Reprints in Theory and Applications of Categories, No. 15, 2006, pp. 1–13.

DIAGONAL ARGUMENTS AND CARTESIAN CLOSED CATEGORIES

F. WILLIAM LAWVERE

Author Commentary

In May 1967 I had suggested in my Chicago lectures certain applications of category theory to smooth geometry and dynamics, reviving a direct approach to function spaces and therefore to functionals. Making that suggestion more explicit led later to elementary topos theory as well as to the line of research now known as synthetic differential geometry. The fuller development of those subjects turned out to involve a truth value object that classifies subobjects, but in the present paper (presented in the 1968 Battelle conference in Seattle) I refer only to weak properties of such an object; it is the other axiom, cartesian closure, that plays the central role.

Daniel Kan had recognized that the function space construction for simplicial sets and other categories is a right adjoint, thus unique. Because this uniqueness property of adjoints implies their main calculational rules, I took the further axiomatic step of defining functor categories as a right adjoint to the finite product construction in my 1963 thesis. In 1965, Eilenberg and Kelly introduced the term *closed* to mean that there is a hom functor valued in the category itself. Such a hom functor is characterized in a relative way as right adjoint to a given tensor product functor; we concentrate here on the absolute case where the tensor is cartesian.

Although the cartesian-closed view of function spaces and functionals was intuitively obvious in all but name to Volterra and Hurewicz (and implicitly to Bernoulli), it has counterexamples within the rigid framework advocated by Dieudonné and others. According to that framework the only acceptable fundamental structure for expressing the cohesiveness of space is a contravariant algebra of open sets or possibly of functions. Even though such algebras are of course extremely important invariants, their nature is better seen as a consequence of the covariant geometry of figures. Specific cases of this determining role of figures were obvious in the work of Kan and in the popularizations of Hurewicz's k-spaces by Kelley, Brown, Spanier, and Steenrod, but in the present paper I made this role a matter of principle: the Yoneda embedding was shown to preserve

Received by the editors 2005-12-05.
Transmitted by R. Paré, R. Rosebrugh and R.J. Wood. Reprint published on 2006-02-24.
2000 Mathematics Subject Classification: 08.10, 02.00.
Originally published as: Lecture Notes in Mathematics, No. 92, pages 134-145, 1969; used by permission.

cartesian closure, and naturality of functionals was shown to be equivalent to Bernoulli's principle. Further, I posed the problem of comparing this principle to practice in the specific cases of smooth and recursive mathematics.

Later detailed work on those particular cases justified the classical intuition embodied in my general definition. In their books Froelicher and Kriegl (1988), and Kriegl and Michor (1997), extensively develop smooth analysis; their higher-order use of the adequacy of figures is based in part on a lower-order result of Boman 1967 (implicit in Hadamard) concerning the adequacy of paths. They cite the result of Lawvere-Schanuel-Zame showing that the natural functionals in this case are indeed the distributions of compact support, as practice would suggest. Nilpotent infinitesimals fall far short of even one-dimensionality, but if taken to be non-commutative, are already adequate for holomorphic functions, as was strikingly shown by Steve Schanuel (1982). The recursive example was studied by Phil Mulry (1982) who constructed a topos that does include as full sub-categories both the Banach-Mazur and the Ersov versions of higher recursive functionals.

I hope that in the future this adequacy of one-dimensional figures will be explained because it occurs in many different examples. Many kinds of cohesion (algebraic geometry, smooth geometry, continuous geometry) are well-expressed as a subtopos of the classifying topos of a finitary single-sorted algebraic theory. But often that algebraic theory is determined by its monoid M of unary operations via naturality only: for example, the binary operations, instead of being independently specified, are just the maps of right M-sets from M^2 to M. If a common explanation can be found (for this adequacy of one-dimensional considerations in the determination of n-dimensional and infinite-dimensional functionals, in so many disparate cases) it would further establish that the Eilenberg-Mac Lane notion of naturality is far more powerful than the mere tautology it is sometimes considered to be.

The original aim of this article was to demystify the incompleteness theorem of Gödel and the truth-definition theory of Tarski by showing that both are consequences of some very simple algebra in the cartesian-closed setting. It was always hard for many to comprehend how Cantor's mathematical theorem could be re-christened as a"paradox" by Russell and how Gödel's theorem could be so often declared to be the most significant result of the 20th century. There was always the suspicion among scientists that such extra-mathematical publicity movements concealed an agenda for re-establishing belief as a substitute for science. Now, one hundred years after Gödel's birth, the organized attempts to harness his great mathematical work to such an agenda have become explicit.

I thank the editors of *TAC*, and especially Christoph Schubert whose dedication and expertise made this reprint possible.

F. William Lawvere
Buffalo, N.Y.
19 February 2006

Introduction

The similarity between the famous arguments of Cantor, Russell, Gödel and Tarski is well-known, and suggests that these arguments should all be special cases of a single theorem about a suitable kind of abstract structure. We offer here a fixed-point theorem in cartesian closed categories which seems to play this role. Cartesian closed categories seem also to serve as a common abstraction of type theory and propositional logic, but the author's discussion at the Seattle conference of the development of that observation will be in part described elsewhere ["Adjointness in Foundations", to appear in Dialectica, and "Equality in Hyperdoctrines and the Comprehension Schema as an Adjoint Functor", to appear in the Proceedings of the AMS Symposium on Applications of Category theory].

1. Exponentiation, surjectivity, and a fixed-point theorem

By a cartesian closed category is meant a category $\mathbf{C}$ equipped with the following three kinds of right-adjoints: a right adjoint 1 to the unique

$$\mathbf{C} \longrightarrow \mathbf{1},$$

a right adjoint $\times$ to the diagonal functor

$$\mathbf{C} \longrightarrow \mathbf{C} \times \mathbf{C},$$

and for each object A in $\mathbf{C}$, a right-adjoint $(\ )^A$ to the functor

$$\mathbf{C} \xrightarrow{A \times (\ )} \mathbf{C}.$$

The adjunction transformations for these adjoint situations, also assumed given, will be denoted by δ, π in the case of products and by λ_A, ϵ_A in case of exponentiation by A. Thus for each X one has

$$X \xrightarrow{X\lambda_A} (A \times X)^A$$

and for each Y one has

$$A \times Y^A \xrightarrow{Y\epsilon_A} Y.$$

Given $f : A \times X \longrightarrow Y$, the composite morphism

$$X \xrightarrow{X\lambda_A} (A \times X)^A \xrightarrow{f^A} Y^A$$

will be called the "λ-transform" of the morphism f. A morphism $h : X \longrightarrow Y^A$ is the λ-transform of f iff the diagram

$$\begin{tikzcd} A \times X \arrow[d, "A\times h"'] \arrow[dr, "f"] & \\ A \times Y^A \arrow[r, "Y\epsilon_A"'] & Y \end{tikzcd}$$

is commutative, showing in particular that f can be uniquely recovered from its λ-transform. Taking the case $X = 1$, one has that every $f : A \longrightarrow Y$ gives rise to a unique $\ulcorner f \urcorner : 1 \longrightarrow Y^A$ and that every $1 \longrightarrow Y^A$ is of that form for a unique f. Since for every $a : 1 \longrightarrow A$ one has (dropping the indices A, Y on ϵ when they are clear)

$$\langle a, \ulcorner f \urcorner \rangle \epsilon = a.f,$$

one calls ϵ the "evaluation" natural transformation; note however that we do *not* assume in general that f is determined by the knowledge of all its "values" $a.f$.

Although we do not make use of it in this paper, the usefulness of cartesian closed categories as algebraic versions of type theory can be further illustrated by assuming that the coproduct

$$2 = 1 + 1$$

also exists in $\mathbf{C}$. It then follows (using the closed structure), that for every object A

$$A \times 2 = A + A$$

and so in particular that 2 is a Boolean-algebra-object in $\mathbf{C}$, i.e. that among the morphisms

$$2 \times 2 \times \ldots \times 2 \longrightarrow 2$$

in $\mathbf{C}$ there are well determined morphisms corresponding to all the finitary (two-valued) truth tables, and that these satisfy all the commutative diagrams expressing the axioms of Boolean algebra. Equivalently, for each X the set

$$P_{\mathbf{C}}(X) = \mathbf{C}(X, 2)$$

of "$\mathbf{C}$-attributes of type X" becomes canonically an actual Boolean algebra, and varying X along any morphism of $\mathbf{C}$ induces contravariantly a Boolean homomorphism of attribute algebras. The morphisms $1 \longrightarrow 2$ form $P_{\mathbf{C}}(1)$ the Boolean algebra of "truth-values"; among these are the two coproduct injections which play the roles of "true" and "false". For any "constant of type X" $x : 1 \longrightarrow X$ and any attribute φ of type X, $x.\varphi$ is then a truth-value. Now noting that

$$X \times 2^X \xrightarrow{(2)\epsilon_X} 2$$

is a "binary operation" we could write it between its arguments, so that we have

$$x \, \epsilon \ulcorner \varphi \urcorner = x.\varphi,$$

an equality of truth-values; thus if we think of $\ulcorner \varphi \urcorner : 1 \longrightarrow 2^X$ as the constant naming the subset of X corresponding to the attribute φ, one sees that the above equation expresses the usual "comprehension" axiom.

Returning to our immediate concern, we define a morphism $g : X \longrightarrow Z$ to be *point-surjective* iff for every $z : 1 \longrightarrow Z$ there exists $x : 1 \longrightarrow X$ with $x.g = z$. This does not imply that g is necessarily "onto the whole of Z", since there may be few morphisms with domain 1; for example if (as in the next section) X and Z are set-valued functors, then a natural transformation g is point-surjective if every element of the *inverse limit* of Z comes from an element of the inverse limit of X. In case Z is of the form Y^A, an even weaker notion of surjectivity can be considered, which in fact suffices for our fixed point theorem. Namely

$$X \xrightarrow{g} Y^A$$

will be called *weakly point-surjective* iff for every $f : A \longrightarrow Y$ there is x such that for every $a : 1 \longrightarrow A$

$$\langle a, xg \rangle \epsilon = a.f$$

Finally we say that an object Y has the *fixed point property* iff for every endomorphism $t : Y \longrightarrow Y$ there is $y : 1 \longrightarrow Y$ with $y.t = y$.

1.1. Theorem. *In any cartesian closed category, if there exists an object A and a weakly point-surjective morphism*

$$A \xrightarrow{g} Y^A$$

then Y has the fixed point property.

Proof. Let $\overline{g}$ be the morphism whose λ-transform is g. Then for any $f : A \longrightarrow Y$ there is $x : 1 \longrightarrow A$ such that for all $a : 1 \longrightarrow A$

$$\langle a, x \rangle \overline{g} = a.f.$$

Now consider any endomorphism t of Y and let f be the composition

$$A \xrightarrow{A\delta} A \times A \xrightarrow{\overline{g}} Y \xrightarrow{t} Y;$$

thus there is x such that for all a

$$\langle a, x \rangle \overline{g} = \langle a, a \rangle \overline{g} t$$

since $a(A\delta) = \langle a, a \rangle$. But then $y = \langle x, x \rangle \overline{g}$ is clearly a fixed point for t. ■

The famed "diagonal argument" is of course just the contrapositive of our theorem. Cantor's theorem follows with $Y = 2$.

1.2. Corollary. *If there exists $t : Y \longrightarrow Y$ such that $yt \neq y$ for all $y : 1 \longrightarrow Y$ then for no A does there exist a point-surjective morphism*

$$A \longrightarrow Y^A$$

(or even a weakly point-surjective morphism).

2. Russell's Paradox is a case of Cantor's theorem; natural functionals in recursive function theory and smooth manifold theory

Russell's Paradox does not presuppose that set theory be formulated as a higher type theory; that is, for A the set-theoretical universe, we do not need 2^A for the argument. In fact we need only apply the *proof* of our theorem, with $\overline{g} : A \times A \longrightarrow 2$ as the set-theoretical membership relation, dispensing with g entirely. That is, more generally, our theorem could have been stated and proved in any category with *only* finite products (no exponentiation) by simply phrasing the notion of (weak) point-surjectivity as a property of a morphism

$$A \times X \longrightarrow Y;$$

however discovering the latter form (or at least calling it surjectivity!) seems to require thinking of such a morphism as a family of morphisms $A \longrightarrow Y$ indexed by the elements of X, suggesting that a closed category is the "natural" setting for the theorem.

In fact the more general form of the theorem just alluded to (for categories with products) follows from the cartesian closed version which we have proved, by virtue of the following remark. Notice that it would suffice to assume $\mathbf{C}$ small (just take the full closure under finite products of the two objects A, Y).

2.1. Remark. *Any small category* $\mathbf{C}$ *can be fully and faithfully embedded in a cartesian closed category in a manner which preserves any products or exponentials that may exist in* $\mathbf{C}$.

