic and F is an automorphism of $\mathbb{S}^W$, then F(A) is also monogenic. Hence the class of monogenic W-semimodules is categorical in $\mathbb{S}^W$.

We know that M_W is monogenic. Furthermore, it is easy to verify that A is monogenic iff there exists an epic morphism $e : M_W \longrightarrow A$. For, A is generated by {s} iff $f_s : M_W \longrightarrow A$ is epic. We find the first sufficient condition for the distinguishability of M_W by insisting that M_W be a "largest" monogenic object of $\mathbb{S}^W$, in the sense to be defined presently. An object M of $\mathbb{S}^W$ is said to be a <u>largest</u> object in a class K of objects of $\mathbb{S}^W$ iff for any object A in K , if $e : A \longrightarrow M$ is an epic morphism, then it is an isomorphism. We prove now that M_W is a largest monogenic object of $\mathbb{S}^W$ iff W is <u>unit-regular</u> in the sense that if uv = 1 in W then vu = 1 as well.

We say that w ε W is a <u>transform of 1</u> iff w = vu where uv = 1 in W . For any w ε W we denote by M(w) the subsystem of M_W which is determined by wW . We prove first the following lemma.

LEMMA 8.3. If w is a transform of 1 in W , then M(w) is isomorphic to M_W .

PROOF. Assume that w_1 = vu where uv = 1 . We define the morphism $e_u : M(w_1) \longrightarrow M_W$ by

$$e_u : w_1W \longrightarrow W : w_1w \longrightarrow uw_1w .$$

Since $uw_1w = uvuw = uw$, it follows that e_u is surjective: for any $w \in W$, $e_u(vw) = e_u(w_1vw) = w$. It is also injective since

$$ve_u(w_1w) = vuw_1w = vuw = w_1w .$$

By the associativity of the multiplication in W, it follows that e_u determines a homomorphism. In conclusion

$$e_u : M(w_1) \longrightarrow M_W$$

is an isomorphism.

Now we have

PROPOSITION 8.4. M_W is the largest monogenic object of $\mathbb{S}^W$ iff W is unit-regular.

PROOF. Assume that M_W is the largest monogenic object in $\mathbb{S}^W$, and assume that $uv = 1$. Since $M(vu)$ is isomorphic to M_W, it is also a largest monogenic object in $\mathbb{S}^W$.

We define $e_{vu} : M_W \longrightarrow M(vu)$ by

$$e_{vu} : W \longrightarrow vuW : w \longrightarrow vuw .$$

Clearly, $e_{vu} : M_W \longrightarrow M(vu)$ is an epic morphism. But $M(vu)$ is a largest monogenic object in $\mathbb{S}^W$, therefore, e_{vu} must be monic. That is,

$$vuw_1 = vuw_2 \text{ implies } w_1 = w_2 .$$

Since $vuvu = vu1$, we must have $vu = 1$, and W is unit-regular.

On the other hand, assume that W is unit-regular. Let A be a monogenic W-semimodule generated by $\{s_b\}$, and let $e : A \longrightarrow M_W$ be an epic morphism. Since $e : S(A) \longrightarrow W$ is surjective, there exists a $v \in W$ such that $1 = e(s_o \cdot v) = e(s_o)v$, and since W is unit-regular, we have $ve(s_o) = 1$ as well. We prove now that e is injective. Assume that $e(s_o \cdot w_1) = e(s_o \cdot w_2)$, then we have

$$w_1 = ve(s_o)w_1 = ve(s_o \cdot w_1) = ve(s_o \cdot w_2) = ve(s_o) \cdot w_2 = w_2 ,$$

and in particular $s_o \cdot w_1 = s_o \cdot w_2$. Thus $e : A \longrightarrow M_W$ is an isomorphism. Hence, W is unit-regular iff M_W is the largest monogenic object in $\mathbb{S}^W$.

From this proposition and from Proposition 8.2 follows now

COROLLARY 8.5. If W is a unit-regular monoid, then M_W is distinguishable in $\mathbb{S}^W$.

9. The free semimodules

We rephrase a part of Lemma 4.1 (cf. Chapter II) into a form which proves that " M_W is free on 1 ". We know from Lemma 4.1 that, for any W-semimodule A and any $s \in S(A)$, there exists a unique morphism $f_s : M_W \longrightarrow A$ with $f_s(1) = s$. Thus, if $j : \{1\} \longrightarrow W$ is the obvious injection, then M_W has the following property. For any W-semimodule A and any function $f : \{1\} \longrightarrow S(A)$, there exists a unique morphism $f^* : M_W \longrightarrow A$ of $\mathbb{S}^W$ such that as functions we have $f^*j = f$ (namely, $f^* = f_{f(j)}$). The general notion of free semimodules is now evident.

DEFINITION 9.1. A W-semimodule A is said to be <u>free with respect to an injective function</u> $g_T : T \longrightarrow S(A)$ iff for any W-semimodule B and for any function $f : T \longrightarrow S(B)$ there exists a unique morphism $f^* : A \longrightarrow B$ of $\mathbb{S}^W$ with $f^*g_T = f$ in $\mathbb{S}$, the category of sets.

As we indicated above, M_W is free with respect to the inclusion function $j : \{1\} \longrightarrow W$. We can prove a more general result.

LEMMA 9.2. For any set T, the W-semimodule $T \circ M_W$ is free with respect to the injection

$$g_T : T \longrightarrow T \times W : t \longrightarrow (t,1) .$$

PROOF. Recall that $T \circ M_W$ is the sum object defined by

$$S(T \circ M_W) = T \times W \quad \text{and} \quad (t,w_1) \circ w_2 = (t,w_1 w_2) .$$

Let A be any W-semimodule and $f : T \longrightarrow S(A)$ be any function. We define

$$f^* : T \times W \longrightarrow S(A) : (t,w) \longrightarrow f(t) \circ w .$$

By its very definition, f^* determines a homomorphism $f^* : T \circ M_W \longrightarrow A$ of W-semimodules; and clearly we have $f^* g_T = f$.

Assume that for $f' : T \circ M_W \longrightarrow A$ we have $f' g_T = f$. Then for any $(t,w) \in T \times W$ we get

$$f'(t,w) = f'(t,1) \circ w = f' g_T(t) \circ w = f(t) \circ w = f^*(t,w) .$$

Hence $f' = f^*$, and so $T \circ M_W$ is free with respect to g_T.

The definition of the free semimodules suggests their characterization as initial objects in categories especially constructed from $\mathbb{S}^W$. The analogy with the free objects in $\mathbb{G}$, the category of groups, and in $\mathbb{M}$, in the category of monoids, is evident. Cohn [5; Chapter III] describes the general notion of free algebras in the same manner which is suggested by our discussions of the free objects in $\mathbb{G}$ and in $\mathbb{M}$, in Part I, Chapter I, Section 4 of this paper. In particular, we derive the following corollary of Lemma 9.2.

COROLLARY 9.3. If A is a free W-semimodule with respect to an injection $h_T : T \longrightarrow S(A)$, then A is isomorphic to $T \circ M_W$.

PROOF. We give here a direct proof which does not depend on the general characterization of the free algebras as initial objects somewhere. Note, however, that this proof reflects the proof to the effect that initial objects in categories are isomorphic (cf. Part I, Chapter I, Section 4).

Since A is free with respect to $h_T : T \longrightarrow S(A)$, there exists a unique morphism $g_T^* : A \longrightarrow T \circ M_W$ with $g_T^* h_T = g_T$, where g_T is the injection specified in Lemma 9.2. Similarly, we have a unique morphism $h_T^* : T \circ M_W \longrightarrow A$ with $h_T^* g_T = h_T$. Hence

$$(h_T^* g_T^*) h_T = h_T \quad \text{and} \quad (g_T^* h_T^*) g_T = g_T .$$

The morphism $h_T^* g_T^* : A \longrightarrow A$ must be the identity, for there exists a unique morphism $i^* : A \longrightarrow A$ for which $i^* h_T = h_T$ where $i = h_T$. Similarly, $g_T^* h_T^* : T \circ M_W \longrightarrow T \circ M_W$ is the identity morphism, and therefore A and $T \circ M_W$ are isomorphic.

This characterization of the free W-semimodules, is obviously not categorical in its appearance, for it is involved with functions into sets of states, and not only with morphisms. Yet, we will refer to it in the characterization of the projective objects in $\mathbb{S}^W$ to be presented in the next section.

10. The structure of the projective W-semimodules.

DEFINITION 10.1. An object P of a category $\mathbb{C}$ is called <u>projective</u> iff for any morphism $g : P \longrightarrow B$ and any epic morphism $e : A \longrightarrow B$ of $\mathbb{C}$, there exists a morphism $f : P \longrightarrow A$ of $\mathbb{C}$ such that $ef = g$.

As we will prove presently, in any category a sum of projective objects is always projective. In $\mathbb{S}^W$, we find that M_W is projective, and therefore every free W-semimodule, by Corollary 9.3, is projective in

$\mathbb{S}^W$. Moreover, we prove that in $\mathbb{S}^W$ every projective object is a sum of monogenic projective W-semimodules. Furthermore, a W-semimodule P is monogenic projective iff it is isomorphic to $M(u)$, the subsystem of M_W determined by uW, where u is an idempotent of W. It follows directly from Lemma 8.3, that if W is a monoid in which every idempotent is a transform of 1 (e.g., if 1 is the only idempotent of W), then M_W is distinguishable in $\mathbb{S}^W$ by being a monogenic projective object of $\mathbb{S}^W$.

LEMMA 10.2. If (in any category) P is a sum of projective objects then P is projective.

PROOF. Let $P = \Sigma\{P_s\}$ be the sum object of $\{P_s\}$, with the sum diagram $\{j_s : P_s \longrightarrow \Sigma\{P_s\}\}$. Let $e : A \longrightarrow B$ be any epic morphism and $g : \Sigma\{P_s\} \longrightarrow B$ be any morphism. If P_s is projective (for all s) then there exists a morphism $f_s : P_s \longrightarrow A$ with $ef_s = gj_s$. Let $f : \Sigma\{P_s\} \longrightarrow A$ be the morphism guaranteed by the sum diagram $\{j_s : P_s \longrightarrow \Sigma\{P_s\}$ and the family of morphisms $\{f_s : P_s \longrightarrow A\}$. That is, we have $fj_s = f_s$ (for all s). Then (for all s) we have $efj_s = ef_s = gj_s$. Hence $ef : \Sigma\{P_s\} \longrightarrow B$ is the <u>unique</u> morphism guaranteed by the sum diagram of $\Sigma\{P_s\}$ for the family $\{gj_s : P_s \longrightarrow B\}$. But $efj_s = gj_s$, hence $ef = g$. And so $\Sigma\{P_s\}$ is also projective.

One can prove directly from the fact that $T \circ M_W$ is free, that it must be projective. We give here a "categorical" proof to the same effect.

LEMMA 10.3. M_W is projective in $\mathbb{S}^W$.

PROOF. Let $e : A \longrightarrow B$ be an epic morphism of $\mathbb{S}^W$. We refer to the commutative diagram of the natural equivalence $\rho : S \longrightarrow \mathbb{S}^W(M_W, -)$ evaluated at $e : A \longrightarrow B$ (cf. Lemma 4.1). We have

$$[\mathbb{S}^W(M_W,-)](e) = \rho(B)e(\rho(A))^{-1} ,$$

and therefore $[\mathbb{S}^W(M_W,-)](e)$ is surjective, which is just a fancy way to say that for any morphism $g : M_W \longrightarrow B$ there exists a morphism $f : M_W \longrightarrow A$ with $ef = g$. Hence M_W is projective.

Note that the argument employed in the proof of Lemma 10.3 can be used to prove the following general statement. Let $\mathbb{C}$ be any category with a functor $S : \mathbb{C} \longrightarrow \mathbb{S}$ which maps epic morphisms on surjective functions. If P is an object of $\mathbb{C}$ such that $\mathbb{C}(P,-)$ is naturally equivalent to S, then P is projective in $\mathbb{C}$.

Now from Corollary 9.3, Lemma 10.3 and Lemma 10.2, follows the next lemma.

LEMMA 10.4. Every free W-semimodule is projective in $\mathbb{S}^W$.