Proof. We consider the usual embedding

$$\mathbf{C} \longrightarrow \mathcal{S}^{\mathbf{C}^{\mathrm{op}}}$$

which identifies an object Y with the contravariant set-valued functor

$$X \longmapsto \mathbf{C}(X, Y).$$

By "Yoneda's Lemma" one has for any functor Y and any object A that the value at A of Y

$$AY \cong \mathcal{S}^{\mathbf{C}^{\mathrm{op}}}(A, Y)$$

where the right hand side denotes the set of all natural transformations from (the functor corresponding to) A into Y, so that in particular the embedding is full and faithful. It is then also clear that the embedding preserves products (in particular if 1 exists in $\mathbf{C}$ it corresponds to the functor that is constantly the one-element set, which is the 1 of $\mathcal{S}^{\mathbf{C}^{\mathrm{op}}}$). For any two functors A, Y the functor

$$C \longmapsto \mathcal{S}^{\mathbf{C}^{\mathrm{op}}}(A \times C, Y)$$

plays the role of Y^A. In particular if B^A exists in $\mathbf{C}$ for a pair of objects A, B in $\mathbf{C}$ then

$$(C)B^A \cong \mathbf{C}(C, B^A) \cong \mathbf{C}(A \times C, B) \cong \mathcal{S}^{\mathbf{C}^{\mathrm{op}}}(A \times C, B)$$

showing that the embedding preserves exponentiation. ∎

2.2. Theorem. *Let A, Y be any objects in any category with finite products (including the empty product 1); then the following two statements cannot both be true*

a) there exists $\overline{g} : A \times A \longrightarrow Y$ such that for all $f : A \longrightarrow Y$ there exists $x : 1 \longrightarrow A$ such that for all $a : 1 \longrightarrow A$

$$\langle a, x \rangle \overline{g} = a.f$$

b) there exists $t : Y \longrightarrow Y$ such that for all $y : 1 \longrightarrow Y$

$$y.t \neq y.$$

Proof. Apply above remark and the proof in the previous section. ∎

Of course the "transcendental" proof just given is somewhat ridiculous, since the incompatibility of a) and b) can be proved directly just as simply as it was proved in the previous section under the more restrictive hypothesis on $\mathbf{C}$. However we wish to take the opportunity to make some further remarks about the above canonical embedding of an arbitrary (small) category into a cartesian closed category $\overline{\mathbf{C}}$ (let the latter denote the smallest full cartesian closed subcategory of $\mathcal{S}^{\mathbf{C}^{\mathrm{op}}}$ which contains $\mathbf{C}$). One of the standard ways of embedding a structure into a higher-order structure is to consider "definable" functionals, operators, etc.; however this is difficult to oversee from a simple-minded point of view since it usually requires enumerating all possible definitions. On the other hand in many situations (e.g. functorial semantics of algebraic theories or functorial semantics of elementary theories if the elementary theories are complete) one has come to expect that natural transformations are identical with definable ones or at least a reasonable substitute for definable ones. The latter alternative seems to be a least partly true in the present case. Thus for example we are led to the following definition. If A, B, C, D are objects in a category $\mathbf{C}$ with finite products, a *natural operator*

$$B^A \xrightarrow{\ \Phi\ } D^C$$

shall be simply a natural transformation between the exponential functors of the (functors corresponding to the) given objects in $\mathcal{S}^{\mathbf{C}^{\mathrm{op}}}$ (hence in $\overline{\mathbf{C}}$). In particular if $C = 1$ we would call a natural operator a natural functional. Note that 1 will not be a generator for all of $\mathcal{S}^{\mathbf{C}^{\mathrm{op}}}$ unless $\mathbf{C} = 1$; however it might conceivably be so for $\overline{\mathbf{C}}$, and we have a partial result in that direction. In fact, in the case that 1 is a generator for $\mathbf{C}$ itself, we can describe in more familiar terms what a natural operator is.

Recall that "1 is a generator for $\mathbf{C}$" simply means that a morphism $f : X \longrightarrow Y$ in $\mathbf{C}$ is determined by its "values" $x.f : 1 \longrightarrow Y$ for $x : 1 \longrightarrow X$. In that case it is sensible to call the elements of the set $\mathbf{C}(1, X)$ of points of X also the *elements of* X. Then a function

$$\mathbf{C}(1, X) \longrightarrow \mathbf{C}(1, Y)$$

is induced by at most one $\mathbf{C}$-morphism $X \longrightarrow Y$, and in case it is, we say by abuse of language that the function *is* a morphism of $\mathbf{C}$.

2.3. PROPOSITION. *Suppose that* $\mathbf{C}$ *is a category with finite products in which* 1 *is a generator, and* A, B, C, D *are objects of* $\mathbf{C}$. *Then*

1) *a natural operator*

$$B^A \xrightarrow{\Phi} D^C$$

is entirely determined by a single function

$$\mathbf{C}(A, B) \xrightarrow{1\Phi} \mathbf{C}(C, D)$$

and

2) *such a function determines a natural operator iff for every object* X *of* $\mathbf{C}$ *and for every* $\mathbf{C}$*-morphism* $f : A \times X \longrightarrow B$, *the function*

$$\mathbf{C}(1, C \times X) \xrightarrow{(f)(X\Phi)} \mathbf{C}(1, D)$$

is a $\mathbf{C}$*-morphism, where* $(f)(X\Phi)$ *is defined by*

$$\langle c, x\rangle\,((f)(X\Phi)) = (c)\,((f_x)(1\Phi))$$

for any $c : 1 \longrightarrow C$, $x : 1 \longrightarrow X$, f_x *denoting the composition*

$$A \cong A \times 1 \xrightarrow{A\times x} A \times X \xrightarrow{f} B.$$

PROOF. We are abusing notations to the extent of identifying a morphism with its λ-transform via the bijections of the form

$$\mathbf{C}(A \times X, B) \cong \overline{\mathbf{C}}(A \times X, B) \cong \overline{\mathbf{C}}(X, B^A).$$

Actually the given operator Φ is a family of functions

$$\mathbf{C}(X, B^A) \xrightarrow{X\Phi} \mathbf{C}(X, D^C)$$

one for each object of $\mathbf{C}$; the "naturalness" condition that this family must satisfy is, via the abuse, that for every morphism $x : X' \longrightarrow X$ of $\mathbf{C}$, the diagram

$$\begin{tikzcd}
\mathbf{C}(A \times X, B) \arrow[r, "X\Phi"] \arrow[d, "x"'] & \mathbf{C}(C \times X, D) \arrow[d, "x"] \\
\mathbf{C}(A \times X', B) \arrow[r, "X'\Phi"'] & \mathbf{C}(C \times X', D)
\end{tikzcd}$$

should commute. Now let $X' = 1$. Since 1 is a generator for $\mathbf{C}$, the value of the function $X\Phi$ at a given $f : A \times X \longrightarrow B$ is determined by the knowledge, for each element x of

X and each element c of C, the result reached in the lower right hand corner by going first across, then down, in the commutative diagram

$$\begin{tikzcd}
\mathbf{C}(A \times X, B) \arrow[r, "X\Phi"] \arrow[d, "x"'] & \mathbf{C}(C \times X, D) \arrow[r, "c"] \arrow[d, "x"] & \mathbf{C}(X, D) \arrow[d, "x"] \\
\mathbf{C}(A, B) \arrow[r, "1\Phi"'] & \mathbf{C}(C, D) \arrow[r, "c"'] & \mathbf{C}(1, D).
\end{tikzcd}$$

But since the same results are obtained by going down, then across, all the functions $X\Phi$ are determined by one function 1Φ, proving the first assertion. The second assertion is then clear, since the definition of $(f)(X\Phi)$ given in the statement of the proposition is just such as to assure naturality of $X\Phi$ provided its values exist. ■

To make the situation perfectly clear, notice that morphisms whose codomain is an exponential object can be discussed even though the exponential object does not exist, just by considering instead morphisms whose domain is a product. There is however then the problem of determining the morphisms whose domain is an exponential, and considering them to be the natural operators is in many contexts the smoothest and most "natural" thing to do. Experts on recursive functions or C^∞ functions between finite-dimensional manifolds may wish to consider the result of taking $\mathbf{C}$ to be these particular categories in the above considerations. They may also wish to consider whether the fixed-point theorem of section one has any applications in those cases.

3. Presentation-free formulations of satisfaction, truth, and provability according to Gödel and Tarski; representability vs. definability

In order to apply the theorem of the previous section to obtain Tarski's theorem concerning the impossibility of defining truth for a theory within the theory itself, we first note briefly how a theory gives rise to a category $\mathbf{C}$ with finite products. Consider two objects A, 2 and let the $\mathbf{C}$-morphisms be equivalence classes of (tuples of) formulas or terms of the theory, where two formulas (or terms) are considered equivalent iff their logical equivalence (or equality) is provable in the theory. Thus the morphisms $1 \longrightarrow A$ are (classes of) constant terms, the morphisms $A \times A \longrightarrow A$ are (classes of) terms with two free variables, while morphisms $A^n \longrightarrow 2$ are (classes of) formulas with n free variables so that in particular morphisms $1 \longrightarrow 2$ are (classes of) sentences of the theory. In particular there is a morphism true : $1 \longrightarrow 2$ corresponding to the class of sentences provable in the theory and similarly a morphism false : $1 \longrightarrow 2$ corresponding to the class of sentences whose negation is provable in the theory. Morphisms $2^n \longrightarrow 2$ would include all propositional operations, but we will make no use of that except for the following case:

> If the theory is consistent there is a morphism not : $2 \longrightarrow 2$ such that $\varphi\,\text{not} \neq \varphi$ for all morphisms $\varphi : 1 \longrightarrow 2$.

In particular we will not need to use the fact that $2 = 1 + 1$, although that determines the nature of those hom-sets not explicitly spelled out above. Defining composition to correspond to substitution (for example a constant $a : 1 \longrightarrow A$ composed with a unary formula $\varphi : A \longrightarrow 2$ composed with not gives the sentence $a\varphi\,\mathrm{not} : 1 \longrightarrow 2$, etc.) we get a category $\mathbf{C}$ with finite products which might be called the Lindenbaum category of the theory. Models of the theory can then be viewed as certain functors $\mathbf{C} \longrightarrow \mathcal{S}$. We make no use here of the operation in $\mathbf{C}$ induced by quantification in the theory, but the categorical description of this operation will be clear to readers of the two papers cited in the introduction. In our construction above of $\mathbf{C}$ we have tacitly assumed that the theory was a first order single-sorted one, in which case all objects of $\mathbf{C}$ are isomorphic to those of the form $A^n \times 2^m$, but with trivial modification we could have started with a higher-order or several-sorted theory with no change of any significance to the arguments below. To make one point somewhat more explicit note that the projection morphisms $A^n \longrightarrow A$ arise from the variables of the theory.

We then say that *satisfaction is definable* in the theory iff there is a binary formula $\mathrm{sat} : A \times A \longrightarrow 2$ in $\mathbf{C}$ such that for every unary formula $\varphi : A \longrightarrow 2$ there is a constant $c : 1 \longrightarrow A$ such that for every constant a the following diagram commutes in $\mathbf{C}$

$$\begin{tikzcd}
1 \arrow[r, "a"] \arrow[d, "\langle a,c\rangle"'] & A \arrow[d, "\varphi"] \\
A \times A \arrow[r, "\mathrm{sat}"'] & 2
\end{tikzcd}$$

Here we imagine taking for c a Gödel number for (one of the representatives of) φ. The condition would traditionally be expressed by requiring that the sentence

$$a\,\mathrm{sat}\,c \iff a\varphi$$

be provable in the theory, but if $\mathbf{C}$ arises from our construction of the Lindenbaum category this amounts to the same thing.

Combining the above notion with our remark about the meaning of consistency and the theorem of the previous section we have immediately the

3.1. Corollary. *If satisfaction is definable in the theory then the theory is not consistent.*

In order to show that Truth cannot be defined we first need to say what Truth would mean; that seems to require some further assumptions on the theory, which are however often realizable. Namely we suppose that there is a binary term

$$A \times A \xrightarrow{\mathrm{subst}} A$$

in $\mathbf{C}$ and a ("metamathematical") binary relation

$$\Gamma \subseteq \mathbf{C}(1, A) \times \mathbf{C}(1, 2)$$

between constants and sentences for which the following holds.

1) For all $\varphi : A \longrightarrow 2$ there is $c : 1 \longrightarrow A$ such that for all $a : 1 \longrightarrow A$

$$(a\,\mathrm{subst}\,c)\Gamma(a\varphi)$$

For example we could imagine that $d\Gamma\sigma$ means that d is the Gödel number of some one of the sentences that represent σ, and that subst is a binary operation which, when applied to a constant a and to a constant c that happens to be the Gödel number of a unary formula φ, yields the Gödel number of the sentence obtained by substituting a into φ.

Given a binary relation $\Gamma \subseteq \mathbf{C}(1, A) \times \mathbf{C}(1, 2)$ we say that *Truth* (of sentences) *is definable* in the theory (relative to Γ) provided there is a unary formula $\mathrm{Truth} : A \longrightarrow 2$ such that

2) For all $\sigma : 1 \longrightarrow 2$ and $d : 1 \longrightarrow A$, if $d\Gamma\sigma$ then $d\mathrm{Truth} = \sigma$.

Again the traditional formulation would require that the sentence

$$\ulcorner\sigma\urcorner\mathrm{Truth} \Longleftrightarrow \sigma, \qquad \text{for } \ulcorner\sigma\urcorner\Gamma\sigma$$

be provable, but in the Lindenbaum category this just amounts to the equation $\ulcorner\sigma\urcorner\mathrm{Truth} = \sigma$.

3.2. THEOREM. *If the theory is consistent and substitution is definable relative to a given binary relation Γ between constants and sentences, then Truth is not definable relative to the same binary relation.*

PROOF. If both 1) and 2) hold then the diagram

$$\begin{tikzcd}
1 \arrow[r, "a"] \arrow[d, "\langle a,c\rangle"'] \arrow[dr, "d"] & A \arrow[dr, "\varphi"] & \\
A \times A \arrow[r, "\mathrm{subst}"'] & A \arrow[r, "\mathrm{Truth}"'] & 2
\end{tikzcd}$$

shows that

$$\begin{tikzcd}
A \times A \arrow[r, "\mathrm{subst}"] \arrow[dr, "\mathrm{sat}"'] & A \arrow[d, "\mathrm{Truth}"] \\
& 2
\end{tikzcd}$$

is a definition of satisfaction, contradicting the previous result. ∎

We will also prove an “incompleteness theorem”, using the notion of a Provability predicate. Given a binary relation Γ between constants and sentences, we say that *Provability is representable in the theory* iff there is a unary formula $\mathrm{Pr} : A \longrightarrow 2$ such that

3) Whenever $d\Gamma\sigma$ then $d\mathrm{Pr} = \mathrm{true}$ iff $\sigma = \mathrm{true}$.

3.3. Theorem. *Suppose that for a given binary relation Γ between constants and sentences of $\mathbf{C}$, substitution is definable and Provability is representable. Then the theory is not complete if it is consistent.*

Proof. Suppose on the contrary that $\mathbf{C}(1,2) = \{\text{false, true}\}$. Our notion of consistency implies that false $\neq$ true. Condition 3) states that for $d\Gamma\sigma$

a) $\sigma = \text{true}$ implies $d\mathrm{Pr} = \text{true}$

b) $\sigma \neq \text{true}$ implies $d\mathrm{Pr} \neq \text{true}$

By completeness b) implies

b') $\sigma = \text{false}$ implies $d\mathrm{Pr} = \text{false}$

But a) and b') together with completeness mean that whenever $d\Gamma\sigma$,

$$\begin{tikzcd} 1 \arrow[d, "d"'] \arrow[dr, "\sigma"] & \\ A \arrow[r, "\mathrm{Pr}"'] & 2 \end{tikzcd}$$

is commutative, i.e. that Pr satisfies condition 2) for a Truth-definition, which by our previous theorem yields a contradiction. ∎

Note: Our proposition in section 2. can be interpreted as a fragment of a general theory developed by Eilenberg and Kelly from an idea of Spanier.

Appendix

Shortly after this article was published I realized that the relation Γ used in section 3. is actually superfluous for the purpose at hand. A mere existential condition suffices because the trace of a specific Gödel numbering is not required. The essential content of the contrast between provability (attainable) and a truth definition (unattainable) can as well be expressed simply by requiring less of a given map of the satisfaction type: for every unary formula ϕ there should exist a number c that B-represents it in the sense that for every number a the satisfaction map gives the result of substituting a into ϕ

$$a \text{ sat } c = a\phi$$

but only in case both sides of the equation are in the specified subset $B \subseteq \mathbf{C}(1,2)$. The usual meaning of representability and provability is thus expressed if B contains at least the two elements true and false; by contrast, completeness is represented if $B = \mathbf{C}(1,2)$, in other words, if all nullary formulas are in B. However, if the present point of view is to be extended to include also Gödel's second incompleteness theorem, a specific relation between formulas and their numbers may be required.

Mathematics Department, State University of New York
244 Mathematics Building, Buffalo, N.Y. 14260-2900 USA
Email: wlawvere@buffalo.edu

This article may be accessed from http://www.tac.mta.ca/tac/reprints or by anonymous ftp at
ftp://ftp.tac.mta.ca/pub/tac/html/tac/reprints/articles/15/tr15.{dvi,ps}

Reprints in Theory and Applications of Categories, No. 11, 2005, pp. 1–35.

AN ELEMENTARY THEORY OF THE CATEGORY OF SETS (LONG VERSION) WITH COMMENTARY

F. WILLIAM LAWVERE

Received by the editors 2005-04-01.
Transmitted by M. Hyland, A. Kock, R. Rosebrugh. Reprint published on 2005-05-23.
2000 Mathematics Subject Classification: 18B05, 00A30, 03A05.
Key words and phrases: Category of sets, Axiom of choice, Mathematical logic and foundations.
This article is an expanded version of 'An elementary theory of the category of sets', *Proceedings of the National Academy of Science of the U.S.A* **52**, 1506–1511.

ETCS and the Philosophy of Mathematics

Colin McLarty

Philosophers and logicians to this day often contrast "categorical" foundations for mathematics with "set-theoretic" foundations as if the two were opposites. Yet the second categorical foundation ever worked out, and the first in print, was a set theory—Lawvere's axioms for the category of sets, called ETCS, (Lawvere 1964). These axioms were written soon after Lawvere's dissertation sketched the category of categories as a foundation, CCAF, (Lawvere 1963). They appeared in the *PNAS* two years before axioms for CCAF were published (Lawvere 1966). The present longer version was available since April 1965 in the Lecture Notes Series of the University of Chicago Department of Mathematics.[1] It gives the same definitions and theorems, with the same numbering as the 5 page PNAS version, but with fuller proofs and explications.

Lawvere argued that set theory should not be based on membership (as in Zermelo Frankel set theory, ZF), but on "isomorphism-invariant structure, as defined, for example, by universal mapping properties" (p. 1). He later noticed that Cantor and Zermelo differed over this very issue. Cantor gave an isomorphism-invariant account of sets, where elements of sets are "mere units" distinct from one another but not individually identifiable. Zermelo sharply faulted him for this and followed Frege in saying set theory must be founded on a membership relation between sets.[2]

Paul Benacerraf made the question prominent for philosophers one year after (Lawvere 1964). The two were no doubt independent since philosophers would not look in the *PNAS* for this kind of thing. Benacerraf argued that numbers, for example, cannot be sets since numbers should have no properties except arithmetic relations. The set theory familiar to him was ZF, where the elements of sets are sets in turn, and have properties other than arithmetic. So he concluded that numbers cannot be sets and the natural number structure cannot be a set.

Benacerraf wanted numbers to be elements of *abstract structures* which differ from ZF sets this way:

> in giving the properties (that is, necessary and sufficient) of numbers you merely characterize an *abstract structure*—and the distinction lies in the fact that the "elements" of the structure have no properties other than those relating them to other "elements" of the same structure. (Benacerraf 1965, p. 70)

[1] Date verified by Lawvere through the University of Chicago math librarian.

[2] Zermelo's complaint about Cantor and endorsement of Frege are on (Cantor 1932, pp. 351, 440-42). See (Lawvere 1994).

The sets of ETCS are *abstract structures* in exactly this sense. An element $x \in S$ in ETCS has no properties except that it is an element of S and is distinct from any other elements of S. The *natural number structure* in ETCS is a triad of a set N, a selected element $0 \in N$, and a successor function $s{:}N{\to}N$. Then $N, 0, s$ expresses the arithmetic relations as for example $m = s(n)$ says m is the successor of n. But $N, 0, s$ simply have no properties beyond those they share with every isomorphic triad.[3] So in relation to 0 and s the elements of N have arithmetic relations, but they have no other properties.

Lawvere carefully says ETCS "provides a foundation for mathematics ... in the sense that much of number theory, elementary analysis, and algebra can apparently be developed within it." This aspect relies entirely on the elementary axioms. But he did not claim that such formal adequacy makes ETCS a formalist or logicist starting point for mathematics, as if there was no mathematics before it. He says that ETCS condenses and systematizes knowledge we already have of "the category of sets and mappings... denoted by $\mathcal{S}$" (pp. 1-2). This aspect deals with models and metatheorems for the axioms. Gödel's theorem says we will never have a proof-theoretically complete description of this category. Lawvere gives various metatheoretic proofs about categorical set theory based on this and on the Löwenheim-Skolem theorem. Yet he takes it that there are sets, we know much about them apart from any formalization, and we can learn much from seeing how formal axioms describe them. The category exists objectively in mathematical experience as a whole—not in a platonic heaven, nor in merely subjective individual experience. This historical (and dialectical) realism continues through all his work and gets clear recent statements in (Lawvere & Rosebrugh 2003) and (Lawvere 2003).

Of course the category of sets is not the only one that exists. It is not the only one formally adequate as a "foundation" nor the first to be described that way. Lawvere concludes this paper, as he also does the *PNAS* version, by saying that more powerful foundations would be more naturally expressed in the category of categories (p.32, or (Lawvere 1964, p. 1510)).

References.

Benacerraf, Paul (1965), 'What numbers could not be', *Philosophical Review* **74**, 47–73.

Cantor, Georg (1932), *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Julius Springer, Berlin.

Lawvere, F. William (1963), Functorial Semantics of Algebraic Theories, PhD thesis, Columbia University. Republished in: *Reprints in Theory and Applications of Categories*, No. 5 (2004) pp 1-121, available on-line at http://tac.mta.ca/tac/reprints/articles/5/tr5abs.html.

Lawvere, F. William (1964), 'An elementary theory of the category of sets', *Proceedings of the National Academy of Science of the U.S.A* **52**, 1506–1511.

[3]For a precise statement and proof see (McLarty 1993).

Lawvere, F. William (1966), The category of categories as a foundation for mathematics, *in* S.Eilenberg et al., eds, 'Proceedings of the Conference on Categorical Algebra, La Jolla, 1965', Springer-Verlag, pp. 1–21.

Lawvere, F. William (1994), 'Cohesive toposes and Cantor's "lauter Einsen"', *Philosophia Mathematica* **2**, 5–15.

Lawvere, F. William (2003), 'Foundations and applications: Axiomatization and education', *Bulletin of Symbolic Logic* **9**, 213–24.

Lawvere, F. William & Robert Rosebrugh (2003), *Sets for Mathematics*, Cambridge University Press.

McLarty, Colin (1993), 'Numbers can be just what they have to', *Noûs* **27**, 487–98.

Author Commentary

F. William Lawvere

Saunders Mac Lane wisely insisted on this more complete version of the work that was summarized in the brief article he communicated in October 1964 to the Proceedings of the National Academy of Science. Accordingly, this manuscript, with its relatively complete proofs and extensive remarks on the metamathematical status of the work, was deposited in the University of Chicago mathematics library at the end of April 1965. I hope that the present publication, made possible thanks to Colin McLarty's efforts in TEX and thanks to the editors of *TAC*, will assist those who wish to apply the Elementary Theory of the Category of Sets to philosophy and to teaching.

Of course, the central theory is more fully developed in the joint textbook with Robert Rosebrugh *Sets for Mathematics* (and even bolder metamathematical comments are offered in its appendix). However, the present document may still be of interest, at least historically, giving as it does a glimpse of how these issues looked even as the crucial need for elementary axiomatization of the theory of toposes was crystallizing. (In July 1964 I noted in Chicago that sheaves of sets form categories with intrinsic properties just as sheaves of abelian groups form abelian categories; in January 1965 Benabou told me that Giraud had axiomatized "toposes"; in June 1965 Verdier's lecture on the beach at LaJolla made clear that those axioms were not yet "elementary" though extremely interesting....).

Among mathematical/logical results in this work is the fact that the axiom of choice implies classical logic. That result was later proved by Diaconescu under the strong hypothesis that the ambient category is a topos (i.e. that the notion of arbitrary monomorphism is representable by characteristic functions) but in this work it is shown that the topos property is itself also implied by choice. Specifically, characteristic functions valued in 2 are constructed by forming the union (proved to exist) of all complemented subobjects that miss a given monomorphism, and then using the choice principle to show that there is no excluded middle. (Presumably with modern technology this calculation could be carried out with more general generators than 1).

This elementary theory of the category of sets arose from a purely practical educational need. When I began teaching at Reed College in 1963, I was instructed that first-year analysis should emphasize foundations, with the usual formulas and applications of calculus being filled out in the second year. Since part of the difficulty in learning calculus stems from the rigid refusal of most textbooks to supply clear, explicit, statements of concepts and principles, I was very happy with the opportunity to oppose that unfortunate trend. Part of the summer of 1963 was devoted to designing a course based on the axiomatics of Zermelo-Fraenkel set theory (even though I had already before concluded that the category of categories is the best setting for "advanced" mathematics). But I soon realized that even an entire semester would not be adequate for explaining all the (for a beginner bizarre) membership-theoretic definitions and results, then translating them into operations usable in algebra and analysis, then using that framework to construct a basis for the material I planned to present in the second semester on metric spaces.