Our main tool for studying the projective objects of $\mathbb{S}^W$ is borrowed from homological algebra, and is the notion of "splitting sequence" (cf. [13: p.68]). A sequence

$$A \xrightarrow{j} B \xrightarrow{e} A$$

of morphisms (in any category) is a <u>splitting sequence</u> iff ej is the identity morphism of A. In this case we say that <u>A splits through B</u>.

LEMMA 10.5. (In any category) P is projective iff P splits through some projective object Q.

PROOF. If P is projective, then P splits trivially through P. On the other hand, assume that Q is projective and that $P \xrightarrow{q} Q \xrightarrow{p} P$ is a splitting sequence. Let $e : A \longrightarrow B$ be any epic morphism and $g : P \longrightarrow B$ any morphism. Since Q is projective, there exists a morphism

$f' : Q \longrightarrow A$ with $ef' = gp$. Hence

$$ef'g = gpq = g ,$$

which shows that P is also projective.

As for the projective W-semimodules, we can be more specific and prove that P is a projective object of $\mathbb{S}^W$ iff P splits through some free W-semimodule. Note that in any category, if P is projective and $A \longrightarrow P$ is epic, then P splits through A . In $\mathbb{S}^W$ we have that the transition function λ_A of any W-semimodule determines an epic homomorphism

$$\lambda_A : S(A) \circ M_W \longrightarrow A$$

from the free W-semimodule $S(A) \circ M_W$ onto A . Hence we have

LEMMA 10.6. A W-semimodule P is projective iff P splits through some free W-semimodules. In particular, if P is projective, then there exists a splitting sequence

$$P \xrightarrow{j_P} S(P) \circ M_W \xrightarrow{\lambda_P} P$$

where the homomorphism λ_P is determined by the transition function of P .

By examining the splitting sequence

$$P \xrightarrow{j_P} S(P) \circ M_W \xrightarrow{\lambda_P} P ,$$

we derive now the following corollary.

COROLLARY 10.7. Every projective W-semimodule is a sum (i.e., disjoint union) of monogenic projective W-semimodules.

PROOF. The idea of the proof is the following. Clearly,

$$S(P)\cdot M_W = \Sigma\{\{s\}\cdot M_W : s \in S(P)\} ,$$

i.e., $S(P)\cdot M_W$ is a sum of monogenic projective objects. Hence the splitting sequence of P through $S(P)\cdot M_W$ decomposes as a disjoint union of splitting sequences of the form

$$P_s \xrightarrow{j_P^s} \{s\}\cdot M_W \xrightarrow{\lambda_P^s} P_s$$

and $P = \Sigma\{P_s\}$. Formally, we proceed as follows.

We define for each $s \in S(P)$, the W-semimodule P_s as the subsystem of P determined by

$$S(P_s) = \{s' \in S(P) : j_p(s') \in s\cdot W\} .$$

Since j_P is monic, we have $P = \Sigma\{P_s : s \in S(P)\}$ and

$$j_P = \Sigma\{j_P^s : P_s \longrightarrow \{s\}\cdot M_W\}$$

where j_P^s is determined by the restriction $j_P|S(P_s)$ of j_P to $S(P_s)$. The decomposition of λ_P as $\lambda_P = \Sigma\{\lambda_P^s : s \in S(P)\}$ is given by

$$\lambda_P^s : s\cdot W \longrightarrow S(P) : (s,w) \longrightarrow s\cdot w .$$

Clearly we have $\lambda_P^s j_P = \lambda_P^s j_P^s$, and

$$i_{P_s} = i_P|S(P_s) = \lambda_P^s j_P|S(P_s) = \lambda_P j_P^s = \lambda_P^s j_P^s ,$$

where i_X denotes the identity morphism of X . Hence, for all $s \in S(P)$, the image of λ_P^s is precisely P_s , and so P_s is monogenic and it splits through $\{s\}\cdot M_W$ by means of the splitting sequence

$$P_s \xrightarrow{j_P^s} \{s\}\cdot M_W \xrightarrow{\lambda_P^s} P_s .$$

Thus P is a sum of monogenic projective W-semimodules. Note that the empty W-semimodule is projective since it is the only free W-semimodule with respect to the empty function. In this case our arguments are vacuously valid; the empty semimodule is the sum of the empty family.

From Crorollary 10.7 we can infer that every summand of a projective object of $\mathbb{S}^W$ is projective. Assume that P is a projective W-semimodule and $P = A + B$ in $\mathbb{S}^W$. The decomposition of P as a sum of monogenic projective W-semimodules $P = \Sigma\{P_s\}$ implies that

$$A = \Sigma\{P_s : s \in S(A)\} \text{ and } B = \Sigma\{P_s : s \in S(B)\} .$$

Hence A and B are sums of projective W-semimodules and therefore, by Lemma 10.2, they are both projective. Another important implication of Corollary 10.7 is

COROLLARY 10.8. A W-semimodule P is monogenic projective iff it is isomorphic to a monogenic projective subsystem of M_W.

PROOF. We give here a proof independent of Corollary 10.7. If P is monogenic, then there exists an epic morphism $e : M_W \longrightarrow P$ (cf. Section 8). If in addition P is projective, then for the identity morphism $P \longrightarrow P$ we have a morphism $j : P \longrightarrow M_W$ of such that ej is the identity morphism of P. Hence

$$P \xrightarrow{j} M_W \xrightarrow{e} P$$

is a splitting sequence and j(P) is a subsystem of M_W isomorphic to P.

Thus, in order to characterize the projective objects of $\mathbb{S}^W$, it is sufficient to characterize the monogenic projective subsystems of M_W. We recall that we denote by M(u) the subsystem of M_W determined by uW. A first characterization of the monogenic projective subsystems of of M_W is given in the following proposition.

PROPOSITION 10.9. A monogenic subsystem $M(u)$ of M_W is projective iff there exists a $j(u) \in W$ with

(i) $uw_1 = uw_2$ iff $j(u)w_1 = j(u)w_2$,

and

(ii) $u = uj(u)$.

PROOF. Assume that $M(u)$ is projective. Then, for the epic morphism $e_u : M_W \longrightarrow M(u)$ which is determines by

$$e_u : W \longrightarrow uW : w \longrightarrow uw ,$$

we have a morphism $j : M(u) \longrightarrow M_W$ with $e_u j$ being the identity morphism of $M(u)$. Now, since j must be monic, we have (i), and since $e_u j$ is the identity of $M(u)$, we have

$$u = (e_u j)(u) = uj(u) .$$

On the other hand, assume that there exists a $j(u) \in W$ with properties (i) and (ii). The function

$$j : uW \longrightarrow j(u)W : uw \longrightarrow j(u)w$$

is well defined as it follows from (i). Obviously it determines a morphism

$$j : M(u) \longrightarrow M_W .$$

By (ii) we have that

$$M(u) \xrightarrow{j} M_W \xrightarrow{e} M(u) ,$$

where $e_u(w) = uw$, is a splitting sequence. Hence $M(u)$ is projective.

For any splitting sequence

$$M(u) \xrightarrow{j} M_W \xrightarrow{e} M(u)$$

we have that j is monic. Hence $M(u)$ is isomorphic to $M(j(u)) = j(M(u))$. If, in particular, we have $e(w) = uw$, then

$$j(u)j(u) = j(uj(u)) = j(e(j(u)) = j(u) \ .$$

That is, $j(u)$ is an idempotent of W , and $M(u)$ is isomorphic to $M(j(u))$. On the other hand if u is an idempotent of W , then Proposition 10.9 implies that $M(u)$ is projective. Note that properties (i) and (ii) are satisfied by $u = j(u)$ iff u is an idempotent of W . In conclusion, we have proved

PROPOSITION 10.10. A W-semimodule P is monogenic projective iff it is isomorphic to a subsystem $M(u)$ of M_W , where u is an idempotent of W .

An alternative proof of Proposition 10.10 can be derived from the isomorphism $\psi_o : W \longrightarrow W^*$ (cf. Corollary 4.2) and Corollary 10.8. Observe that if $P \xrightarrow{j} M_W \xrightarrow{e} P$ is a splitting sequence, then $je : M_W \longrightarrow M_W$ is an idempotent of W^* .

As for the characterization of M_W , we can modify now Proposition 8.4. Lemma 8.3 and the proof of Proposition 8.4 imply

PROPOSITION 10.11. M_W is the largest monogenic projective object of $\mathcal{S}^W$ iff W is unit-regular.

PROOF. Note that if $uv = 1$, then vu is an idempotnet of W . By Proposition 10.10, the proof of Proposition 8.4 yields the proof of Proposition 10.11.

For example, if 1 is the only idempotent of W , then W is unit-regular. Yet, M_W is distinguished in $\mathbb{S}^W$ by being a monogenic projective object of $\mathbb{S}^W$, as it is implied by Proposition 10.10 and Lemma 8.3.

11. The generators of $\mathbb{S}^W$

We recall from the introductory discussions of this chapter, the following definition.

DEFINITION 11.1. An object G of a category $\mathbb{C}$ is called a generator iff $\mathbb{C}(G,-) : \mathbb{C} \longrightarrow \mathbb{S}$ is an embedding.

Thus, G is a generator of $\mathbb{C}$ iff for any two morphisms $g_1, g_2 : A \longrightarrow B$ of $\mathbb{C}$, if the two functions

$$[\mathbb{C}(G,-)](g_1) : \mathbb{C}(G,A) \longrightarrow \mathbb{C}(G,B) : f \longrightarrow g_1 f ,$$

and

$$[\mathbb{C}(G,-)](g_2) : \mathbb{C}(G,A) \longrightarrow \mathbb{C}(G,B) : f \longrightarrow g_2 f$$

are identical, then $g_1 = g_2$.

LEMMA 11.2. M_W is a generator of $\mathbb{S}^W$.

PROOF. By Lemma 4.1, we have the natural equivalence $\rho : S \longrightarrow \mathbb{S}^W(M_W,-)$ and from Section 3, we know that S is an embedding. We prove now the general statement: if $T_1, T_2 : \mathbb{C}_1 \longrightarrow \mathbb{C}_2$ are two naturally equivalent functors and T_1 is an embedding, then T_2 is also an embedding.

So let $\rho : T_1 \longrightarrow T_2$ be a natural equivalence, and assume that for some morphisms $g_1, g_2 : A \longrightarrow B$ of $\mathbb{C}_1$, we have $T_2(g_1) = T_2(g_2)$. Then for the commutative diagram of ρ, evaluated once at g_1 and once at g_2 we get

$$\rho(B)T_1(g_1)(\rho(A))^{-1} = T_2(g_1) = T_2(g_2) = \rho(B)T_1(g_2)(\rho(A))^{-1} ,$$

and therefore $T_1(g_1) = T_1(g_2)$. But T_1 is an embedding, hence $g_1 = g_2$.

Hence $\mathbf{S}^W(M_W,-)$ is an embedding, and therefore M_W is a generator of $\mathbf{S}^W$.

In order to characterize the generators of $\mathbf{S}^W$, we follow here Freyd's discussion [8; pp.68-69] on the generators of abelian categories which admit sums. In particular, we find that like in categories of modules an object G is a generator of $\mathbf{S}^W$ iff M_W splits through G .

Let G be any object of $\mathbf{S}^W$. For any object A of $\mathbf{S}^W$, we denote by $(G,A)\circ G$ the W-semimodule $\mathbf{S}^W(G,A)\cdot G$ which can be defined as the sum of the family $\{G_f : G_f = G$ and $f \in \mathbf{S}^W(G,A)\}$. Explicitly, $(G,A)\circ G$ is defined by $S((G,A)\circ G) = \mathbf{S}^W(G,A) x S(G)$ and $(f,s)\circ w = (f,s\cdot w)$ for all $f \in \mathbf{S}^W(G,A)$ and $w \in W$.

Note that $\mathbf{S}^W(G,A)$ may be empty even if G and A are not empty. For any $f \in \mathbf{S}^W(G,A)$ we have the canonical inclusion

$$j_f : G \longrightarrow (G,A)\circ G$$

determined by $j_f(s) = (f,s)$. Since $(G,A)\circ G$ is a sum object, its sum diagram (i.e., the family $\{j_f : G \longrightarrow (G,A)\cdot G\}$), guarantees the existence of a morphism

$$\sigma_{(G,A)} : (G,A)\circ G \longrightarrow A ,$$

determined by the family $\mathbf{S}^W(G,A)$ of all the morphisms $f : G \longrightarrow A$. Hence we have for any morphism $f : G \longrightarrow A$ the equality

$$\sigma_{(G,A)} j_f = f .$$

We prove now an analogue of a proposition of Freyd's [8; Prop. 3.36].

PROPOSITION 11.3. (Freyd) A W-semimodule G is a generator of $\mathbb{S}^W$ iff the morphism

$$\sigma_{(G,A)} : (G,A)\cdot G \longrightarrow A$$

is epic for all objects A of $\mathbb{S}^W$.

PROOF. Let $f_1, f_2 : A \longrightarrow B$ be any two morphisms of $\mathbb{S}^W$. Assume that for all $f : G \longrightarrow A$ we have

$$f_1 f = f_1 \sigma_{(G,A)} j_f = f_2 \sigma_{(G,A)} j_f = f_2 f .$$

Then, for all $f : G \longrightarrow A$ and for all $s \in S(G)$, since $j_f(s) = (f,s)$, we have

$$[f_1 \sigma_{(G,A)}](f,s) = [f_2 \sigma_{(G,A)}](f,s) .$$

Hence, if for all $f : G \longrightarrow A$ we have $f_1 f = f_2 f$, then $f_1 \sigma_{(G,A)} = f_2 \sigma_{(G,A)}$. On the other hand, since $\sigma_{(G,A)} j_f = f$, the equality $f_1 \sigma_{(G,A)} = f_2 \sigma_{(G,A)}$ implies that for all $f : G \longrightarrow A$, $f_1 f = f_2 f$. Thus

$$f_1 \sigma_{(G,A)} = f_2 \sigma_{(G,A)} \quad \text{iff} \quad [\mathbb{S}^W(G,-)](f_1) = [\mathbb{S}^W(G,-)](f_2) ;$$

i.e., $\sigma_{(G,A)}$ is epic iff $\mathbb{S}^W(G,-)$ is an embedding, that is, iff G is a generator.

Proposition 11.3 implies

COROLLARY 11.4. A W-semimodule G is a generator of $\mathbb{S}^W$ iff M_W splits through G.

PROOF. We prove first that if G is a generator of $\mathbb{S}^W$, then there exists an epic morphism $e : G \longrightarrow M_W$. Since M_W is projective, it follows that M_W splits through G.

If G is a generator of $\mathbb{S}^W$ then

$$\sigma_{(G,M_W)} : (G,M_W)\cdot G \longrightarrow M_W$$

is epic. Hence there exists a state (e,s_o) of $(G,M_W)\cdot G$ (i.e., $e : G \longrightarrow M_W$ and $s_o \in S(G)$) such that

$$\sigma_{(G,M_W)}(e,s_o) = 1 .$$

Thus

$$e(s_o) = [\sigma_{(G,M_W)} j_e](s_o) = \sigma_{(G,M_W)}(e,s_o) = 1 ,$$

which proves that $e : G \longrightarrow M_W$ is epic.

On the other hand, assume that there exists an epic morphism $e : G \longrightarrow M_W$. We prove that for all W-semimodules A, $\sigma_{(G,A)}$ is epic. For the empty W-semimodule $\emptyset_W$, we have that

$$\sigma_{(G,\emptyset_W)} : (G,\emptyset_W)\cdot G \longrightarrow \emptyset_W$$

is epic since it is the identity morphism of $\emptyset_W$.

Let A be any non-empty W-semimodule. We want to show that for any $s \in S(A)$ there exists a $(f,s') \in S((G,A)\cdot G)$ such that

$$\sigma_{(G,A)}(f,s') = s .$$

Since $\sigma_{(G,A)}(f,s') = f(s')$, we want to find a morphism $f : G \longrightarrow A$ and $s' \in S(G)$ such that $f(s') = s$. So let $s' \in S(G)$ be the state of G for which $e(s') = 1$, and let $f_s : M_W \longrightarrow A$ (i.e., $f_s(1) = s$), then $f_s e : G \longrightarrow A$ is a morphism for which

$$[f_s e](s') = f_s(1) = s .$$

Hence $\sigma_{(G,A)}$ is epic and, therefore, G is a generator of S^W.

We can express now another characterization of the unit-regular monoids which is equivalent to Proposition 8.4 or Proposition 10.11, as follows.

PROPOSITION 11.5. The monoid W is unit-regular iff all the splitting sequences of monogenic generators or of monogenic projective objects in $\mathbb{S}^W$ are made of isomorphisms only.

We note in passing the "polarity" between the monogenic projective objects of $\mathbb{S}^W$ and the generators of $\mathbb{S}^W$. Ones are the objects of $\mathbb{S}^W$ which split through M_W and the others are the objects of $\mathbb{S}^W$ for which M_W splits through them. Moreover, every summand of a projective object in $\mathbb{S}^W$ is projective, and as one can easily prove, every sum of a generator (of any category) with any object is also a generator.

12. The endomorphism monoids of monogenic projective W-semimodules

We give here an explicit description of the endomorphism monoid $E(M(u))$ of the morphisms $M(u) \longrightarrow M(u)$, where u is an idempotent element of W and $M(u)$ is the subsystem of M_W determined by uW.

Let $h : M(u) \longrightarrow M(u)$ be any homomorphism and suppose $h(u) = uv$ for $v \in W$. Then v has the property that $uw_1 = uw_2$ implies $uvw_1 = uvw_2$ for all $w_1, w_2 \in W$; for $h(uw) = h(u)w = uvw$. Conversely, if v satisfies the requirement that $uw_1 = uw_2$ always implies $uvw_1 = uvw_2$, then the function

$$h_v : uW \longrightarrow uW : uw \longrightarrow uvw$$

determines a homomorphism $hv : M(u) \longrightarrow M(u)$. Clearly $h_{v_1} = h_{v_2}$ iff $uv_1 = uv_2$. So let us denote by $R(u)$ the set of all elements v of W such that $uw_1 = uw_2$ always implies $uvw_1 = uvw_2$. Assume now that u is an idempotent of W. Then it is easily verified that $R(u)$ is a submonoid of W, $uR(u)$ is a subsemigroup of W with u as an identity element, and that $v \longrightarrow uv$ establishes a surjective homomorphism of monoids.

$$\varepsilon_u : R(u) \longrightarrow uR(u) .$$

PROPOSITION 12.1. Let u be an idempotent of W. The function

$$r : uR(u) \longrightarrow S^W(M(u),M(u)) : uv \longrightarrow h_v ,$$

where h_v is determined by

$$h_v : uW \longrightarrow uW : uw \longrightarrow uvw ,$$

determines an isomorphism of monoids

$$r : uR(u) \longrightarrow E(M(u)) .$$

PROOF. By our previous observations we know that r is a bijection. Furthermore (and here we give a proof to the effect that $uR(u)$ is a monoid), for any $v_1,v_2 \in R(u)$, we have

$$uv_1uv_2 = uv_1v_2 .$$

For since $v_1 \in R(u)$ and $uv_2 = u(uv_2)$, we must have $uv_1v_2 = uv_1uv_2$. This yields $h_{v_1}h_{v_2} = h_{v_1v_2}$.

Thus the endomorphism monoid of a monogenic projective W-semimodule is isomorphic to a homomorphic image of a submonoid of W. If we force the monogenic projective W-semimodules to have endomorphism monoids isomorphic to W in a simple fashion, we end up again with unit-regular monoids. As examples of this phenomenon, consider the following lemmata.

LEMMA 12.2. If W does not have a proper submonoid V with a surjective homomorphism onto W, or if W does not have a proper submonoid isomorphic to W, then W is unit-regular.

PROOF. Assume that $v_1v_2 = 1$ holds in W. Then by Lemma 8.3, $M(v_2v_1)$ is isomorphic to M_W and v_2v_1 is an idempotent of W. Hence

$(v_2v_1)R(v_2v_1)$ is isomorphic to W. But W does not have a proper submonoid with a surjective homomorphism onto W, hence $R(v_2v_1) = W$. Thus $v_2v_1v = 1$ for some $v \in W$, and therefore $v_2v_1 = 1$.

COROLLARY 12.3. Every finite monoid is unit-regular.

REMARK. Note that one can prove directly (i.e., without reference to semimodules) that every finite monoid is unit-regular.

LEMMA 12.4. If for any idempotent u of W we have $uR(u) = uW$, then $R(u) = W$ and W is unit-regular.

PROOF. We prove first that the equality $uR(u) = uW$ is equivalent to $R(u) = W$. So let w be any element of W, and we want to show that if $uR(u) = uW$, then for any $w_1, w_2 \in W$, $uw_1 = uw_2$ implies $uww_1 = uww_2$. Now from $uw \in uR(u)$ it follows that $uw = uv$ for some $v \in R(u)$. Hence if $uw_1 = uw_2$, then $uvw_1 = uvw_2$; i.e., $uww_1 = uww_2$. Obviously, if $R(u) = W$ then $uR(u) = uW$.

We prove now that if for any idempotent u of W, $R(u) = W$, then W is unit-regular. So assume that $v_1v_2 = 1$ holds in W, then v_2v_1 is an idempotent of W, and therefore $R(v_2v_1) = W$. We prove now that $v_2 = v_2v_2v_1$ holds in W. From $R(v_2v_1) = W$ and $v_2v_1v_2v_1 = v_2v_1$ we infer

$$v_2v_2v_1 = v_2v_1(v_2)v_2v_1 = v_2v_1(v_2) = v_2 ,$$

because $v_1v_2 = 1$. But, because of the same reason, $v_2 = v_2v_2v_1$ implies

$$1 = v_1v_2 = v_1v_2v_2v_1 = v_2v_1 ,$$

and in conclusion W is unit-regular.

LEMMA 12.5. If for any idempotent u of W we have $uR(u) = W$, then 1 is the only idempotent of W, and W is, therefore, unit-regular.

PROOF. From $uR(u) = W$ it follows that $uv = 1$ for some $v \in W$. If, in addition, u is an idempotent, then

$$u = u1 = uuv = uv = 1 .$$

13. On the characterization of M_W as a monogenic projective generator of $\mathbf{S}^W$

Let us say that W satisfies the m.p.g. property iff M_W is distinguishable in $\mathbf{S}^W$ by being a monogenic projective generator in $\mathbf{S}^W$. In this section we derive a structural characterization of the monoids with the m.p.g. property. Unfortunately, there are monoids which do not satisfy the m.p.g. property, and the problem whether M_W is distinguishable in $\mathbf{S}^W$ for any monoid W, is open.

We know from Corollary 11.4, that a subsystem $M(u)$ of M_W is a generator of $\mathbf{S}^W$ iff there exists an epic morphism $e : M(u) \longrightarrow M_W$. Assume that $e(u) = v$, then we have

(i) $vW = W$, and

(ii) $uw_1 = uw_2$ always implies $vw_1 = vw_2$.

Conversely, if there exists a $v \in W$ with these two properties, then property (ii) insures that

$$e : uW \longrightarrow W : uw \longrightarrow vw$$

determines a morphism $e : M(u) \longrightarrow M_W$ which is epic by property (i).

Furthermore, $M(u)$ is isomorphic to M_W iff there exists a $v \in W$ with $vW = W$ and such that for any $w_1, w_2 \in W$, $uw_1 = uw_2$ is equivalent to $vw_1 = vw_2$.

We prove now

PROPOSITION 13.1. A W-semimodule is a monogenic projective generator of $\mathcal{S}^W$ iff it is isomorphic to a subsystem $M(u)$ of M_W with

(i) $u^2 = u$,

and

(ii) there exists a $v \in W$ such that $vW = W$, and $vu = v$.

PROOF. Let $M(u)$ be a subsystem of M_W with properties (i) and (ii). From (i) it follows that $M(u)$ is a monogenic projective object of $\mathcal{S}^W$. Define now the function

$$e : uW \longrightarrow W : uw \longrightarrow vw ,$$

where $v \in W$ is the element of W whose existence and properties are stated by (ii). Since $vu = v$, it follows that e is well defined and that it determines a morphism $e : M(u) \longrightarrow M_W$, which is by property (ii) (i.e., $vW = W$) an epic morphism. Hence $M(u)$ is also a generator of $\mathcal{S}^W$.