However I found a way out of the ZF impasse and the able Reed students could indeed be led to take advantage of the second semester that I had planned. The way was to present in a couple of months an explicit axiomatic theory of the mathematical operations and concepts (composition, functionals, etc.) as actually needed in the development of the mathematics. Later, at the ETH in Zurich, I was able to further simplify the list of axioms.

April 1, 2005

An Elementary Theory of the Category of Sets (Long Version)

F. William Lawvere [4]

The elementary theory presented in this paper is intended to accomplish two purposes. First, the theory characterizes the category of sets and mappings as an abstract category in the sense that any model for the axioms which satisfies the additional (non-elementary) axiom of completeness (in the usual sense of category theory) can be proved to be equivalent to $\mathcal{S}$. Second, the theory provides a foundation for mathematics which is quite different from the usual set theories in the sense that much of number theory, elementary analysis, and algebra can apparently be developed within it even though no relation with the usual properties of $\in$ can be defined.

Philosophically, it may be said that these developments partially support the thesis that even in set theory and elementary mathematics it is also true as has long been felt in advanced algebra and topology, namely that the substance of mathematics resides not in Substance (as it is made to seem when $\in$ is the irreducible predicate, with the accompanying necessity of defining all concepts in terms of a rigid elementhood relation) but in Form (as is clear when the guiding notion is isomorphism-invariant structure, as defined, for example, by universal mapping properties). As in algebra and topology, here again the concrete technical machinery for the precise expression and efficient handling of these ideas is provided by the Eilenberg-Mac Lane theory of categories, functors, and natural transformations.

The undefined terms of our theory are *mappings*, *domain*, *co-domain*, and *composition*. The first of these is merely a convenient name for the elements of the universe over which all quantifiers range, the second two are binary relations in that universe, and the last is a ternary relation. The heuristic intent of these notions may be briefly explained as follows: a *mapping* f consists of three parts, a set S, a set S', and a "rule" which assigns to every element of S exactly one element of S'. The identity rule on S to S determines the *domain* of f and the identity rule on S' to S' determines the *codomain* of f. Thus every mapping has a unique domain and codomain, and these are also mappings. Mappings which appear as domains or codomains are also called *objects* and are frequently denoted by capital letters to distinguish them from more general mappings. If A, f, B are mappings, and if A is the domain of f while B is the codomain of f, we denote this situation by $A \xrightarrow{f} B$.

[4]Work partially supported by AF Contract No. AFOSR-520-64 at the University of Chicago and partially by a NATO Postdoctoral Fellowship at the ETH, Zurich.

A binary operation called *composition* is defined in the obvious way for precisely those pairs of mappings f, g such that the codomain of f is the domain of g. The result of such composition is denoted by juxtaposition fg, and we have

$$A \xrightarrow{f} B \xrightarrow{g} C \qquad \Rightarrow \qquad A \xrightarrow{fg} C \ .$$

It is clear that composition is associative insofar as it is defined, and that objects behave as neutral elements with respect to composition.

The heuristic class of all mappings, with the above described domain, codomain, and composition structure, will be denoted by $\mathcal{S}$. The axioms of our theory may be regarded as a listing of some of the basic and mathematically useful properties of $\mathcal{S}$. There are three groups of axioms and two special axioms.

The first group of axioms (to which we do not assign numbers) consists of the axioms for an abstract *category* in the sense of (Eilenberg and Mac Lane 1945). We take these axioms as known but note that these axioms are obviously elementary (first-order) and that they have been stated informally in the third paragraph. We remark that a map $A \xrightarrow{f} B$ in a category is called an *isomorphism* if there exists a map $B \xrightarrow{g} A$ in the category such that $fg = A$ and $gf = B$; g is then a uniquely determined isomorphism called the *inverse* of f.

The second group of axioms calls for the existence of solutions to certain "universal mapping problems". As usual in such situations, it will be clear that the objects and operations thus asserted to exist are actually unique up to an isomorphism which is itself uniquely determined by a "naturalness" condition suggested by the structure of the particular axiom. This naturalness condition is simply that the isomorphism in question commutes with all the "structural maps" occurring in the statement of the universal mapping problem.

Axiom 1. All finite roots exist. Explicitly, this is guaranteed by assuming, in the sense explained below, that a terminal object 1 and an initial object 0 exist; that the product $A \times B$ and the coproduct (sum) $A + B$ of any pair of objects exists; and that the equalizer $E \xrightarrow{k} A$ and the coequalizer $B \xrightarrow{q} E^*$ of any pair $A \underset{g}{\overset{f}{\rightrightarrows}} B$ of maps exist.

Axiom 2. The exponential B^A of any pair of objects exists. The defining universal mapping property of exponentiation, discussed below, is closely related to the concept of λ-conversion and requires, in essence, that for each object A, the functor $(\ )^A$ is co-adjoint to the functor $A \times (\ )$.

Axiom 3. There exists a Dedekind-Pierce object N. This plays the role of our axiom of infinity; the defining property of the natural numbers is here taken to be the existence and uniqueness of sequences defined by a very simple sort of recursion, as explained below.

We now explain in more detail the universal properties associated with the terminology occurring in the existential axioms above. A *terminal* object is an object 1 such that for any object A there is a unique mapping $A \to 1$. An *initial* object is an object 0 such that for any object B there is a unique mapping $0 \to B$. It is clear that in our heuristic picture of $\mathcal{S}$, any singleton set is a terminal object and the null set is an initial object. (More precisely, we refer here to the identity mappings on these sets.) Because nothing of substance that we wish to do in our theory depends on the size of the isomorphism classes of objects, we find it a notational convenience to assume that there is exactly one terminal object 1 and exactly one initial object 0; however, we do not make this convention a formal axiom as it can be avoided by complicating the notation somewhat.

Definition 1. x is an *element* of A, denoted $x \in A$, iff $1 \xrightarrow{x} A$.

Remark 1: Elementhood, as defined here, is a special case of membership, to be defined presently. Neither of these notions has many formal properties in common with the basic relation of the usual set theories; $x \in y \in z$, for example, never holds except in trivial cases. However, we are able (noting that the evaluation of a mapping at an element of its domain may be viewed as a special case of composition) to deduce the basic property required by our heuristic definition of mapping, namely that if $A \xrightarrow{f} B$ then

$$\forall x \; [x \in A \;\Rightarrow\; \exists! y \; [y \in B \;\&\; xf = y]]$$

We must still describe the universal mapping properties of product, coproduct, equalizer and coequalizer in order to explicate our first axiom. Given two objects A and B their *product* is an object $A \times B$, together with a pair of mappings

$$\begin{tikzcd} & A \\ A \times B \arrow[ur, "p_A"] \arrow[dr, "p_B"'] & \\ & B \end{tikzcd}$$

such that for any object X and pair of mappings

$$\begin{tikzcd} & A \\ X \arrow[ur, "f_A"] \arrow[dr, "f_B"'] & \\ & B \end{tikzcd}$$

there is a unique mapping $X \xrightarrow{h} A \times B$ such that $hp_A = f_A$ and $hp_B = f_B$, i.e. such that the following diagram commutes:

$$\begin{tikzcd}
& & A \\
X \arrow[urr, "f_A"] \arrow[r, dashed, "h"] \arrow[drr, "f_B"'] & A \times B \arrow[ur, "p_A"'] \arrow[dr, "p_B"] & \\
& & B
\end{tikzcd}$$

We use the notation $h = \langle f_A, f_B \rangle$. Thus taking $X = 1$, we deduce that the elements x of $A \times B$ are uniquely expressible in the form $x = \langle x_A, x_B \rangle$ where $x_A \in A$ and $x_B \in B$. Returning to consideration of an arbitrary X and pair of maps f_A, f_B it then follows that the "rule" of $\langle f_A, f_B \rangle$ is given by

$$x\langle f_A, f_B \rangle \;=\; \langle x f_A, x f_B \rangle \qquad \text{for all } x \in X.$$

The "structural maps" p_A and p_B are called *projections.* There are unique natural (commuting with projections) isomorphisms

$$A \times 1 \;\cong\; A \;\cong 1 \times A$$
$$(A \times B) \times (C \times D) \;\cong\; (A \times C) \times (B \times D)\ .$$

Given mappings $A \xrightarrow{f} A'$, $B \xrightarrow{g} B'$, there is a unique natural mapping

$$A \times B \xrightarrow{\;f \times g\;} A' \times B'$$

and this product operation on mappings is functorial in the sense that given further mappings $A' \xrightarrow{\overline{f}} A''$, $B' \xrightarrow{\overline{g}} B''$, one has

$$(f\,\overline{f}) \times (g\,\overline{g}) \;=\; (f \times g)(\overline{f} \times \overline{g})$$

The *coproduct* $A + B$ of two objects is characterized by the dual universal mapping property

$$\begin{tikzcd}
A \arrow[dr, "i_A"'] \arrow[drr, "f_A"] & & \\
& A + B \arrow[r, dashed, "h"] & X \\
B \arrow[ur, "i_B"] \arrow[urr, "f_B"'] & &
\end{tikzcd}$$

$\forall X\ \forall f_A\ \forall f_B\ \exists! h\ [i_A h \;=\; f_A\ \&\ i_B h \;=\; f_B]$

(where the quantifiers range only over diagrams of the above form). Here the structural maps i_A and i_B are called *injections*. In the usual discussions of $\mathcal{S}$ and hence in our theory, the coproduct is called the *sum*. The functorial extension of this operation from objects to all mappings follows just as in the case of the product operation, but the deduction of the nature of the elements of $A+B$ is not so easy; in fact we will find it necessary to introduce further axioms to describe the elements of a sum. This is not surprising in view of the great variation from category to category in the nature of coproducts. Suffice it for the moment to note that the existence of sums in $\mathcal{S}$ can be verified by one of the usual "disjoint union" constructions; the universal mapping property corresponds to the heuristic principle that a mapping is well defined by specifying its rule separately on each piece of a finite partition of its domain.

The *equalizer* of a pair of mappings $A \underset{g}{\overset{f}{\rightrightarrows}} B$ is a mapping $E \xrightarrow{k} A$ such that $kf = kg$ and which satisfies the universal mapping property: for any $X \xrightarrow{u} A$ such that $uf = ug$ there is a unique $X \xrightarrow{z} E$ such that $u = zk$. It follows immediately that the elements a of A such that $af = ag$ are precisely those which factor through k, and the latter are in one-to-one correspondence with the elements of E. It also follows that k is, in a sense to be defined presently, a subset of A; clearly it is just the subset on which f and g agree. The equalizer construction, like the coequalizer to be described next, has certain functorial properties which we do not bother to state explicitly.

The *coequalizer* of a pair of mappings $A \underset{g}{\overset{f}{\rightrightarrows}} B$ is a mapping $B \xrightarrow{q} E^*$ such that $fq = gq$ and which satisfies the dual universal mapping property:

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r, "q"] \arrow[dr, "u"'] & E^* \arrow[d, dashed, "z"] \\ & & X \end{tikzcd}$$

$$\forall X\ \forall u\ [fu = gu\ \Rightarrow\ \exists!\ z\ [u = qz]\ ]$$

(where the quantified variables range only over diagrams of the sort pictured). As in the case of sums, an elementwise investigation of coequalizers is not immediate; the discussion of the elements of E^* can be reduced to the consideration of the pairs of elements of B which are identified by q, but the latter is non-trivial. However, using our full list of axioms we will be able to prove a theorem to the effect that q is the quotient map obtained by partitioning B in the finest way which, for each $a \in A$, puts af and ag in the same cell of the partition.

This completes the explication of our axiom "finite roots exist" except for the remark that in category theory "finite roots" refers to a more general class of operations including inverse images and intersections and their duals as well as more complicated constructions.

However, all these can be proved to exist (within the first order language) once the existence of the six special cases we have assumed is affirmed.

Before proceeding to the discussion of the exponentiation axiom, let us note that the usual formal definition of mapping can be dualized. The usual notion of a mapping f from A to B describes f as a subset of $A \times B$, namely the *graph* g of f in the diagram below.

$$\begin{tikzcd}
& & A \\
A \arrow[urr, "A"] \arrow[r, "g"] \arrow[drr, "f"'] & A \times B \arrow[ur, "p_A"'] \arrow[dr, "p_B"] & \\
& & B
\end{tikzcd}$$

Dualizing the diagram, we obtain the *cograph* g^* of f defined to make the following diagram commutative.

$$\begin{tikzcd}
A \arrow[dr, "i_A"'] \arrow[drr, "f"] & & \\
& A + B \arrow[r, dashed, "g^*"] & B \\
B \arrow[ur, "i_B"] \arrow[urr, "B"'] & &
\end{tikzcd}$$

This suggests the alternative concept of a mapping from A to B as a partition of the disjoint union with the property that each cell in the partition contains exactly one member of B. The concept of cograph corresponds to the intuitive diagrams sometimes used to represent mappings in which both A and B are displayed (disjointly!) and a line is drawn from each member of A to its image in B; two members of $A + B$ can then be thought of as identified if they can be connected by following these lines.

Whatever formal definition is taken, it is intuitively clear that for each pair of sets A, B there exists a set whose elements are precisely the mappings from A to B or at least "names" for these mappings in some sense. This is one of the consequences of our axiom that "exponentiation exists".

The meaning of the exponentiation axiom is that the operation of forming the product has a co-adjoint operation in the sense that given any pair of objects A, B there exists an object B^A and a mapping $A \times B^A \xrightarrow{e} B$ with the property that for any object X and any mapping $A \times X \xrightarrow{f} B$ there is a unique mapping $X \xrightarrow{h} B^A$ such that $(A \times h)e = f$. This universal mapping property is partly expressed by the following diagram.

$$\begin{tikzcd} A \times X \arrow[d, dashed, "A \times h"'] \arrow[dr, "f"] & \\ A \times B^A \arrow[r, "e"'] & B \end{tikzcd}$$

Here the structural mapping e is called the *evaluation* map. By taking $X = 1$, and noting that the projection $A \times 1 \to A$ is an isomorphism, it follows immediately that the elements of B^A are in one-to-one correspondence with the mappings from A to B. Explicitly, a mapping $A \xrightarrow{f} B$ and its "name" $[f] \in B^A$ are connected by the commutativity of the following diagram.

$$\begin{tikzcd} A \times 1 \arrow[r, "p_A"] \arrow[d, "A \times [f]"'] & A \arrow[d, "f"] \\ A \times B^A \arrow[r, "e"'] & B \end{tikzcd}$$

It follows immediately that for $a \in A$, $A \xrightarrow{f} B$, one has

$$\langle a, [f] \rangle e \;=\; af \;,$$

i.e., the rule of the evaluation map is evaluation.

When extended to a functor in the natural (i.e., coherent with evaluation) fashion, exponentiation is actually contravariant in the exponent. That is, given $A' \xrightarrow{f} A$, $B \xrightarrow{g} B'$, the induced map goes in the direction opposite of f:

$$B^A \xrightarrow{\;g^f\;} B'^{A'}$$

Functoriality in this case just means that if $A'' \xrightarrow{\overline{f}} A'$, $B' \xrightarrow{\overline{g}} B''$ are further mappings then $(g\overline{g})^{(\overline{f}f)} = (g^f)(\overline{g}^{\overline{f}})$. The rule of g^f is of course that for $A \xrightarrow{u} B$,

$$[u](g^f) \;=\; [fug] \;.$$

A simple example of such an induced map is the "diagonal" $B \to B^A$ which is induced by $A \to 1$ and $B \xrightarrow{B} B$ and which assigns to each $b \in B$ the "name" of the corresponding constant mapping $A \to B$.

Also deducible from the exponentiation axiom is the existence, for any three objects A, B, C, of a mapping

$$B^A \times C^B \xrightarrow{\;\gamma\;} C^A$$

such that for any $A \xrightarrow{f} B$, $B \xrightarrow{g} C$ one has

$$\langle [f], [g] \rangle \gamma \ = \ [fg]$$

(Of course the usual "laws of exponents" can be proved as well.)

Another consequence of the exponentiation axiom is the distributivity relation

$$A \times B \ + \ A \times C \ \cong \ A \times (B + C)$$

Actually a canonical mapping from the left hand side to the right hand side can be constructed in any category in which products and coproducts exist; in a category with exponentiation the inverse mapping can also be constructed. A related conclusion is that if q is the coequalizer of f with g, then $A \times q$ is the coequalizer of $A \times f$ with $A \times g$. Both these conclusions are special cases of the fact that for each A the functor $A \times (\ )$, like any functor with a co-adjoint, must preserve all right hand roots. We remark that in the usual categories of groups, rings, or modules, exponentiation does not exist, which follows immediately from the fact that distributivity fails.

The third axiom asserts the existence of a special object N equipped with structural maps $1 \xrightarrow{0} N \xrightarrow{s} N$ with the property that given any object X and any $x_0 \in X$, $X \xrightarrow{t} X$ there is a unique $N \xrightarrow{x} X$ such that the following diagram commutes:

$$\begin{tikzcd}
 & N \arrow[r, "s"] \arrow[dd, "x"] & N \arrow[dd, "x"] \\
1 \arrow[ur, "0"] \arrow[dr, "x_0"'] & & \\
 & X \arrow[r, "t"] & X
\end{tikzcd}$$

As usual, the universal mapping property determines the triple $N, 0, s$ up to a unique isomorphism. A mapping $N \xrightarrow{x} X$ is called a *sequence* in X, and the sequence asserted to exist in the axiom is said to be defined by simple recursion with the starting value x_0 and the transition rule t. Most of Peano's postulates could now be proved to hold for $N, 0, s$ but we delay this until all our axioms have been stated because, for example, the existence of an object with more than one element is needed to conclude that N is infinite. However we can now prove the

THEOREM 1 (Primitive Recursion). Given a pair of mappings

$$A \xrightarrow{f_0} B$$
$$N \times A \times B \xrightarrow{u} B$$

there is a mapping

$$N \times A \xrightarrow{f} B$$

such that for any $n \in N$, $a \in A$ one has

$$\begin{aligned} \langle 0, a\rangle f &= af_0 \\ \langle ns, a\rangle f &= \langle n, a, \langle n, a\rangle f\rangle u \end{aligned}$$

(It will follow from the next axiom that f is uniquely determined by these conditions.)

Proof. In order to understand the idea of the proof, notice that primitive recursion is more complicated than simple recursion in essentially two ways. First, the values of f depend not only on $n \in N$ but also on members of the parameter object A. However, the exponentiation axiom enables us to reduce this problem to a simpler one involving a sequence whose values are functions. That is, the assertion of the theorem is equivalent to the existence of a sequence

$$N \xrightarrow{y} B^A$$

such that $0y = [f_0]$ and such that for any $n \in N,\ a \in A$, one has

$$\langle a, (ns)y\rangle e = \langle n, a, \langle a, ny\rangle e\rangle u$$

where e is the evaluation map for B^A.

The second way in which primitive recursion (and the equivalent problem just stated) is more complicated than the simple recursion of our axiom is that the transition rule u depends not only on the value calculated at the previous step of the recursion but also on the number of steps that have been taken. However, this can be accomplished by first constructing by simple recursion a sequence whose values are ordered pairs in which the first coordinate is used just to keep track of the number of steps. Thus the proof of this theorem consists of constructing a certain sequence

$$N \xrightarrow{x} N \times B^A$$

and then defining y as required above to be simply x followed by projection onto the second factor. That is, we define by simple recursion the *graph* of y.

Explicitly, the conditions on the sequence x are that

$$0x = \langle 0, [f_0]\rangle 0$$

and that for each $n \in N$ and $a \in A$

$$\begin{aligned} (ns)xp_N &= ns \\ \langle a, (ns)xp_{B^A}\rangle e &= \langle n, a, \langle a, nxp_{B^A}\rangle e\rangle u \end{aligned}$$

By our axiom, the existence of such x is guaranteed by the existence of a mapping

$$N \times B^A \xrightarrow{t} N \times B^A$$

such that for any $n \in N$, $h \in B^A$, $a \in A$

$$\begin{aligned} \langle n, h\rangle t p_N &= ns \\ \langle a, \langle n, h\rangle t p_{B^A}\rangle e &= \langle n, a, \langle a, h\rangle e\rangle u \end{aligned}$$

But such a t can be constructed by defining

$$t = \langle p_N s, t_2\rangle$$

where $N \times B^A \xrightarrow{t_2} B^A$ is the mapping corresponding by exponential adjointness to the composite

$$A \times N \times B^A \longrightarrow N \times A \times A \times B^A \longrightarrow N \times A \times B \xrightarrow{u} B$$

where the first is induced by the diagonal map $A \to A \times A$ and where the second is induced by the evaluation map $A \times B^A \to B$. (Here we have suppressed mention of certain commutativity and associativity isomorphisms between products.) This completes the proof of the Primitive Recursion theorem. ∎

Our first three axioms, requiring the existence of finite roots, exponentiation, and the Dedekind-Pierce object N, actually hold in certain fairly common categories other than $\mathcal{S}$, such as the category $\mathcal{C}$ of small categories and functors or any functor category $\mathcal{S}^{\mathbb{C}}$ for a fixed small category $\mathbb{C}$ (examples of the latter include the category $\mathcal{S}^{\mathbb{G}}$ whose objects are all permutation representations of a given group $\mathbb{G}$ and whose maps are equivariant maps as well as the category $\mathcal{S}^2$ whose objects correspond to the mappings in $\mathcal{S}$ and whose maps are commutative squares of mappings in $\mathcal{S}$). However, these categories do not satisfy our remaining axioms.

We state now our two special axioms

Axiom 4. 1 is a *generator*. That is, if $A \overset{f}{\underset{g}{\rightrightarrows}} B$ then

$$f \neq g \;\Rightarrow\; \exists a\, [a \in A \;\&\; af \neq ag]$$

In other words, two mappings are equal if they have the same domain and codomain and if they have the same value at each element of their domain. We mention the immediate consequence that if A has exactly one element then $A = 1$.

Axiom 5(Axiom of Choice). If the domain of f has elements, then there exists g such that $fgf = f$.

Remark 2: This axiom has many uses in our theory even at the elementary level, as will soon become apparent. That, in relation to the other axioms, the axiom of choice is stronger in our system than in the usual systems is indicated by the fact that the axiom of choice (as stated above) is obviously independent. Namely, the category $\mathcal{O}$ of partially ordered sets and order-preserving maps is a model for all our axioms (even including the final group of axioms not yet described) with the exception that the axiom of choice is false in $\mathcal{O}$. To see the latter we need only take f to be the "identity" map from a set with trivial partial ordering to the same set with a stronger ordering. Most of the basic theorems that we will prove fail in $\mathcal{O}$.

Proposition 1. 2 is a cogenerator.

Proof. Here we define $2 = 1+1$, and by the statement that 2 is a *cogenerator* we mean that if $A \overset{f}{\underset{g}{\rightrightarrows}} B$ are different then there is $B \xrightarrow{u} 2$ such that $fu \neq gu$. Since 1 is a generator we are immediately reduced to the case where $A = 1$ and f, g are elements of B. By the universal property of sums there is an induced map $2 \xrightarrow{h} B$ which by the axiom of choice has a "quasi-inverse" u

$$\begin{tikzcd}
1 \arrow[dr, "i_0"'] \arrow[drr, "f"] & & \\
& 2 \arrow[r, shift left, "h"] & B \arrow[l, dashed, shift left] \\
1 \arrow[ur, "i_1"] \arrow[urr, "g"'] & &
\end{tikzcd} \qquad huh = h$$

But this u clearly separates f, g if they are different, since

$$fu = gu \;\Rightarrow\; fuh = guh \;\Rightarrow\; i_0huh = i_1huh$$
$$\Rightarrow\; i_0h = i_1h \;\Rightarrow\; f = g$$

■

In order to state our next proposition, we must define subset, inclusion, and member.

Definition 2. a is a *subset* of A iff a is a monomorphism with codomain A.

Here a monomorphism a is a mapping such that

$$\forall b\, \forall b' \; [ba = b'a \;\Rightarrow\; b = b']$$

By the axiom of choice it is clear that in $\mathcal{S}$ (indeed in any model of our theory) every monomorphism, except for trivial cases, is a retract, i.e. has a right inverse. Also, since 1 is a generator, it suffices in verifying that a is a monomorphism to consider b, b' which are elements of the domain of a. Note that the subset is a itself, not just the domain of a. The domain of a is an object, and an object in $\mathcal{S}$ is essentially just a cardinal number, whereas the notion of subset contains more structure than just number. Namely, a subset of A

involves a specific way of identifying its members as elements of A, hence the mapping a.

Remark 3: A mapping $A \xrightarrow{f} B$ is an epimorphism iff

$$\forall y \in B \; \exists x \in A \; [xf = y] \; ,$$

a monomorphism iff

$$\forall x \in A \; \forall x' \in A \; [xf = x'f \;\; \Rightarrow \;\; x = x'] \; ,$$

an isomorphism iff it is both an epimorphism and a monomorphism. The proof (using the axiom of choice) is left to the reader. (The concept of epimorphism is defined dually to that of monomorphism.)