On the other hand, if G is a monogenic projective generator of $\mathcal{S}^W$, then it is isomorphic to a subsystem $M(u)$ of M_W , where u is an idempotent of W , and is such that there exists an epic morphism $e : M(u) \longrightarrow M_W$. Assume that $e(u) = v$, then $vW = W$, and since $uu = u$ we have

$$vu = e(u)u = e(uu) = e(u) = v .$$

From our observations before Proposition 13.1, and from Proposition 13.1, follows now

PROPOSITION 13.2. A monoid W satisfies the m.p.g. property (i.e., M_W is distinguishable in $\mathcal{S}^W$ by being a monogenic projective generator of $\mathcal{S}^W$) iff W satisfies the following condition:

for any idempotent u of W , if

1. there exists a $v \in W$ with $vW = W$ and $vu = v$,

then

2. there exists a $v' \in W$ with $v'W = W$ and $v'u = v'$ such that $v'w_1 = v'w_2$ always implies $uw_1 = uw_2$.

PROOF. The condition 1. is equivalent to $M(u)$ being a monogenic projective generator of S^W, and the consequence 2. is equivalent to $M(u)$ being also isomophic to M_W.

We prove now that the class of all monoids which satisfy the m.p.g. property properly includes the class of all unit-regular monoids, and that it is a proper subclass of all monoids.

PROPOSITION 13.3. If W is unit-regular, then W satisfies the m.p.g. property.

PROOF. Let u be an idempotent of W, and $v \in W$ such that $vW = W$ and $vu = v$. From $vW = W$ it follows that $uv = 1$ for some $w \in W$, and since W is unit-regular, we have $wv = 1$ as well. Assume that $vw_1 = vw_2$ holds for some $w_1, w_2 \in W$. Then we have

$$uw_1 = uwvw_1 = uwvw_2 = uw_2 .$$

Hence $vw_1 = vw_2$ always implies $uw_1 = uw_2$. Hence by Proposition 13.1 and the discussions preceding it, every monogenic projective generator of S^W is isomorphic to M_W.

On the other hand we have

LEMMA 13.4. If every idempotent of W is a transform of 1, then W satisfies the m.p.g. property.

PROOF. By Lemma 8.3 we have that if every idempotent of W is a transform of 1, then every monogenic projective object of S^W is isomorphic to M_W.

For example, let W be the monoid generated by the two elements a and b with the single relation $ab = 1$. Then the elements of W are all of the form $b^{k_1}a^{k_2}$ where $k_1, k_2 \geq 0$. Simple arithmetic shows that the idempotents of W are all of the form $b^k a^k$, and since $a^k \circ b^k = 1$ holds in W, all the idempotents of W are transforms of 1. Clearly W is not unit-regular, for $ba \neq 1$. In conclusion, we have that the class of all monoids with the m.p.g. property properly includes the class of unit-regular monoids. Another example, suggested by A. Brumer, shows that there are monoids which do not satisfy the m.p.g. property.

Let W be the monoid with u, v, and w as generators, and with the following relations which insure by Proposition 13.1 that $M(u)$ is a monogenic projective generator in $\mathbb{S}^W$:

$$u^2 = u \ , \ vw = 1 \ , \ vu = v \ .$$

Combinatorial arguments show that $x \cdot y = 1$ holds in W, for some $y \in W$, iff $x = v^k$ for some $k \geq 0$; and furthermore, if $x \circ y = 1$ holds in W, then $x \circ w \circ v = x \circ 1$ holds as well. But $wv \neq 1$ in W, and therefore, by our observations that precede Proposition 13.1, we infer that $M(u)$ is not isomorphic to M_W in $\mathbb{S}^W$. Hence W does not satisfy the m.p.g. property. We do not know, for this particular W, whether M_W is distinguishable in $\mathbb{S}^W$ at all.

We conclude this section, and this part of the paper, with a review of some well known monoids which are unit-regular. To begin with, we note without a proof that the class of unit-regular monoids, and also the class of monoids whose idempotents are all transforms of 1, are closed under direct and free products of monoids (i.e., under the products and sums in $\mathbb{M}$, the category of monoids). Now, the cancellative monoids (of either side) are obviously monoids with 1 as their only idempotent, and therefore, they are unit-regular. Thus the groups and the free monoids are unit-regular. Needless to say, the abelian monoids are unit-regular, and we know by Corollary 12.3 that all the finite monoids are unit-regular.

PART III

APPLICATIONS TO AUTOMATA THEORY

INTRODUCTION TO PART III

Let us summarize those of our results whose relevance to automata theory will be discussed in this part of the work. If W is a monoid which satisfies the m.p.g. property (e.g., if W is unit-regular), then $\mathrm{Aut}(\mathbb{S}^W)$ is isomorphic to $\mathrm{Aut}(W)$ and to the group of the automorphisms of $\mathbb{S}^W$ which are induced by the automorphisms of W. We recall (cf. Part II, Chapter II, Section 6) that the automorphism of $\mathbb{S}^W$ induced by $\alpha \in \mathrm{Aut}(W)$ is the automorphism

$$T_\alpha : \mathbb{S}^W \longrightarrow \mathbb{S}^W$$

defined as follows. For any W-semimodule A, the W-semimodule $T_\alpha(A)$ is defined by

$$S(T_\alpha(A)) = S(A) \quad \text{and} \quad \lambda_{T_\alpha(A)}(s,w) = \lambda_A(s,\alpha(w)) \ .$$

For any morphism $f : A \longrightarrow B$ of $\mathbb{S}^W$, the morphism $T_\alpha(f) : T_\alpha(A) \longrightarrow T_\alpha(B)$ has the same underlying function as $f : A \longrightarrow B$. The particular isomorphism that exists between $\mathrm{Aut}(\mathbb{S}^W)$ and the group $\mathcal{A}$ of these automorphisms of $\mathbb{S}^W$ induced by $\mathrm{Aut}(W)$, whenever M_W is distinguishable in $\mathbb{S}^W$, implies the following characterization of the categorical classes in $\mathbb{S}^W$ (cf. Part II, Chapter II, Theorem 7.4). A natural class of morphisms of $\mathbb{S}^W$ is categorical iff it is closed under $\mathcal{A}$. In particular, a typital class of W-semimodules (i.e., a property of isomorphism types in $\mathbb{S}^W$) is categorical iff it is closed under $\mathcal{A}$.

In this part of this work, we examine the significance of these results to actual automata theory. First we note that most of the monoids encountered in automata theory, either as input monoids or otherwise, are unit-regular. For example, in the theory of finite automata [2, 4, 6, 9, 10, 14, 15, 16, 17, 19] mostly free monoids are considered as input monoids. Sometimes direct products of free monoids are considered, for example in the study of multi-tape automata [6], and even cancellative monoids, as in Ginsburg's [9] theory of abstract machines. Thus, it is meaningful to wonder about the significance of our study of S^W to actual automata theory.

A support for the significance of our results should consist at least in the examination of the following problems:

1. In the course of study in automata theory, several specific properties of automata are defined and studied. We will refer to these properties as the actual properties of automata. We will be concerned with the following double edged problem. Are the actual properties of automata categorical? Are there, or could there be, any property of W-semimodules which is, or may be, significant for the development of automata theory, and yet not categorical? We will discuss these problems in Chapter I.

2. Categorical algebra is not only a convenient language (in the broad sense of this term) for expressing certain ideas. While there is quite a common agreement on the fact that the notions of category and functor provide a convenient language for many branches of algebra, the realization that categorical algebra is also a useful tool of mathematical study is still debateful. It is interesting to note that certain authors are quite ambivalent about this point. For example, MacLane writes "The notions of category and functor provide not profound theorems, but a convenient language." [14; p.33]. Also, "By now it is expected that each definition of a type of mathematical system be accompanied by a definition of the morphisms of this system." [14; p.34], and later, in the same page, while refering to categories and functors, he writes "They have proved useful in the formulation of axiomatic homology, etc."

Such an ambivalence cannot be but expected from one of the fathers of categorical algebra. As for the usefulness of the categorical study of S^W to automata theory, it should be clear that the main goal of this particular work is not to prove that categorical algebra methods or "categorical thinking" are useful in automata theory. We wanted to determine what properties of W-semimodules can be determined by the algebraic properties of the composition of the homomorphisms of W-semimodules. Still, we would like to indicate the advantage of using categorical methods in the development of automata theory. Only experience and the future developments of automata theory can judge whether categorical algebra methods will be useful in automata theory as they have been proven to be so in many branches of mathematics (cf. [13; p.102]). In Chapter II we give an example of a categorical analysis of some basic properties of strongly connected and abelian semimodules which are due to Fleck [7] and Weeg [18]. It is the author's belief that categorical algebra methods applied to the study of automata are the known best methods suitable for the understanding of the mathematical ideas that take place in automata theory.

CHAPTER I: ACTUAL AUTOMATA THEORETIC PROPERTIES OF SEMIMODULES

During the development of automata theory, various properties were introduced because of their interest or their usefulness for the study of automata. Here, we will examine four types of such properties, and indicate the proof that they are all categorical in the categories of the form $\mathbb{S}^W$, provided that M_W is distinguishable in $\mathbb{S}^W$. From now on we assume tacitly (unless stated otherwise) that we are concerned only with input monoids W for which M_W is distinguishable in $\mathbb{S}^W$.

From Theorem 7.4 of Part I we know that a class K of W-semimodules is categorical iff

(i) K is closed under the automorphisms of $\mathbb{S}^W$ which are naturally equivalent to the identity functor of $\mathbb{S}^W$, and

(ii) K is closed under $\mathcal{A}$, the automorphisms of $\mathbb{S}^W$ induced by Aut(W).

Hence, in order to prove that a given class K of W-semimodules is categorical, it is sufficient to prove two properties of K. By proving first that K is closed under the isomorphisms inside $\mathbb{S}^W$ (i.e., that K is a typical class in $\mathbb{S}^W$), we prove that K is natural. Any automorphism which is naturally equivalent to the identity functor of maps every object onto an isomorphic object. If in addition to this we prove that K is closed under $\mathcal{A}$, we can infer that K is categorical. In what follows, in proving that any given class of W-semimodules is categorical, we indicate, therefore, only that K has these two properties.

1. Graph theoretic properties of semimodules.

The category $\mathbb{DG}$ of directed graphs is defined as follows. An object G of $\mathbb{DG}$ is a system of the form

$$G = \langle N(G), Ad(G) \rangle$$

where $N(G)$ is a set, the set of <u>nodes</u> of G, and $Ad(G)$ is any subset of $N(G) \times N(G)$, the set of <u>edges</u> of G. The morphisms of $\mathbb{D}G$ are of the form

$$f : G_1 \longrightarrow G_2$$

where $f : N(G_1) \longrightarrow N(G_2)$ is a function which "maps $A(G_1)$ into $A(G_2)$" in the sense that for any $(u_1, u_2) \in A(G_1)$ we have $(f(u_1), f(u_2)) \in A(G_2)$. The composition rule of $\mathbb{D}G$ is evident, and we have the obvious forgetful functor $N : \mathbb{D}G \longrightarrow \mathbb{S}$.

We define a functor

$$UG : \mathbb{S}^W \longrightarrow \mathbb{D}G$$

which maps W-semimodules onto their "underlying graphs", as follows. For any W-semimodule A, the directed graph $UG(A)$ is defined by

$$N(UG(A)) = S(A) \text{ , and } (s_1, s_2) \in Ad(UG(A)) \text{ iff } s_2 = s_1 \circ w \text{ for some } w \in W .$$

Since every function $f : S(A) \longrightarrow S(B)$ which determines a homomorphism $f : A \longrightarrow B$ of W-semimodules determines a morphism of $\mathbb{D}G$ as well, we define $UG(f) : UG(A) \longrightarrow UG(B)$ to be precisely the morphism of $\mathbb{D}G$ determined by the underlying function of $f : A \longrightarrow B$.

We indicate now the proof of

PROPOSITION 1.1. For any typical class K of directed graphs, the class $UG^{-1}(K)$ of all W-semimodules A with $UG(A)$ in K is a categorical typical class in $\mathbb{S}^W$.

PROOF. Since for any isomorphism j of W-semimodules, $UG(j)$ is an isomorphism in $\mathbb{D}G$, we have that the class $UG^{-1}(K)$ is natural in $\mathbb{S}^W$.

Since the automorphisms of $\mathbb{S}^W$ induced by Aut(W) obviously do not affect the "underlying graphs" of semimodules (that is, we have the equality $UG \circ T_\alpha = UG$ for all $\alpha \in$ Aut(W)), the class $UG^{-1}(K)$ is also closed under $\mathcal{A}$. Hence it is categorical typical in $\mathbb{S}^W$.

Examples of actual automata theoretic properties which are covered by this proposition are the following.

1. Set theoretic properties. Every typical class of K of sets in $\mathbb{S}$ (i.e., properties of sets which depend only on cardinalities) determines a typical class of directed graphs with set of nodes belonging to K . Hence every property of W-semimodules which depends only on the cardinality of their sets of states (e.g., infinite, finite, etc.) are categorical in $\mathbb{S}^W$.

2. Connected semimodules. A semimodule A is called connected iff UG(A) is a connected graph. A directed graph G is called connected iff the transitive closure of the symmetric reflexive closure of the binary relation Ad(G) is the total relation N(G)xN(G) on N(G) . Clearly, the class of all connected directed graph is typical in $\mathbb{DG}$. Hence the class of all connected W-semimodules is categorical typical in $\mathbb{S}^W$.

3. Strongly connected semimodules. A W-semimodule A is called strongly connected iff for any $s_1, s_2 \in S(A)$, there exists a $w \in W$ such that $s_2 = s_1 \circ w$. Clearly, a semimodule A is strongly connected iff UG(A) is strongly connected, that is, iff Ad(UG(A)) is the total binary relation on N(UG(A)) . Now, since the class of all strongly connected directed graphs (i.e., all complete directed graphs) is obviously typical in $\mathbb{DG}$, it follows that the class of all strongly connected W-semimodules is categorical typical in $\mathbb{S}^W$.

4. Semimodules with a reset state. A W-semimodule A is said to

have a reset state iff there exists an $s_o \in S(A)$ such that for any $s \in S(A)$ there exists a $w \in W$, possibly depending on s, such that $s \circ w = s_o$. Clearly, A has a reset state iff $UG(A)$ has a node s_o such that for any node s of $UG(A)$, $(s,s_o) \in Ad(UG(A))$. By analogy, we say in this case that s_o is a reset node of $UG(A)$. If s_o is a reset node of a directed graph G_1 and $j : G_1 \longrightarrow G_2$ is an isomorphism in $\mathbb{D}G$, then $j(s_o)$ is obviously a reset node of G_2. Hence the class of all directed graphs with reset nodes is typical in $\mathbb{D}G$, and, therefore, the class of all W-semimodules with reset states is categorical typical in $\mathbb{S}^W$.

2. Monoid theoretic properties of semimodules

Recent developments in automata theory (e.g. [12, 19]) bring to light the importance of the transition monoid associated with a finite automaton. For any W-semimodule A, the transition monoid $Mon(A)$ of A, is defined as the monoid of all the functions of the form

$$\tau_w^A : S(A) \longrightarrow S(A) : s \longrightarrow s \circ w ,$$

where $\tau_{w_1}^A = \tau_{w_2}^A$ iff for all $s \in S(A)$, $s \circ w_1 = s \circ w_2$. Here we use the "right-wise" notation of functions and denote by $(s)\tau_w^A$ the image $s \circ w$ of s under τ_w^A. Clearly we have

$$(s)\tau_{w_1}^A \tau_{w_2}^A = (s)\tau_{w_1 w_2}^A ,$$

and therefore, every W-semimodule A determines a surjective homomorphism of monoids

$$\tau^A : W \longrightarrow Mon(A)$$

which is determined by $\tau^A(w) = \tau_w^A$. Furthermore, for every epic morphism

$e : A \longrightarrow B$ of $\mathbb{S}^W$, there exists a unique surjective homomorphism of monoids

$$\mathrm{Mon}(e) : \mathrm{Mon}(A) \longrightarrow \mathrm{Mon}(B)$$

with $\tau^B = \mathrm{Mon}(e) \circ \tau^A$. The homomorphism Mon(e) is determined by $[\mathrm{Mon}(e)](\tau_w^A) = \tau_w^B$, and it is an isomorphism of monoids whenever $e : A \longrightarrow B$ is an isomorphism in $\mathbb{S}^W$. (Note that for any two epic morphisms $e_1, e_2 : A \longrightarrow B$, we have $\mathrm{Mon}(e_1) = \mathrm{Mon}(e_2)$.) Since also for any $\alpha \in \mathrm{Aut}(W)$ the function

$$\tau^\alpha : \mathrm{Mon}(A) \longrightarrow \mathrm{Mon}(T_\alpha(A)) : \tau_w^A \longrightarrow \tau_{\alpha(w)}^{T_\alpha(A)}$$

determines an isomorphicm of monoids

$$\tau^\alpha : \mathrm{Mon}(A) \longrightarrow \mathrm{Mon}(T_\alpha(A)) ,$$

we infer the following expected proposition.

PROPOSITION 2.1. If K is a typical class of monoids in $\mathbb{M}$, the category of monoids, then the class of all W-semimodules A with Mon(A) belonging to K is categorical typical in $\mathbb{S}^W$.

A very important class of actual properties of automata is covered by a corollary to Proposition 2.1. Let K be a typical class of groups in $\mathbb{G}$, the category of groups. Denote by $\overline{K}$ the class of all monoids M with the following property:

if G is a subsemigroup of M which is a group, then G belongs to K. Since every subsemigroup of M is mapped by an isomorphism j of M onto a subsemigroup of the range of j, and every image of a group under a homomorphism of semigroups is also a group, we infer that $\overline{K}$ is typical in $\mathbb{M}$. Hence we have

COROLLARY 2.2. Let K be any typical class of groups in G . The class of all W-semimodules A with Mon(A) belonging to $\bar{K}$, where $\bar{K}$ is the class of monoids with group-subsemigroups in K , is categorical typical in S^W .

For the significance of this classification of the W-semimodules, the reader is referred to the literature (e.g., [12, 19]). Other properties of semimodules which are covered by Proposition 2.1 are the following (cf. [7, 18]).

1. Abelian semimodules. A semimodule A is called abelian iff Mon(A) is an abelian monoid. Since, obviously, the class of all abelian monoids is typical in M , the class of all W-semimodules with abelian transition monoids is categorical typical in S^W .

2. Semimodules with a group action. A semimodule A is said to have a group-action iff Mon(A) is a group. Obviously, the class of all W-semimodules with a group action is categorical typical in S^W .

3. Analytical constructions in S^W

The problems of composition and decomposition of automata, and hence the problem of construction of automata from components, seem to be essential problems of automata theory, even if it is only recently that they have been seriously attacked. Unlike in group theory or other areas of algebra, these problems are not raised by the desire to understand the objects under study, but by the concrete interpretations and applications of the theory; i.e., by machine design and construction.

The semimodules are too poor in structure for studying composition of automata which are strongly involved with outputs. However, they are rich enough for the study of analytical construction to be defined presently. Certain recent results in automata theory [10, 19] indicate that even decomposability of automata with output can be reduced in some significant cases to properties of the semimodules involved. We define analytical constructions as multifunctors on $\mathbb{S}^W$ (i.e., on some power $(\mathbb{S}^W)^\xi$ of $\mathbb{S}^W$ where ξ is any ordinal) which do not depend very much on the particular labeling of the input monoid W. More precisely

DEFINITION 3.1. By an <u>analytical construction</u> on $\mathbb{S}^W$ we mean any functor

$$T : (\mathbb{S}^W)^\xi \longrightarrow \mathbb{S}^W ,$$

where $(\mathbb{S}^W)^\xi$ is the ξ-th power category of $\mathbb{S}^W$ and ξ is any ordinal, such that for any $\alpha \in \mathrm{Aut}(W)$ there exists a natural equivalence

$$\tau_\alpha : T_\alpha \circ T \longrightarrow T \circ T_\alpha^\xi ,$$

where

$$T_\alpha^\xi : (\mathbb{S}^W)^\xi \longrightarrow (\mathbb{S}^W)^\xi$$

is the obvious power of T_α : $T_\alpha^\xi(x,y,\ldots) = (T_\alpha(x),T_\alpha(y),\ldots)$. If ξ is finite, then we say that T is a finitary analytical construction.

It is easy to verify that functors derived from (finitary) analytical constructions by means of substitutions are also (finitary) analytical constructions. This follows directly from the properties of $\mathbb{N}$at (cf. Part I, Chapter I, Section 6). The properties of $\mathbb{N}$at imply also that any functor naturally equivalent to an (finite) analytical construction is also an (finite) analytical construction. For example, products and sums in $\mathbb{S}^W$ yield multifunctors which are analytical constructions. The subset construction employed in [16] in the reduction of non-deterministic automata to deterministic automata, when it is applied to $\mathbb{S}^W$, yields a functor

$$P : \mathbb{S}^W \longrightarrow \mathbb{S}^W$$

which is a finitary analytical construction. The explicit definition of the subset construction functor P is as follows. For any W-semimodule A, the W-semimodule P(A) is defined by

$$S(P(A)) = \{T : T \subseteq S(A)\} \text{ and } T \circ W = \{s \circ w : s \in T\} .$$

For any morphism $f : A \longrightarrow B$, the morphism

$$P(f) : P(A) \longrightarrow P(B)$$

is the homomorphism determined by the obvious function

$$P(f) : S(P(A)) \longrightarrow S(P(B)) : T \longrightarrow \{f(s) : s \in T\} .$$

Every set T determines two functors

$$\Pi^T : \mathbb{S}^W \longrightarrow \mathbb{S}^W \text{ and } \Sigma^T : \mathbb{S}^W \longrightarrow \mathbb{S}^W$$

which map every W-semimodule A onto A^T and $T \circ A$ respectively (cf. Part II, Chapter I, Section 2). Let us call these functors the T-power and the T-multiple functors. Clearly both these functors are analytical constructions for any set T. Thus every substitution combination of sums, products, power and multiple constructions and subset construction functors yields only analytical constructions on $\mathbb{S}^W$.

In order to determine whether certain problems related to analytical constructions give rise to categorical predicates, we prove the following lemma.

LEMMA 3.2. If $T : (\mathbb{S}^W)^\xi \longrightarrow \mathbb{S}^W$ is an analytical construction on $\mathbb{S}^W$, then for every automorphism F of $\mathbb{S}^W$, the functors $F \circ T$ and $T \circ F^\xi$ are naturally equivalent.

PROOF. By our tacit assumption on W, F is naturally equivalent to T_α for some $\alpha \in \mathrm{Aut}(W)$. Hence $F\circ T$ is naturally equivalent to $T_\alpha \circ T$, and $T\circ F^\xi$ is naturally equivalent to $T\circ T_\alpha^\xi$. Since T is analytical, $F\circ T$ and $T\circ F^\xi$ are naturally equivalent.

COROLLARY 3.3. Let $T : (\mathbb{S}^W)^\xi \longrightarrow \mathbb{S}^W$ be any analytical construction, and K be any categorical typical class in $\mathbb{S}^W$. Then the class of all ξ-tuples $(A_1, A_2, \ldots)$ of W-semimodules such that $T(A_1, A_2, \ldots)$ belongs to K is categorical typical in $\mathbb{S}^W$.

The examples of analytical constructions on $\mathbb{S}^W$ discussed above have the property that for any $\alpha \in \mathrm{Aut}(W)$, $T_\alpha \circ T = T\circ T_\alpha^\xi$. Let us call such constructions regular. The regular constructions have the following property.

LEMMA 3.4. Let $T : (\mathbb{S}^W)^\xi \longrightarrow \mathbb{S}^W$ be a regular construction. If $F : \mathbb{S}^W \longrightarrow \mathbb{S}^W$ is a functor with a natural equivalence $\tau : F \longrightarrow T_\alpha$, for some $\alpha \in \mathrm{Aut}(W)$, such that

$$T(\tau^\xi) = \tau\circ T ,$$

then $F\circ T = T\circ F^\xi$.