Definition 3. x is a member of a [notation $x \in a$] iff for some A (which will be unique) x is an element of A, a is a subset of A, and there exists $\overline{x}$ such that $\overline{x}a = x$.

$$\begin{tikzcd}
& 1 \arrow[dl, dashed, "\overline{x}"'] \arrow[dr, "x"] & \\
\bullet \arrow[rr, "a"'] & & A
\end{tikzcd}$$

Note that $\overline{x}$ is an element of the domain of a and is uniquely determined. Note also that x is an element of a iff x is a member of a and a is an object; the use of the same notation for the more general notion causes no ambiguity.

Definition 4. $a \subseteq b$ iff for some A, a and b are both subsets of A and there exists h such that $a = hb$.

$$\begin{tikzcd}
\bullet \arrow[rr, dashed, "h"] \arrow[dr, "a"'] & & \bullet \arrow[dl, "b"] \\
& A &
\end{tikzcd}$$

Note that h is a uniquely determined monomorphism if it exists. The inclusion relation thus defined for subsets of A is clearly reflexive and transitive. Also $x \in a$ iff $x \subseteq a$ and the domain of x is 1, and a is a subset of A iff $a \subseteq A$ and A is an object.

Proposition 2. Let a, b be subsets of A. Then $a \subseteq b$ iff

$$\forall x \in A \; [x \in a \;\; \Rightarrow x \in b]$$

(For the proof we assume that the domain of b has elements; however the next axiom makes this restriction unnecessary.)

Proof. If $a \subseteq b$ and $1 \xrightarrow{x} A$, then clearly $x \in a \;\; \Rightarrow \;\; x \in B$.

$$\begin{tikzcd}
& 1 \arrow[dl] \arrow[dr, dashed] & \\
\bullet \arrow[rr, "h"] \arrow[dr, "a"'] & & \bullet \arrow[dl, "b"] \\
& A &
\end{tikzcd}$$

Conversely, suppose $\forall x \in A\ [x \in a \;\Rightarrow\; x \in B]$. By the axiom of choice there is $A \xrightarrow{g} dom\ b$ such that $bgb = b$. Since b is a monomorphism, $bg = dom\ b$ (the identity mapping). We define $h = ag$, and attempt to show $a = hb$. It suffices to show that $\overline{x}a = \overline{x}agb$ for every element $\overline{x}$ of the domain of a. But given $\overline{x}$, then $x = \overline{x}a$ is an element of A for which $x \in a$, so by hypothesis $x \in b$, i.e., $\exists y\ [x = yb]$. Then

$$\overline{x}hb = \overline{x}agb = xgb = ybgb = yb = x = \overline{x}a$$

Since $\overline{x}$ was arbitrary, $hb = a$ because 1 is a generator. This shows $a \subseteq b$. ∎

Our final group of axioms is the following:

Axiom 6. Each object other than 0 (i.e. other than initial objects) has elements.

Axiom 7. Each element of a sum is a member of one of the injections.

Axiom 8. There is an object with more than one element.

We remark that Axiom 8 is independent of all the remaining axioms, and that, in fact, the full set of axioms with 8 excluded is easily proved consistent, for all these axioms are verified in the finite category

$$0 \longrightarrow 1$$

(In this case $N = 1$.) However, the question of independence for Axioms 4,6,7 is still unsettled.

Proposition 3. 0 has no elements.

Proof. If $1 \to 0$, then $1 = 0$, so that every object has exactly one element, contradicting axiom 8. ∎

Proposition 4. The two injections $1 \overset{i_0}{\underset{i_1}{\rightrightarrows}} 2$ are different and are the only elements of 2.

Proof. The second assertion is immediate by Axiom 7. Suppose $i_0 = i_1$ and let S be an object with two distinct elements x, y. Then there is t such that

$$\begin{tikzcd}
1 \arrow[dr, "i_0"'] \arrow[drr, "x"] & & \\
& 2 \arrow[r, "t"] & S \\
1 \arrow[ur, "i_1"] \arrow[urr, "y"'] & &
\end{tikzcd}$$

commutes by the sum axiom, so $x = i_0 t = i_1 t = y$, a contradiction. Thus $i_0 \neq i_1$. ∎

It is also easy to establish, by using the axiom of choice, that if A has exactly two elements, then $A \cong 2$.

Proposition 5. To any two objects A, B, the two injections

$$\begin{tikzcd} A \arrow[dr, "i_A"] & \\ & A+B \\ B \arrow[ur, "i_B"'] & \end{tikzcd}$$

have no members in common.

Proof. Consider $A + B \to 2$ induced by the unique maps $A \to 1$, $B \to 1$. Suppose there are a, b such that $ai_A = bi_B$ are equal elements of $A + B$. Then the diagram

$$\begin{tikzcd} & A \arrow[rr] \arrow[dr, "i_A"] & & 1 \arrow[dr, "i_0"] & \\ 1 \arrow[ur, "a"] \arrow[dr, "b"'] & & A+B \arrow[rr] & & 2 \\ & B \arrow[rr] \arrow[ur, "i_B"'] & & 1 \arrow[ur, "i_1"'] & \end{tikzcd}$$

is commutative. But the composite maps $1 \xrightarrow{a} A \to 1$ and $1 \xrightarrow{b} B \to 1$ must both be the identity, yielding $i_0 = i_1$, a contradiction. ∎

We now prove five theorems which, together with Theorem 1, are basic to the development of mathematics within our theory.

Theorem 2. Peano's postulates hold for N.

Proof. That the successor mapping $N \xrightarrow{s} N$ is well-defined is already implicit. To show that s is injective, note that the predecessor mapping can be defined by primitive recursion; one of the defining equations of the predecessor is that it be right inverse to s. The other defining equation for the predecessor p is that $0p = 0$, so that if 0 were a successor, say $ns = 0$, then $n = n(sp) = (ns)p = 0p = 0$. But if $0s = 0$, then *every* object has only the identity endomorphism. For let $X \xrightarrow{t} X$ and $x_0 \in X$; then there is $N \xrightarrow{x} X$ such that $0x = x_0$ and $sx = xt$, hence $x_0 t = 0xt = 0sx = 0x = x_0$. But x_0 being arbitrary, we can conclude $t = X$ since 1 is a generator. However, Axiom 8 implies that 2, for example, has non-identity endomorphisms; this contradiction shows that 0 is not the successor of any $n \in N$.

Finally we prove Peano's induction postulate. Let $A \xrightarrow{a} N$ be a subset such that $0 \in A$ and

$$\forall n \in N\ [n \in a \ \Rightarrow\ ns \in a]$$

The last means that a is included in its own inverse image under s, so there is $A \xrightarrow{t} A$ such that $as = ta$.

$$\begin{tikzcd}
A \arrow[dr, dashed] \arrow[drr, "t\ (\text{def})"] \arrow[ddr, "a"'] & & \\
& \bullet \arrow[r] \arrow[d] \arrow[dr, phantom, "\substack{s^{-1}[a] \\ \text{``meet''} \\ \text{diagram}}"] & A \arrow[d, "a"] \\
& N \arrow[r, "s"'] & N
\end{tikzcd}$$

The first means there is $\overline{0} \in A$ such that $\overline{0}a = 0$. By simple recursion $\overline{0}, t$ determine $N \xrightarrow{x} A$ such that $0x = \overline{0}$ and $sx = xt$. This implies

$$\begin{aligned} 0(xa) &= 0 \\ s(xa) &= (xa)s \end{aligned}$$

so that by the uniqueness of mappings defined by simple recursion, $xa = N$. Thus for any $n \in N$, $n = (nx)a$, so that $n \in a$. ■

The following theorem has the consequence that every mapping can be factored into an epimorphism followed by a monomorphism. However, from the viewpoint of general categories, the result is stronger. (It does not hold in the category of commutative rings with unit, for example.)

THEOREM 3. For any mapping f, the regular co-image of f is canonically isomorphic to the regular image of f. More precisely, the unique h rendering the following diagram commutative is an isomorphism, where k is the equalizer of $p_0 f, p_1 f$ and q is the coequalizer of kp_0, kp_1, while k^* is the coequalizer of fi_0, fi_1 and q^* is the equalizer of $i_0 k^*, i_1 k^*$.

$$\begin{tikzcd}
R_f \arrow[r, "k"] & A \times A \arrow[r, shift left, "p_0"] \arrow[r, shift right, "p_1"'] & A \arrow[r, "f"] \arrow[d, "q"'] & B \arrow[r, shift left, "i_0"] \arrow[r, shift right, "i_1"'] & B + B \arrow[r, "k^*"] & R_f^* \\
& & I^* \arrow[r, dashed, "h"'] & I \arrow[u, "q^*"'] & &
\end{tikzcd}$$

PROOF. We first show that hq^* is a monomorphism. This is clear if $A = 0$, so we assume A has elements. Then by the axiom of choice, q has a left inverse t: $tq = I^*$. Suppose $uhq^* = u'hq^*$. Then

$$\langle ut, u't\rangle p_0 f \;=\; utf \;=\; uhq^* \;=\; u'hq^* \;=\; u'tf \;=\; \langle ut, u't\rangle p_1 f$$

Hence, k being the equalizer of $p_0 f, p_1 f$, there is w such that $\langle ut, u't\rangle = wk$. But then, q being a coequalizer,

$$u \;=\; utq \;=\; \langle ut, u't\rangle p_0 q \;=\; wkp_0 q \;=\; wkp_1 q$$
$$= \langle ut, u't\rangle p_1 q \;=\; u'tq \;=\; u'$$

The above shows hq^* is a monomorphism, from which it follows that h is a monomorphism. A dual argument shows that h is an epimorphism and hence an isomorphism by Remark 3. ∎

The above theorem implies that any reasonable definition of image gives the same result, so we may drop the word "regular". There is no loss in assuming that $I* = h = I$; we then refer to the equation $f = qq*$ as the factorization of f through its image.

For the discussion of the remaining theorems we assume that a fixed labelling of the two maps $1 \overset{i_0}{\underset{i_1}{\rightrightarrows}} 2$ has been chosen. It is clear heuristically that the (equivalence classes of) subsets of any given set X are in one-to-one correspondence with the functions $X \to 2$. It is one of our goals to prove this. More exactly, we wish to show that every subset of X has a *characteristic function* $X \to 2$ in the sense of

DEFINITION 5. The mapping $X \xrightarrow{\varphi} 2$ is the *characteristic function* of the subset $A \xrightarrow{a} X$ iff

$$\forall x \in X\ [x \in a \;\Leftrightarrow\; x\varphi = i_1]$$

It is easy to see (using equalizers) that every function $X \to 2$ is the characteristic function of some subset. We will refer to those subsets which have characteristic functions as *special* subsets, until we have shown that every subset is special. This assertion will be made to follow from the construction of a "complement" $A' \xrightarrow{a'} X$ for every subset $A \xrightarrow{a} X$.

Essentially, the complement of a will be constructed as the union of all special subsets which do not intersect a. We first prove a theorem to the effect that such unions exist.

THEOREM 4. Given any "indexed family of special subsets of X"

$$I \xrightarrow{\quad\alpha\quad} 2^X$$

there exists a subset

$$\bigcup_\alpha \xrightarrow{\quad a\quad} X$$

which is the union of the α_j, $j \in I$, in the sense that for any $x \in X$,

$$x \in a \;\Leftrightarrow\; \exists j \in I\ [\langle x, j\alpha\rangle e_{2^X} = i_1]$$

where e_{2^X} is the evaluation $X \times 2^X \to 2$.

Proof. By exponential adjointness, α is equivalent to a mapping

$$X \times I \xrightarrow{\overline{\alpha}} 2$$

and the desired property of $\bigcup_\alpha$ is equivalent to

$$x \in a \;\Leftrightarrow\; \exists j \in I\ [\langle x, j\rangle \overline{\alpha} = i_1]$$

We construct $\bigcup_\alpha$ as follows:

$$\begin{tikzcd}
\sum_\alpha \arrow[r, "k"] \arrow[d, "q"'] & X \times I \arrow[r] \arrow[rr, bend left, "\overline{\alpha}"] \arrow[d, "p_X"] & 1 \arrow[r, "i_1"'] & 2 \\
\bigcup_\alpha \arrow[r, "a"'] & X & &
\end{tikzcd}$$

Here k is the equalizer of $\overline{\alpha}$ with the composite $X \times I \to 1 \xrightarrow{i_1} 2$, and $kp_X = qa$ is the factorization of kp_X through its image. Suppose $x \in a$. Then since q is an epimorphism there exists $\overline{x} \in \sum_\alpha$ such that $\overline{x}kp_X = x$. Define $j = \overline{x}kp_I$. Then

$$\langle x, j\rangle \overline{\alpha} \;=\; \overline{x}k\overline{\alpha} \;=\; \overline{x}k(X \times I \to 1 \xrightarrow{i_1} 2) \;=\; i_1$$

since k is an equalizer and since $\overline{x}k(X \times I \to 1) = 1$. The converse is even easier: if there is $j \in I$ such that $\langle x, j\rangle \overline{\alpha} = i_j$, then by the universal mapping property of the equalizer k, $\langle x, j\rangle$ must come from $\overline{x} \in \sum_\alpha$, so that on applying q to $\overline{x}$ one finds that $x \in a$. ■

Remark 4: Using the above construction of $\sum_\alpha$ we can prove a perhaps more familiar form of the axiom of choice, namely, given

$$I \xrightarrow{\alpha} 2^X$$

if every α_j is non-empty ($\alpha_j \neq 0$), there is $I \xrightarrow{f} X$ such that $jf \in \alpha_j$ for each $j \in I$. (Here by α_j we mean the subset of X obtained by equalizing with $X \to 1 \xrightarrow{i_1} 2$ the map $X \to 2$ whose "name" is $j\alpha$.) We leave the details to the reader, but note that the essential step is the application of the axiom of choice to the composite mapping

$$\sum_\alpha \longrightarrow X \times I \xrightarrow{p_I} I$$

We also leave to the reader the construction of the "product" $\prod_\alpha$ which is the subset of $(\sum_\alpha)^I$ consisting of all choice functions f as above. Notice that, although in our heuristic model $\mathcal{S}$, $\sum_\alpha$ and $\prod_\alpha$ are actual (infinite) coproduct and product in the categorical sense, there is no way to guarantee by first order axioms that this will be true in every first order model of our theory. (See also Remark 11.)