PROOF. For any ξ-tuple $(A_1, A_2, \ldots)$ of W-semimodules, the family

$$T(\tau^{-1}(A_1), \tau^{-2}(A_2), \ldots)\circ\tau(T(A_1), T(A_2), \ldots) : [F\circ T[(A_1, A_2, \ldots) \to T(F(A_1), F(A_2), \ldots)$$

is the identity natural transformation.

COROLLARY 3.5. Let $T : (\mathbb{S}^W)^\xi \longrightarrow \mathbb{S}^W$ be a regular construction on $\mathbb{S}^W$ such that for any automorphism F of $\mathbb{S}^W$ there exists a natural equivalence $\tau : F \longrightarrow T_\alpha$ for some $\alpha \in \mathrm{Aut}(W)$ with $T(\tau^\xi) = \tau\circ T$. Then the class of all $\xi+1$ tuples $(A_1, A_2, \ldots, T(A_1, A_2, \ldots))$ is a categorical predicate in $\mathbb{S}^W$.

While it is easily verified that all the specific analytical constructions discussed above are regular, it is not known whether they satisfy the hypothesis of Corollary 3.5.

4. Event properties of semimodules

Our terminology here is adapted from the study of finite automata as recognition devices (e.g., [2, 4, 6, 15, 16, 17]). These automata are finite semimodules over finitely generated free monoids provided with a specification of initial and final states. Even though we avoid assigning designated states to semimodules, we can still reduce some properties of automata viewed as recognition systems, to properties of semimodules. A similar procedure can reduce some properties of sequential machines [4, 9, 12] to properties of semimodules. In this section we will briefly discuss some properties of semimodules which have some bearing on their potentialities as components of recognition devices.

Generally and vaguely speaking, the event-properties of W-semimodules are those properties which are involved with the manner by which W is acting on the states. Hence, unless they are involved with properties of the actions of W which are invariant under the automorphisms of W, they would not be categorical.

For example, let w be any element of W which is not kept fixed under all the automorphisms of W. Then the class of all W-semimodules A for which there exists an $s_o \in S(A)$ such that $s \circ w = s_o$ for all $s \in S(A)$ is obviously not categorical. On the other hand, let E be a subset of elements of W which is closed under all the automorphisms of W (e.g., E = W). Then the class of all W-semimodules A for which there exists an $s_o \in S(A)$ and $w_o \in E$ such that for all $s \in S(A)$ we have $s \circ w_o = s_o$ (i.e., w_o is a reset input of A) is categorical in $\$^W$.

We concentrate now on properties of W-semimodules regarded as potential recognition systems for subsets of W. A W-semimodule A is said to potentially define the event $E \subseteq W$ iff there are $s_o \in S(A)$ and $F \subseteq S(A)$ such that

$$E = \{w : s_o \circ w \in F\} .$$

A family K of subsets of W is said to be symmetrical iff for any $E \in K$ and any $\alpha \in Aut(W)$, $\alpha(E) \in K$.

PROPOSITION 4.1. Let K be a symmetrical family of subsets of W. Then the class of all W-semimodules which potentially define events belonging to K is categorical typical in $\$^W$. Similarly, the class of all W-semimodules which potentially define only events of K is also categorical typical in $\$^W$.

PROOF. Isomorphic semimodules behave identically with respect to potential definition of events. That is, if $j : A \longrightarrow B$ is an isomorphism in $\$^W$, $s_o \in S(A)$ and $F \in S(A)$, then

$$\{w : \lambda_A(s_o,w) \in F\} = \{w : \lambda_B(j(s_o),w) \in j(F)\} .$$

If $\alpha \in Aut(W)$, then we have

$$\alpha(\{w : \lambda_A(s_o,w) \in F\}) = \{w : \lambda_{T_\alpha(A)}(s_o,w) \in F\} .$$

The following are some examples of properties of W-semimodules which are related to actual event properties of finite automata and which fall under Proposition 4.1.

1. Ideals in W. A subset E of W is said to be a left ideal in W iff for any $w_1 \varepsilon W$ and $w_2 \varepsilon E$ we have $w_1w_2 \varepsilon E$. Right ideals are defined similarly. A W-semimodule is said to be potentially definite iff it potentially defines a finitely generated right ideal in W. Since the family of right ideals in W is obviously symmetrical, the class of potentially definite W-semimodules is categorical typical in $\$^W$.
Similarly, potentially co-definite semimodules are those which potentially define left-ideals, and the class of potentially co-definite W-semimodules is categorical typical in $\$^W$.

2. Infinite and finite recognitions. A W-semimodule A is said to have a potentially finite recognition (respectively, infinite recognition) iff it potentially defines a finite (respectively, an infinite) subset of W.

For some studies of finite automata, with respect to ideal recognitions and finite recognitions, the reader is referred to the literature [15, 16]. Out last example is related to the notion of star-height of regular events introduced by Eggan [6].

3. Potential star-height and star-height of finite semimodules over a finitely generated free monoid. Let W be a finitely generated free monoid. A W-semimodule A is said to have the potential star-height k iff A potentially defines an event E with star-height k (Eggan [6]). A is said to have the star-height k iff k is the maximal potential star-height that A has. Since the family of subsets E of W with star-height k is symmetrical, we have that the class of all W-semimodules with a potential star-height k, and the class of all W-semimodules with star-height k, are both categorical typical in $\$^W$.

We now can make some concluding remarks with respect to our first problem discussed in the Introduction to Part III. The categorical predicates in $\mathcal{S}^W$ are those predicates which are not affected by an automorphism of W. We reviewed several actual properties of automata and proved that they are categorical in $\mathcal{S}^W$. However, with respect to the other side of the fence, we should be aware of the following two facts. First, a significant part of automata theory is concerned with automata with output. Consequently, the categories $\mathcal{S}^W$ are not suitable for the representation of such automata. One must consider and examine categories whose objects are the automata under study. Finally, even if we are interested only in semimodules, the significance of the non-categorical properties of W-semimodules is not obvious. Sometimes, the input is dictated in such a way that we are not free to relabel it under automorphisms. In this case, it is very probable that we have to deal with non-categorical properties of semimodules. The usual case in automata theory is, however, that only properties which are preserved under automorphisms of the input monoid are studied.

CHAPTER II: A CATEGORICAL VERSION OF THE STUDY OF SEMIMODULES

In this chapter we indirectly discuss our second problem: how useful categorical algebra is in the understanding and in the study of automata? The answer to such a question is not a matter of proof. What we plan to do is to show by means of a brief example that once the right questions are asked with respect to semimodules, categorical algebra can serve as a good guide in the search for their solution. Roughly speaking, a categorical study of any mathematical domain is any study which is directed towards a solution of any kind of problem via the study of functors and natural transformations. The example that we are going to discuss here is quite elementary from the point of view of categorical algebra, because it does not rise above the level of the notion of functor. Yet, it seems that in spite of the effort in "categorizing" the arguments of a given study (and not to mention the effort in becoming acquainted with categorical algebra), we get a better understanding of the mathematical phenomena under study.

In addition to the categorical approach, we want to illustrate a categorical interpretation of the customary notation used to refer to automata or to semimodules. This interpretation is based on our most useful lemma in this paper. From Lemma 4.1 (in Part II, Chapter II, Section 4) we have the equality

$$gf_s = f_{g(s)} ,$$

valid for any morphism $g : A \longrightarrow B$ of $\mathbb{S}^W$ and any $s \in S(M)$, and where $f_s : M_W \longrightarrow A$ is given by $f_s(1) = s$, and $f_{g(s)} : M_W \longrightarrow B$ is given by $f_{g(s)}(1) = s$. From this equality we have inferred the equality

$$f_s f_w = f_{s \circ w} ,$$

valid for all $f_s : M_W \longrightarrow A$ and $f_w : M_W \longrightarrow M_W$, and the equality

$$f_{w_1} f_{w_2} = f_{w_1 w_2}$$

valid for any $f_{w_1} f_{w_2} : M_W \longrightarrow M_W$. These equalities lead us to change our notation as follows. We will identify any $s \in S(A)$ with $f_s : M_W \longrightarrow A$. Hence $w \in W$ will mean also $f_w : M_W \longrightarrow M_W$. For any morphism $f : A \longrightarrow B$ and $s \in S(A)$ we will denote by fs, either the state $f(s)$, or the morphism ff_s. Under this change of notation, $\lambda_A(s,w)$ will be denoted by sw, of course. Any function $f : S(A) \longrightarrow S(B)$ will be taken as a function $f : \mathbf{S}^W(M_W,A) \longrightarrow \mathbf{S}^W(M_W,B)$, but once it is recognized that it is an underlying function of a homomorphism (i.e., $f(sw) = f(s)w$), it will be identified with $f : A \longrightarrow B$. The above listed equalities prove that this change of notation is consistent, and in the forthcoming discussion we will hardly mention its existence. Above all this we reserve the right to use the old notion without any previous warning.

5. A categorical reformulation and examination of Fleck's theory of perfect automata

In this section we reformulate some of Fleck's [7] results concerning perfect automata (cf. also Weeg [18]). Perfect semimodules are defined as strongly connected abelian semimodules. It is proven that if A is a perfect W-semimodule, then the transition monoid $Mon(A)$ of A is identical with $Aut(A)$, the automorphism group of A in $\mathbf{S}^W$, and it has the same cardinality as $S(A)$. Furthermore, it is proven that a perfect semimodule is a product of two semimodules B and C iff $Aut(A)$ is a product of $Aut(B)$ and $Aut(C)$. All these results are due to Fleck [7].

DEFINITION 5.1. A W-semimodule A is <u>strongly connected</u> iff every $s \in S(A)$ is epic.

The projectivity of M_W implies the equivalence of this characterization with the customary definition of strongly connected semimodules (cf. Chapter 1, Section 5). For the projectivity of M_W means precisely that for any $s_1 \in S(B)$ and any epic $f : A \longrightarrow B$ there exists an $s_2 \in S(A)$ with $fs_2 = s_1$. Hence, if A is strongly connected, then for any $s_1, s_2 \in S(A)$ there exists a $w \in W$ with $s_1 w = s_2$. On the other hand assume that this holds for a W-semimodule A and let $s \in S(A)$. Assume also that for $f_1, f_2 : A \longrightarrow B$ we have $f_1 s = f_2 s$; hence for all $w \in W$, $f_1 sw = f_2 sw$. Therefore $f_1 = f_2$, which implies that $s \in S(A)$ is epic. The projectivity of M_W also implies that if A is strongly connected and $e : A \longrightarrow B$ is epic, then B is strongly connected. For let $s_2 \in S(B)$, then there exists an $s_1 \in S(A)$ with $es_1 = s_2$. But A is strongly connected and so s_1 is epic, and therefore $s_2 = es_1$, is also epic.

DEFINITION 5.2. If A is a W-semimodule, then Aut(A) is the group of the two-sided units (i.e., the invertible morphisms) in the endomorphism monoid E(A) of A in S^W. Mon(A) is the monoid of all the functions $\tau_w^A : S(A) \longrightarrow S(A)$ with $(s)\tau_w^A = sw$, for $w \in W$ (cf. Section 2).

We say that a function $h^* : S(A) \longrightarrow S(B)$ is the <u>result</u> of $h : A \longrightarrow B$ iff $h^*(s) = hs$ for all $s \in S(A)$. Since we will mostly be concerned with functions $S(A) \longrightarrow S(A)$ which belong to Mon(A), we will consistently use the "right-side" notation for <u>functions</u> $S(A) \longrightarrow S(B)$. Thus $h^* : S(A) \longrightarrow S(B)$ is the result of $h : A \longrightarrow B$ iff

$$(s)h^* = hs \quad \text{for all} \quad s \in S(A) .$$

It is clear that a function $h^* : S(A) \longrightarrow S(A)$ is the result of a morphism $h : A \longrightarrow A$ iff for any $w \in W$ we have $h^*\tau_w^A = \tau_w^A h^*$.

Now, for $h_1, h_2 : A \longrightarrow A$ and $s \in S(A)$, if s is epic, then by definition $h_1 s = h_2 s$ implies $h_1 = h_2$. Hence if A is strongly connected, $h_1, h_2 : A \longrightarrow A$ and $h_1 s = h_2 s$ for some $s \in S(A)$, then $h_1 = h_2$. In particular, for $h : A \longrightarrow A$, if $hs = s$ for some $s \in S(A)$, then h is the identity morphism of A. In conclusion we have proved the following lemma.

LEMMA 5.3. If A is a strongly connected W-semimodule, then for any $s \in S(A)$ the function

$$s^* : \mathrm{Aut}(A) \longrightarrow S(A) : h \longrightarrow hs$$

is injective. In particular, the cardinality of $\mathrm{Aut}(A)$ cannot exceed that of $S(A)$.

Obviously, the injections

$$s^* : \mathrm{Aut}(A) \longrightarrow S(A) ,$$

as defined in Lemma 5.3, determine the orbits of $\mathrm{Aut}(A)$ as acting on $S(A)$ (cf. [3: pp.35-39]). Next we prove the well known property of groups of transformations, that the orbits of $\mathrm{Aut}(A)$ in $S(A)$ partition $S(A)$ into disjoint and isomorphic subsets. Denote by $\mathrm{orb}(s)$ the image of s^* in $S(A)$; i.e.,

$$\mathrm{orb}(s) = \{hs : h \in \mathrm{Aut}(A)\} .$$

LEMMA 5.4. Let A be strongly connected W-semimodule and $s_1, s_2 \in S(A)$. If $\mathrm{orb}(s_1) \cap \mathrm{orb}(s_2) \neq \emptyset$, then $\mathrm{orb}(s_1) = \mathrm{orb}(s_2)$.

PROOF. If $orb(s_1) \cap orb(s_2) \neq \emptyset$, then $h_1s_1 = h_2s_2$ for some $h_1, h_2 \in Aut(A)$. Hence $h_2^{-1}h_1s_1 = s_2$. Letting $h = h_2^{-1}h_1$, we have

$$\begin{aligned} orb(s_2) &= \{gs_2 : g \in Aut(A)\} \\ &= \{ghs_1 : g \in Aut(A)\} \\ &= \{g's_1 : g' \in Aut(A)\} \\ &= orb(s_1) . \end{aligned}$$

COROLLARY 5.5. Let A^* be the quotient set of $S(A)$ under the partition of $S(A)$ into the orbits $orb(s)$, $s \in S(A)$. Then there exists a bijection

$$A^* x Aut(A) \longrightarrow S(A) .$$

PROOF. If $c : A^* \longrightarrow A$ is any choice function then

$$\phi_c : A^* x Aut(A) \longrightarrow S(A) : (a,h) \longrightarrow hc(a)$$

(where $a \in A^*$) is by Lemmata 5.3 and 5.4 a bijection.

Thus, in particular, if $S(A)$ is a finite strongly connected semi-module, then the cordinality of $Aut(A)$ divides that of $S(A)$. We add now the requirement that A be abelian.

DEFINITION 5.6. A W-semimodule A is <u>abelian</u> iff $Mon(A)$ is abelian; i.e., iff for any $s \in S(A)$ and any $w_1, w_2 \in W$, $sw_1w_2 = sw_2w_1$. In particular, an abelian strongly connected semimodule is called <u>perfect</u>.

From Chapter I, Section 2, we know that an epic morphism $e : A \longrightarrow B$ induces a surjective monoid homomorphism

$$Mon(e) : Mon(A) \longrightarrow Mon(B) .$$

Hence if A is abelian and $e : A \longrightarrow B$ is epic, then B is abelian too; and if A is perfect, then B is perfect too. We prove now that if A is perfect, then the transition monoid is the group of the underlying functions of $\mathrm{Aut}(A)$. With an abuse of notation we write

LEMMA 5.7. If A is a perfect W-semimodule then $\mathrm{Mon}(A) = \mathrm{Aut}(A)$.

PROOF. Since $\mathrm{Mon}(A)$ is abelian, every $\tau_w^A : S(A) \longrightarrow S(A)$ is the result of some endomorphism $t_w^A : A \longrightarrow A$ with $t_w^A s = (s)\tau_w^A$. Since A is strongly connected, τ_w^A is surjective and so t_w^A is epic. From now on we drop the superscript A from τ_w^A , and we therefore write $\tau_w s = sw$. Assuming that for some $s_1, s_2 \in S(A)$, $\tau_w s_1 = \tau_w s_2$, then, since A is strongly connected, there exists a $w_1 \in W$ such that $s_1 = s_2 w = \tau_w s_2 = \tau_w s_1$; i.e., $s_1 = \tau_w s_1$. This implies that τ_w is the identity morphism of A and $s_1 = s_2$. Hence every τ_w is an isomorphism of A onto A ; i.e., $\mathrm{Mon}(A) \subseteq \mathrm{Aut}(A)$.

On the other hand, let $h \in \mathrm{Aut}(A)$ and $s_o \in S(R)$. Since A is strongly connected, we have $hs_o = s_o w_h$ for some $w_h \in W$. We prove now that for all $s \in S(A)$, $hs = sw_h$. So let $w \in W$ be such that $s = s_o w$, then

$$hs = hs_o w = s_o w_h w = s_o w w_h = s w_h .$$

Hence $h = \tau_{w_h}$. In conclusion, $\mathrm{Mon}(A) = \mathrm{Aut}(A)$.

COROLLARY 5.8. If A is a perfect semimodule, then for any $s \in S(A)$, $\mathrm{orb}(s) = S(A)$. Thus we have a bijection

$$s^* : \mathrm{Aut}(A) \longrightarrow S(A) : h \longrightarrow hs .$$

The identity of $\mathrm{Mon}(A)$ with $\mathrm{Aut}(A)$, for perfect semimodules A , implies directly that $\mathrm{Aut}(A)$ is a (monoid) homomorphic image of W , and that epic morphisms $e : A \longrightarrow B$ induce surjective monoid homomorphisms $Aut(A) \longrightarrow Aut(B)$ which are compatible with the homomorphisms from W onto $Aut(A)$ and $Aut(B)$. We prove now a partial converse to Corollary 5.8.

PROPOSITION 5.9. If A is a strongly connected semimodule such that $orb(s) = S(A)$ for some $s \in S(A)$ and if $Aut(A)$ is abelian, then A is perfect.

PROOF. It suffices to show that A is abelian, so let $w_1, w_2 \in W$. Since $orb(s) = S(A)$, there are $h_1, h_2 \in Aut(A)$ with

$$h_1 s = s w_1 \quad \text{and} \quad h_2 s = s w_2 .$$

Now since $Aut(A)$ is abelian, we get

$$s w_1 w_2 = h_1 s w_2 = h_1 h_2 s = h_2 h_1 s = h_2 s w_1 = s w_2 w_1 ,$$

and therefore A is abelian.

COROLLARY 5.10. If A is a finite strongly connected semimodule and $Aut(A)$ is abelian with the same cordinality as $S(A)$, then A is perfect.

In general, for an arbitrary semimodule A, a morphism from A does not induce any group homomorphism of $Aut(A)$. We are inclined to say that in order to derive from the assignment $A \longrightarrow Aut(A)$ a functor of a suitable subcategory, one does not have much choice, and has to consider perfect semimodules. Furthermore, the previous results precisely establish the fact that the assignment $A \longrightarrow Aut(A)$, as applied to perfect W-semimodules, determines a functor

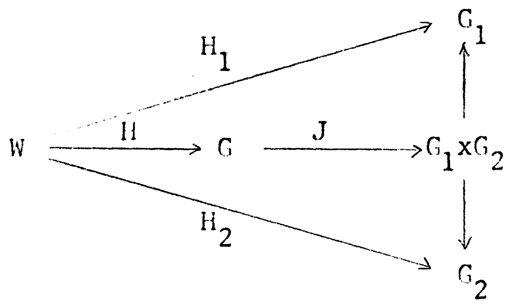
$$Tau : \mathfrak{P}er(W) \longrightarrow \mathfrak{G}im(W)$$

from a category of perfect W-semimodules to a category of surjective homomorphisms of W with group images. To be precise, the category Per(W) has all the perfect W-semimodules as objects and all the possible morphisms (they must be epic!) among them. For each object A of Per(W) , we define

$$\mathrm{Tau}(M) = (\tau^A : W \longrightarrow G(A))$$

where $G(A) = \mathrm{Mon}(A) = \mathrm{Aut}(A)$, and τ^A is the surjective monoid homomorphism defined by $\tau^A(w) = \tau^A_w$. We noted before that every epic morphism of perfect W-semimodules, that is, every morphism $e : A \longrightarrow B$ induces a monoid homomorphism $\mathrm{Mon}(e) : G(A) \longrightarrow G(B)$ with $\mathrm{Mon}(e) \circ \tau^A = \tau^B$. It is natural, therefore, to consider the category Gim(W) whose objects are surjective monoid homomorphisms $H : W \longrightarrow G$ onto abelian groups. There is a morphism of Gim(W) from $H_1 : W \longrightarrow G_1$ to $H_2 : W \longrightarrow G_2$ whenever there exists a homomorphism $H' : G_1 \longrightarrow G_2$ with $H'H_1 = H_2$. Hence, for any morphism $e : A \longrightarrow B$ of Per(W) , the morphism $\mathrm{Tau}(e) : \mathrm{Tau}(A) \longrightarrow \mathrm{Tau}(B)$ is defined to be the one determined by Mon(e) .

Now our strategy is the following. We prove that the functor $\mathrm{Tau} : \mathrm{Per}(W) \longrightarrow \mathrm{Gim}(W)$ relate products. That is, every product diagram in Per(W) is mapped under Tau onto a product diagram in Gim , and every product diagram in Gim(W) which comes from a diagram in Per(W) via Tau comes from a product diagram in Per(W) . But before we do so we have first to be convinced that products in Per(W) yield the same product objects as in S^W . For this, it suffices to show that a product object in the category of all epic morphisms of S^W is precisely a product object in S^W . Put differently, that a diagram $A_2 \longleftarrow P \longrightarrow A_1$ of epic morphisms is a product diagram in S^W iff it is a product diagram in the subcategory of S^W of all the epic morphisms. We also have to be convinced that if $H : W \longrightarrow G$ is a product object of $H_1 : W \longrightarrow G_1$ and $H_2 : W \longrightarrow G_2$ in $\mathbb{G}\mathrm{im}(W)$, then G is a product of G_1 and G_2 in $\mathbb{G}^{ab}$. And, conversely, for any objects $H_1 : W \longrightarrow G_1$ and $H_2 : W \longrightarrow G_2$ of $\mathbb{G}\mathrm{im}(W)$, there exists another object $H : W \longrightarrow G_1 x G_2$ of $\mathbb{G}\mathrm{im}(W)$ such that $H : W \longrightarrow G_1 x G_2$ is a product object of the given two objects in $\mathbb{G}\mathrm{im}(W)$. In order to prove this, it suffices to show that $H : W \longrightarrow G$ is a product of $H_1 : W \longrightarrow G_1$ and $H_2 : W \longrightarrow G_2$ in $\mathbb{G}\mathrm{im}$ iff there exists an isomorphism $J : G \longrightarrow G_1 x G_2$ with



commutative, where

$$G_1 \xleftarrow{P_1} G_1 x G_2 \xrightarrow{P_2} G_2$$

is a product diagram in $\mathbb{G}^{ab}$. The proof of these connections between products of $\mathbb{G}\mathrm{im}(W)$ and $\mathbb{G}^{ab}$, and between $\mathbb{P}\mathrm{er}(W)$ and $\mathbb{S}^W$, are straightforward product diagram chasing, and it is left for the reader. We prove however that the functor

$$\mathrm{Tau} : \mathbb{P}\mathrm{er}(W) \longrightarrow \mathbb{G}\mathrm{im}(W)$$

relates products.