Theorem 5. Given any subset $A \xrightarrow{a} X$, there is a subset $A' \xrightarrow{a'} X$ such that $X \cong A + A'$, with a, a' as the injections. Thus $x \in a'$ iff $x \in X$ and $x \notin a$, and a has a characteristic function $X \to 2$.

We will need for the proof the following

Lemma 1. Given a subset $A \xrightarrow{a} X$ and an element $x \in X$ such that $x \notin a$, then there is $X \xrightarrow{\varphi} 2$ such that $x\varphi = i_1$ and $a\varphi \equiv i_0$; that is, there is a special subset containing x but intersecting a vacuously.

Proof of Lemma 1. Define $\binom{a}{x}$ by the universal mapping property of sums so that

$$\begin{tikzcd} A \arrow[dr, "i_A"'] \arrow[drr, "a"] & & \\ & A+1 \arrow[r, "\binom{a}{x}"] & X \\ 1 \arrow[ur, "i_1"] \arrow[urr, "x"'] & & \end{tikzcd}$$

and let u be a "quasi-inverse" for $\binom{a}{x}$ as guaranteed by the axiom of choice. Then define φ to be the composition

$$X \xrightarrow{u} A + 1 \xrightarrow{(A\to 1)+1} 1 + 1 \ = \ 2$$

By axiom 7, $\binom{a}{x}$ is a monomorphism since a is and since $x \notin a$, so we actually have that $\binom{a}{x}\, u = \ A + 1$. Therefore (setting $t_A \ = \ (A \to 1) + 1$)

$$\begin{aligned} x\varphi \ &= \ xut_A \ = \ i_1\tbinom{a}{x}ut_A \ = \ i_1 t_A \ = \ i_1 \\ a\varphi \ &= \ aut_A \ = \ i_A\tbinom{a}{x}ut_A \ = \ i_A t_A \ = \ i_0 \end{aligned}$$

■

Proof of Theorem 5. Let $j_0 \in 2^A$ be the element corresponding by exponential adjointness to the composition $A \times 1 \xrightarrow{p_1} 1 \xrightarrow{i_0} 2$, and then define $a \xrightarrow{\mu} 2^X$ to be the equalizer of 2^a with the composition $2^X \to 1 \xrightarrow{j_0} 2^A$. Thus intuitively μ is the subset of 2^X whose members are just the special subsets of X which intersect a vacuously. Define a' to be the union of μ. Thus

$$\begin{tikzcd}
\sum_\mu \arrow[r, "k"] \arrow[d, "q"'] & X \times a \arrow[r] \arrow[d, "p_X"] \arrow[rr, bend left, "\overline{\mu}"] & 1 \arrow[r, "i_1"'] & 2 \\
A' = \bigcup_\mu \arrow[r, "a'"'] & X & &
\end{tikzcd}$$

where $\overline{\mu}$ corresponds to μ and k is the equalizer of $\overline{\mu}$ with the function constantly i_1, and where q is an epimorphism while a' is a monomorphism. There is, of course a mapping f such that the diagram

$$\begin{tikzcd}
A \arrow[dr, "i_A"'] \arrow[drr, "a"] & & \\
& A + A' \arrow[r, "f"] & X \\
A' \arrow[ur, "i_{A'}"] \arrow[urr, "a'"'] & &
\end{tikzcd}$$

commutes. We must show that f is an isomorphism. For this it suffices to show that f is a monomorphism and an epimorphism. That f is a monomorphism follows from axiom 7 and the fact that, by construction, a and a' have no members in common, since $x \in a$ implies that $\langle x, t\rangle e_{2^X} = i_0$ for every $t \in \mu$, whereas $x \in a'$ implies that $\exists t \in \mu\ [\langle x, t\rangle e_{2^X} = i_1]$. On the other hand, to show that f is an epimorphism, it suffices to show that f is surjective, so consider an arbitrary $x \in X$. If $x \notin A$ then by the lemma there is $[\varphi] \in 2^X$ such that $x\varphi = i_1$ while $a\varphi \equiv i_0$. Since $a\varphi \equiv i_0$, we actually have $[\varphi] \in \mu$ by construction of $a \xrightarrow{\mu} 2^X$. Because $\langle x, [\varphi]\rangle e_{2^X} = i_1$, $\langle x, [\varphi]\rangle$ is actually a member of $k(X \times \mu)$ where k is the equalizer of $\overline{\mu}$ with the map constantly i_1, so finally applying $\sum_\mu \xrightarrow{q} A'$ to the element of $\sum_\mu$ which corresponds to $\langle x, [\varphi]\rangle$, we find that $x \in a'$. Thus, since we have shown that x comes either from A or A', f is surjective. ∎

The final "mathematical" theorem we shall prove concerns a more "internal" description of the nature of coequalizers. Actually the theorem only refers to an important special case, but we will remark that the general case can also be handled, leaving details to the reader. We consider a pair

$$\begin{tikzcd}
R \arrow[r, shift left, "f_0"] \arrow[r, shift right, "f_1"'] & A
\end{tikzcd}$$

of mappings, and will sometimes denote by $f = \langle f_0, f_1\rangle$ the corresponding single mapping $R \to A \times A$, from which f_0 and f_1 can be recovered.

DEFINITION 6. We say that f is *reflexive* iff

$$\exists d\ [A \xrightarrow{d} R \quad \& \quad df_0 = A = df_1]$$

symmetric iff

$$\exists t\ [R \xrightarrow{t} R \quad \& \quad tf_0 = f_1 \quad \& \quad tf_1 = f_0]$$

transitive iff

$$\forall h_0\ \forall h_1\ \{X \underset{h_1}{\overset{h_0}{\rightrightarrows}} R \quad \& \quad h_0 f_1 = h_1 f_0$$
$$\Rightarrow \exists u\ [uf_0 = h_0 f_0 \quad \& \quad uf_1 = h_1 f_1]\}$$

It has previously been pointed out (Lawvere 1963a) that the following theorem holds in every "algebraic" category and is in a sense characteristic of such categories. Since $\mathcal{S}$ is the basic algebraic category, it is essential that we should be able to prove the theorem from our axioms for $\mathcal{S}$. Notice that although our proof depends on the axiom of choice, and although the validity of the theorem for algebraic categories depends on its validity for $\mathcal{S}$, nevertheless the axiom of choice does not hold in most algebraic categories.

THEOREM 6. If $R \underset{f_1}{\overset{f_0}{\rightrightarrows}} A$ is such that $f = \langle f_0, f_1 \rangle$ is reflexive, symmetric, transitive, and a subset of $A \times A$, then f is the equalizer of $p_0 q$ with $p_1 q$, where $A \xrightarrow{q} Q$ is the coequalizer of f_0 with f_1.

PROOF. We actually show that if f is reflexive, symmetric, and transitive, and if $a_0 \in A,\ a_1 \in A$, then

$$a_0 q = a_1 q \ \Leftrightarrow\ \exists r \in R\ [rf_0 = a_0\ \&\ rf_1 = a_1]$$

which clearly implies the assertion in the theorem. We accomplish this by constructing a mapping $A \xrightarrow{g} 2^A$ which is intuitively just the classification of elements of A into f-equivalence classes. For the construction of g we need two lemmas:

LEMMA 2. for any A there is a mapping $A \xrightarrow{\{\}} 2^A$ such that for any $x \in A,\ y \in A$, y is a member of the subset whose characteristic function has the name $\{x\}$ iff $y = x$.

PROOF. Let $\{\}$ be the mapping corresponding by exponential adjointness to the characteristic function of the diagonal subset

$$A \xrightarrow{\langle A, A \rangle} A \times A$$

∎

Lemma 3. Given any mapping $R \xrightarrow{h} A$ there is a mapping

$$2^R \xrightarrow{\overline{h}} 2^A$$

with the property that given any subset of R with characteristic function ψ, then $[\psi]\overline{h}$ is the name of the characteristic function of the subset a of A having the property that for any $x \in A$,

$$x \in a \;\Leftrightarrow\; \exists r \; [r\psi = i_1 \quad \& \quad rh = x]$$

Briefly, $\overline{h}$ is the direct image mapping on subsets determined by the mapping h on the elements.

Proof. Consider the equalizer Ψ of e_{r^R} with i_1:

$$\begin{tikzcd} R' \arrow[r, "\Psi"] & R \times 2^R \arrow[rr, "e_{2^R}"] \arrow[r] & 1 \arrow[r, "i_1"'] & 2 \end{tikzcd}$$

Thus Ψ is a subset whose members are just the pairs $\langle r, [\psi]\rangle$ where ψ is the characteristic function of a subset of R of which r is a member. Form the composition of Ψ with

$$R \times 2^R \xrightarrow{h \times 2^R} A \times 2^R$$

Then the image of this composition is a subset of $A \times 2^R$ and hence has a characteristic function $A \times 2^R \xrightarrow{\Phi} 2$. Let $\overline{h}$ be the mapping $2^R \to 2^A$ corresponding by exponential adjointness to Φ. To show that $\overline{h}$ has the desired property, let $R \xrightarrow{\psi} 2$ and let a be the subset of A whose characteristic function φ has the name $[\psi]\overline{h}$, i.e $[\varphi] = [\psi]\overline{h}$ and φ is the characteristic function of a. Suppose $r\psi = i_1$ and $rh = x$; then we must show that $x \in a$, i.e. that $x\varphi = i_1$. However, since $\langle r, [\psi]\rangle \in \Psi$, one has also $\langle r, [\psi]\rangle(h \times 2^R) \in \Psi(h \times 2^R)$, which is the same as to say $rh, [\psi]\rangle\Phi = i_1$, which since $x = rh$, implies by definition of $\overline{h}$ that x is a member of the subset whose characteristic function is $[\psi]\overline{h} = [\varphi]$; but the latter subset is a, as required. Conversely, assume $x \in a$. Then $x\varphi = i_1$, so that $\langle x, [\varphi]\rangle\Phi = i_1$ and hence $\langle x, [\varphi]\rangle \in \Psi(h \times 2^R)$.

$$\begin{tikzcd} 1 \arrow[dr, dashed] \arrow[ddr, "{\langle x,[\varphi]\rangle}"'] & & \\ & R' \arrow[r, "\Psi"] & R \times 2^R \arrow[dl, "h \times 2^R"] \\ & A \times 2^R & \end{tikzcd}$$

Therefore there is $r \in R \ni rh = x$ and $\langle r, [\psi]\rangle \in \Psi$. But this means $r\psi = i_1$. ∎

Remark 5: It is easy to verify that in these lemmas we have actually constructed a covariant functor and a natural transformation $\{\}$ from the identity functor into this functor. There are actually three different "power-set" functors: the contravariant one $f \rightsquigarrow 2^f$, the covariant "direct image" functor $h \rightsquigarrow \overline{h}$ just constructed, and a dual covariant functor related to universal quantification in the same way that the direct image functor is related to existential quantification.

We now return to the proof of the theorem. Given the two mappings

$$\begin{tikzcd} R \arrow[r, shift left, "f_0"] \arrow[r, shift right, "f_1"'] & A \end{tikzcd}$$

we can by the lemmas construct two mappings $A \to 2^A$ as follows

$$\begin{tikzcd} & & 2^R \arrow[dr, "\overline{f}_1"] & \\ A \arrow[r, "\{\}"] & 2^A \arrow[ur, "2^{f_0}"] \arrow[dr, "2^{f_1}"'] & & 2^A \\ & & 2^R \arrow[ur, "\overline{f}_0"'] & \end{tikzcd}$$

We claim that if $f = \langle f_0, f_1 \rangle$ is symmetric, then the two composites $A \to 2^A$ are actually equal (to a map which we will call g), that if f is symmetric and transitive, then $f_0 g = f_1 g$, and that if f is reflexive, symmetric and transitive, then the induced map $Q \to 2^A$ (where $A \xrightarrow{q} Q$ is the coequalizer of f_0 with f_1) is actually the image of g (so in particular a subset of 2^A). When reflexivity does not hold, the difference between q and the co-image of g is that q keeps separate all elements of A which are not related to any element, whereas g maps all such elements to the empty subset of A.