We define now a second functor

$$\mathrm{Act} : \mathbb{G}\mathrm{im}(W) \longrightarrow \mathbb{P}\mathrm{er}(W)$$

as follows. Every object $H : W \longrightarrow G$ of $\mathbb{G}im(W)$ determines a perfect W-semimodule G^H, which amounts to the action of W on G via $H : W \longrightarrow G$. That is,

$$S(G^H) = G \quad \text{and} \quad g \cdot w = gH(w) .$$

Since H is surjective, G^H is strongly connected, and since G is abelian, G^H is abelian too. Furthermore, $\mathrm{Aut}(G^H)$ is isomorphic to G. Every morphism h from $H_1 : W \longrightarrow G_1$ to $H_2 : W \longrightarrow G_2$ of $\mathbb{G}im(W)$ is determined by a homomorphism $H' : G_1 \longrightarrow G_2$ with $G'G_1 = G_2$. Hence $H' : G_1 \longrightarrow G_2$ determines a morphism $H' : G_1^{H_1} \longrightarrow G_2^{H_2}$, and so we define $\mathrm{Act}(h) = H'$.

We prove the following two lemmata, which express interesting relationships between the functors Act and Tau. These lemmata will imply that Tau relates products.

LEMMA 5.11. Let A be a perfect W-semimodule and $H : W \longrightarrow G$ an object of $\mathbb{G}im(W)$ with a morphism of $\mathbb{G}im(W)$ into $\mathrm{Tau}(A)$; i.e., with an epimorphism $E : G \longrightarrow \mathrm{Mon}(A)$ such that $EH = \tau^A$. Then for any $s \in S(A)$ there exists an epic morphism

$$e_s : G^H \longrightarrow A$$

of $\mathbb{P}er(W)$.

PROOF. Define $e_s : G^H \longrightarrow A$ by $e_s(g) = E(g)(s)$. Then for all $w \in W$,

$$e_s(g \cdot w) = e_s(gH(w)) = E(gH(w))(s) = E(g)EH(w)(s)$$

$$= E(g)\tau^A(w)(s) = E(g)sw = e_s(g)w .$$

Similarly, we have

LEMMA 5.12. Let A and B be perfect W-semimodules with a morphism of $\mathbb{G}$im(W) from Tau(A) into Tau(B) ; i.e., with an epimorphism

$$E : \mathrm{Mon}(A) \longrightarrow \mathrm{Mon}(B)$$

such that $E\tau^A = \tau^B$. Then for any $s_1 \in S(A)$ and $s_2 \in S(B)$ there exists an (epic) morphism

$$e_{s_1,s_2} : A \longrightarrow B$$

of $\mathbb{P}$er(W) with

$$e_{s_1,s_2}(s_1 w) = s_2 w .$$

PROOF: We define $e_{s_1,s_2} : A \longrightarrow B$ by

$$e_{s_1,s_2}(s_1 w) = s_2 w .$$

Since $s_1 w_1 = s_1 w_2$ implies $\tau_{w_1} s_1 = \tau_{w_2} s_1$, and therefore $\tau_{w_1} = \tau_{w_2}$. Hence $s_1 w_1 = s_1 w_2$ implies $s_2 w_1 = s_2 w_2$, and therefore e_{s_1,s_2} is well defined. Obviously it determines a morphism of $\mathbb{P}$er(W) .

Thus Lemma 5.11 states that if $\mathbb{G}$im(H,Tau(A)) is not empty, then $\mathbb{P}$er(Act(H),A) is note empty. Similarly, Lemma 5.12 states that if $\mathbb{G}$im(Tau(A),Tau(B)) is not empty, then $\mathbb{P}$er(A,B) is not empty. Our final theorem follows now directly.

THEOREM 5.13. The functor

$$\mathrm{Tau} : \mathbb{P}\mathrm{er}(W) \longrightarrow \mathbb{G}\mathrm{im}(W)$$

relates products. That is, a family $\{p_j : P \longrightarrow A_j\}$ of morphisms of $\mathbb{P}$er(W) is a product diagram in $\mathbb{P}$er(W) iff the induced family

$$\{\mathrm{Tau}(p_j) : \mathrm{Tau}(P) \longrightarrow \mathrm{Tau}(A_j)\}$$

is a product diagram in $\mathbb{G}im(W)$.

Now, because of the asserted relations between the products in $\mathbb{G}im(W)$ and the products in $\mathbb{G}^{ab}$, and between the products in $\mathbb{P}er(W)$ and the products in $\mathbb{S}^{W}$, we have

COROLLARY 5.14. (Fleck) A perfect W-semimodule A is a product object of the family $\{A_j\}$ in $\mathbb{P}er(W)$ iff Mon(A) is the direct product of $\{Mon(A_j)\}$ in the category of abelian groups.

BIBLIOGRAPHY

1. Birkhoff, G., "Structure of Abstract Algebras," Proc. Cambridge Philos. Soc., 29, 441-464 (1935).

2. Büchi, J. R., "Mathematical Theory of Automata," Notes on material presented by J. B. Wright and J. R. Büchi, Communication Sciences 403, Fall 1960, The University of Michigan (Unpublished).

3. Chevalley, C., Fundamental Concepts of Algebra, Acad. Press, New York (1956).

4. Chomsky, N. and Schützenberger, M. P., "The Algebraic Theory of Context-Free Languages," in Computer Programming and Formal Systems, Edited by P. Braffort and D. Hirschberg, North Holland, Amsterdam (1963).

5. Cohn, P. M., Universal Algebra, Harper's Series in Modern Mathematics, Harper & Row, New York (1965).

6. Eggan, L. C., "Transition Graphs and the Star-Height of Regular Events," Michigan Math. J., 10, 385-397 (1963).

7. Fleck, A. C., "Isomorphism Groups of Automata," J. ACM, 9, 469-476 (1962).

8. Freyd, P., Abelian Categories: An Introduction to the Theory of Functors, Harper's Series in Modern Mathematics, Harper & Row, New York (1964).

9. Ginsburg, S., An Introduction to Mathematical Machine Theory, Addison-Wesley, Reading, Mass. (1962).

10. Hartmanis, J. and Stearns, R. E., "Pair Algebra and Automata Theory," Inf. Con., 7, 485-507 (1964).

11. Kleene, S. C., Introduction to Metamathematics, Van-Nostrand, Princeton (1952).

12. Krohn, K. B. and Rhodes, J. L., "Algebraic Theory of Machines," Proc. Symp. Math. Theory of Automata, Polytechnic Press, Brooklyn, New York, 341-384 (1962).

13. MacLane, S., "Categorical Algebra," Bull. AMS, 71, 40-106 (1965).

14. MacLane, S., Homology, Academic Press, New York (1963).

15. Perles, M., Rabin, M. O., and Shamir, E., "The Theory of Definite Automata," IEEE Transactions on Electronic Computers, Vol. EC-12, No.3, June, 1963, 233-243.

16. Rabin, M. O. and Scott, D., "Finite Automata and Their Decision Problems," IBM J. 3, 114-125 (1959).

BIBLIOGRAPHY (Concluded)

17. Thatcher, J. W., "Notes on Mathematical Automata Theory," Tech. Note 05662, 03105-26-T, ORA, The University of Michigan (1963).

18. Weeg, G. P., "The Group and Semigroup Associated with Automata," Proc. Symp. Math. Theory of Automata, Polytechnic Press, Brooklyn, New York, 257-266 (1962).

19. Zeiger, H. P., "Loop Free Synthesis of Finite State Machines," Thesis, M.I.T. (1964).

INDEX

of the most re-occurring symbols ordered
according to their first occurrence

DISTRIBUTION LIST

(One copy unless otherwise noted)

Technical Library
Director Defense Res. & Eng.
Room 3C-128, The Pentagon
Washington, D.C. 20301

Defense Documentation Center 20
Cameron Station
Alexandria, Virginia 22314

Chief of Naval Research 2
Department of the Navy
Washington 25, D.C.
Attn: Code 437, Information Systems Branch

Director, Naval Research Laboratory 6
Technical Information Officer
Washington 25, D.C.
Attention: Code 2000

Commanding Officer 10
Office of Naval Research
Navy 100, Fleet Post Office Box 39
New York, New York 09599

Commanding Officer
ONR Branch Office
207 West 24th Street
New York 11, New York

Office of Naval Research Branch Office
495 Summer Street
Boston, Massachusetts 02110

Naval Ordnance Laboratory
White Oaks, Silver Spring 19
Maryland
Attn: Technical Library

David Taylor Model Basin
Washington,D.C. 20007
Attn: Code 042, Technical Library

Naval Electronics Laboratory
San Diego 52, California
Attn: Technical Library

Dr. Daniel Alpert, Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois

Air Force Cambridge Research Labs
Laurence C. Hanscom Field
Bedford, Massachusetts
Attn: Research Library, CRMXL R

U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448
Attn: G. H. Gleissner, Code K4
Asst. Dir. for Computation

National Bureau of Standards
Data Processing Systems Division
Room 239, Building 10
Washington 25, D.C.
Attn: A. K. Smilow

George C. Francis
Computing Laboratory, BRL
Aberdeen Proving Ground, Maryland

Office of Naval Research
Branch Office, Chicago
230 North Michigan Avenue
Chicago, Illinois 60601

Commanding Officer
ONR Branch Office
1030 E. Green Street
Pasadena, California

Commanding Officer
ONR Branch Office
1000 Geary Street
San Francisco 9, California

DISTRIBUTION LIST (Concluded)

The University of Michigan
Department of Philosophy
Attn: Professor A. W. Burks

National Physical Laboratory
Teddington, Middlesex, England
Attn: Dr. A. M. Uttley, Supt.
Autonomics Division

Commanding Officer
Harry Diamond Laboratories
Washington, D.C. 20438
Attn: Library

Commanding Officer and Director
U. S. Naval Training Device Center
Port Washington
Long Island, New York
Attn: Technical Library

Department of the Army
Office of the Chief of Research
and Development
Pentagon, Room 3D442
Washington 25, D.C.
Attn: Mr. L. H. Geiger

National Security Agency
Fort George G. Meade, Maryland
Attn: Librarian, C-332

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington 73, Massachusetts
Attn: Library

Office of Naval Research
Washington 25, D.C.
Attn: Code 432

Dr. Kenneth Krohn
Krohn Rhodes Research Institute, Inc.
328 Pennsylvania Avenue, S. E.
Washington 13, D. C.

Dr. Larry Fogel
Decision Science, Inc.
6508 Pacific Highway
San Diego, California

National Bureau of Standards
Applications Engineering Section
Washington 25, D. C.
Attn: Miss Mary E. Stevens

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Logic of Computers Group The University of Michigan Ann Arbor, Michigan 48104	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ON SOME CATEGORICAL ALGEBRA ASPECTS OF AUTOMATA THEORY: THE CATEGORICAL PROPERTIES OF TRANSITION SYSTEMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (Last name, first name, initial)

Give'on, Yehoshafat

6. REPORT DATE February 1966	7a. TOTAL NO. OF PAGES 173	7b. NO. OF REFS 19
8a. CONTRACT OR GRANT NO. Nonr 1224(21) b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 03105-43-T 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Department of the Navy Washington, D.C.

13. ABSTRACT

The ubiquity and usefulness of homomorphisms in various studies of automata lead us to consider the following problem. What can be said on automata by referring only to homomorphisms of automata? In the report we present a study of this problem with respect to a special type of automaton, namely with respect to transition systems.

Categorical algebra methods are applied to the precise formulation of this problem and to its solution. We find that if W is a monoid belonging to a broad class of monoids, then the categorical abstract properties of transition systems with input W are determined by the automorphisms of the monoid W. In particular, any property of automata without output is categorical iff it does not depend on the particular labeling of the input alphabet.

This study of the categorical properties of automata has two additional outcomes. First, we realize that categorical algebra methods can be applied to automata with arbitrary input monoids, with results pertinent to the theory of monoids. On the other hand it indicates a possible usefulness in the study of automata, in particular, in getting a better understanding of the mathematical ideas employed in automata theory.

In order to support this point of view with respect to automata theory, we show that many actually studied properties of automata are categorical. And we give an example of a categorical examination and formulation of a particular study of perfect automata.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
1. algebraic automata theory						
2. categorical algebra						
3. homomorphisms						
4. transition systems						
5. monoids						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
________________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
________________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
________________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNIVERSITY OF MICHIGAN
3 9015 02825 9987