First we show that symmetry implies $\{\}2^{f_0}\overline{f}_1 = \{\}2^{f_1}\overline{f}_0$ and hence a unique definition of g. Let $a, a' \in A$. Then a' is a member of the subset of A whose characteristic function has the name $\{a\}2^{f_0}\overline{f}_1$ (briefly $a' \underset{\sim}{\in} \{a\}2^{f_0}\overline{f}_1$) iff there is $r \in R$ such that $rf_0 = a$ and $rf_1 = a'$. But this condition is clearly symmetrical in f_0, f_1. (And of course "members" (in the sense of $\underset{\sim}{\in}$) determine elements of 2^A.)

Next we show that $f_0 g = f_1 g$, where $g = 2^{f_0}\overline{f}_1 = 2^{f_1}\overline{f}_0$, if symmetry and transitivity hold. Let $r \in R$. Then the desired relation $rf_0 g = rf_1 g$ holds iff

$$\forall a' \in A \; [a' \underset{\sim}{\in} rf_0 g \;\Leftrightarrow\; a' \underset{\sim}{\in} rf_1 g]$$

So let $a' \in A$. Then what we must show is that

$$\exists \overline{r} \in R \; [\overline{r} f_0 = a' \;\&\; \overline{r} f_1 = rf_0] \quad \Leftrightarrow \quad \exists \overline{\overline{r}} \in R \; [\overline{\overline{r}} f_0 = a' \;\&\; \overline{\overline{r}} f_1 = rf_1]$$

That the left hand side implies the right hand side follows immediately from transitivity ($a' \equiv rf_0 \;\&\; rf_0 \equiv rf_1 \;\Rightarrow\; a' \equiv rf_1$, where $\equiv$ is defined as the image of f) and the converse follows in the same way after first applying symmetry. Thus $f_0 g = f_1 g$ and so there is a unique induced map

$$\begin{tikzcd} R \arrow[r, shift left, "f_0"] \arrow[r, shift right, "f_1"'] & A \arrow[r, "q"] \arrow[dr, "g"'] & Q \arrow[d, dashed] \\ & & 2^A \end{tikzcd}$$

The third claim made above was that under the additional hypothesis of reflexivity, q is actually the co-image of g. This is essentially just the assertion that $a_0 g = a_1 g \Rightarrow a_0 q = a_1 q$. But $a_0 g = a_1 g$ means just that

$$\forall a' \in A\ [a' \equiv a_0 \Leftrightarrow a' \equiv a_1]$$

(where the relation $\equiv$ is just the image of f as above). By reflexivity there is at least one a' such that $a' \equiv a_0$ (namely a_0 itself) and by symmetry $a_0 \equiv a'$; but since $a_0 g = a_1 g$, it also follows from $a' \equiv a_0$ that $a' \equiv a_1$, so by transitivity $a_0 \equiv a_1$. This means $\exists r \in R\ [rf_0 = a_0\ \&\ rf_1 = a_1]$, and thus $a_0 q = rf_0 q = rf_1 q = a_1 q$ because $f_0 q = f_1 q$, q being the coequalizer of f_0 with f_1.

Finally we can assert our theorem

$$a_0 q = a_1 g \Leftrightarrow a_0 \equiv a_1$$

The implication $a_0 \equiv a_1 \Rightarrow a_0 q = a_1 q$ is trivial (and we just used it), while the converse was proved in the preceding paragraph. ■

Remark 6: It follows easily now that for an *arbitrary* pair of mappings

$$\begin{tikzcd} R \arrow[r, shift left, "f_0"] \arrow[r, shift right, "f_1"'] & A \end{tikzcd}$$

the equalizer $\tilde{f}$ of $p_0 q$ with $p_1 q$, where q is the coequalizer, is the smallest RST subset of $A \times A$ containing the image of f, and that the coequalizer of $\tilde{f}$ is the same as the coequalizer of f. Thus an "internal" construction of an arbitrary coequalizer will be obtained if one knows a way to construct the RST hull of an arbitrary relation. The construction below involves the natural numbers (see Remark 8 in connection with our metatheorem below). Since the symmetric hull is easy to construct we may assume that f is symmetric and reflexive. The reader may verify that the following then describes a construction of the RST hull $\tilde{f}$.

Let $\mathscr{S}_f \to R^N$ be the subset whose members are just the sequences $N \xrightarrow{r} R$ such that $srf_0 = rf_1$, s being the successor mapping (that is, form the equalizer of f_0^s with f_1^N). Then define $\tilde{f}$ to be the image of the following composite mapping $N \times N \times \mathscr{S}_f \to A \times A$:

$$\begin{tikzcd}
N \times N \times \mathscr{S}_f \arrow[r, "diag"] \arrow[dd, "\mathrm{epi}"'] & (N \times \mathscr{S}_f) \times (N \times \mathscr{S}_f) \arrow[r] & (N \times R^N) \times (N \times R^N) \arrow[d, "(N\times f_0^N)\times(N\times f_1)^N"] \\
 & & (N \times A^N) \times (N \times A^N) \arrow[d, "e \times e"] \\
\bullet \arrow[rr, "\tilde{f}"'] & & A \times A
\end{tikzcd}$$

(Here e is the evaluation mapping for sequences $N \to A$.) That $\tilde{f}$ is the RST hull of the reflexive and symmetric f then follows from the fact that $\langle a_0, a_1 \rangle \in \tilde{f}$ iff there exists a sequence r in R such that $srf_0 = rf_1$ and there exist n_0, n_1 such that $a_0 = n_0 x$ and $a_1 = n_1 x$ where $x = rf_0$.

Remark 7: For the development of mathematics in our theory it is also convenient to make use of a "Restricted Separation Schema" the best form of which the author has not yet determined. This theorem schema is to the effect that given any formula involving sets and maps as well as a free variable $x \in A$, if the quantifiers in the formula are suitably restricted (say to maps between sets which result from certain numbers of applications of the operations of Axioms 1-3 to the given sets and maps), then a subset of A exists whose members are precisely the elements of A which satisfy the formula. (Note that A itself could be a product set $B^3 \times C$, etc.) The Restricted Separation Schema is equivalent to a schema which asserts that there exists a mapping having any given (suitably restricted) rule. This theorem schema is surely strong enough to guarantee the existence of all the constructions commonly encountered in analysis. Without using the theorem schema, one can of course derive particular constructions using equalizers, exponentiation, images, etc, and this can often be illuminating. For example, the ring of Cauchy sequences of rational numbers and the ideal of sequences converging to 0 can be constructed and calculus developed.

The reader should also be able now to construct proofs of Tarski's fixed point theorem, the Cantor-Schroeder-Bernstein theorem, and Zorn's lemma.

Finally, we wish to prove a metatheorem which clarifies the extent to which our theory characterizes $\mathcal{S}$. Though our proof is informal it could easily be formalized within a sufficiently strong set theory of the traditional type or within a suitable theory of the category of categories. [For a preliminary description of the latter, see the author's doctoral dissertation (Lawvere 1963b).]

Let $\mathbb{C}$ be any category which satisfies the eight axioms of our theory. (To be more precise, we also assume that for each pair of objects C, C' in $\mathbb{C}$, the class (C, C') of maps $C \to C'$ is "small", i.e. is a "set".) There is then a canonical functor

$$\mathbb{C} \xrightarrow{\mathrm{H}^1} \mathcal{S}$$

which assigns to each object C in $\mathbb{C}$ the (identity map of the) set $(1, C)$ of all maps $1 \to C$ in $\mathbb{C}$.

The functor H^1 is clearly left exact (i.e. preserves 1, products, and equalizers) and faithful (i.e. if $C \rightrightarrows C'$ are identified by H^1 then they are equal). Our metatheorem states that under the additional (non-first-order) axiom of completeness, H^1 is an equivalence of categories, i.e. there is a functor $\mathcal{S} \to \mathbb{C}$ which, up to natural equivalence, is inverse to H^1. Thus our axioms serve to separate the structure of $\mathcal{S}$ from the structure of all other complete categories, such as those of topological spaces, vector spaces, partially ordered sets, groups, rings, lattices, etc.

Remark 8: For any model $\mathbb{C}$, H^1 also preserves sums and coequalizers of RST pairs of maps. However it seems unlikely that it is necessary that H^1 be right exact, since as remarked above, the construction of the RST hull involves N and by Gödel's theorems the nature of the object N may vary from one model $\mathbb{C}$ of any theory to another model. That is, H^1 need not preserve N. Also H^1 will be *full at* 1 in the sense that the induced mapping

$$(1, A) \longrightarrow (1, A\mathrm{H}^1)$$

must be surjective for each object A in $\mathbb{C}$. However, it is certainly not necessary that H^1 be full, because there are countable models by the Skolem-Lowenheim theorem, yet $(1, N)$ is always infinite; thus (N, N) can be countable but $(N\mathrm{H}^1, N\mathrm{H}^1)$ must be uncountable, so that e.g. the induced map

$$(N, N) \longrightarrow (N\mathrm{H}^1, N\mathrm{H}^1)$$

need not be surjective. Of course the countability of $(N, N) = (1, N^N)$ could not be demonstrated by maps $N \to N^N$ in $\mathbb{C}$. For the same reason, H^1 need not preserve exponentiation.

Actually, we will derive our result (that our axioms together with completeness characterize $\mathcal{S}$) from the following more general metatheorem concerning the category of models for our theory. We will use some of the basic facts about adjoint functors.

Metatheorem. Let $\mathbb{C} \xrightarrow{\mathrm{T}} \mathbb{C}'$ be a functor such that:

1. Both $\mathbb{C}$ and $\mathbb{C}'$ satisfy axioms 1-8.
2. Both $\mathbb{C}$ and $\mathbb{C}'$ have the property that for each object A the lattice of subobjects of A is complete.
3. T has an adjoint $\check{\mathrm{T}}$.

4. T is full at 1, i.e. for each object A in $\mathbb{C}$ the induced mapping
$$(1, A)_{\mathbb{C}} \longrightarrow (1\mathrm{T}, A\mathrm{T})_{\mathbb{C}'}$$
is surjective.

Then T is an equivalence of categories, i.e. $\mathrm{T}\check{\mathrm{T}}$ and $\check{\mathrm{T}}\mathrm{T}$ are naturally equivalent to the identity functors of $\mathbb{C}$ and $\mathbb{C}'$, respectively.

Remark 9: Completeness of the lattice of subobjects of A in $\mathbb{C}$ means that every family $A_\alpha \xrightarrow{a_\alpha} A$, $\alpha \in \mathcal{F}$ of subobjects of A has an intersection $D \to A$ and a union $V \to A$. Here *every* family is to be understood in an absolute sense, i.e., relative to the universe in which we are discussing model theory; neither $\mathcal{F}$ nor the mapping which takes $\alpha \rightsquigarrow a_\alpha$ need be an object or map in $\mathbb{C}$. We can show of course (Theorems 4 & 5) that in any model $\mathbb{C}$, any family $I \xrightarrow{\alpha} 2^A$ in $\mathbb{C}$ has a uniquely determined union and intersection in $\mathbb{C}$. However, as with any first-order set theory, we cannot guarantee that, in a given model $\mathbb{C}$ every absolute family of subsets of A is represented by a single mapping in $\mathbb{C}$ with codomain 2^A. Note that, in particular, lattice-completeness of $\mathbb{C}$ implies that to every (absolute) family of elements of A there is a subset of A with precisely those elements as members.

Note that we clearly need to assume that T is full at 1, since for any model $\mathbb{C}$ and any object I in $\mathbb{C}$, $A \rightsquigarrow A^I$ is a functor $\mathbb{C} \to \mathbb{C}$ which has an adjoint; but which is surely not an equivalence if, e.g., I is infinite.

Proof of Metatheorem. Since T has an adjoint it must preserve products and equalizers and in particular $1\mathrm{T} = 1$. Since T is full at 1, $0\mathrm{T}$ can have no elements, and so $0\mathrm{T} = 0$ by Axiom 6.

We claim that T is faithful, i.e., that each induced mapping
$$(A, B) \longrightarrow (A\mathrm{T}, B\mathrm{T})$$
is injective. Suppose $A \underset{g}{\overset{f}{\rightrightarrows}} B$ are such that $f\mathrm{T} = g\mathrm{T}$ and consider $1 \xrightarrow{x} A$. We must show $xf = xg$, so consider the equalizer $E \to 1$ of xf with xg. We must have $E = 1$ or $E = 0$. If $E = 0$, then $E\mathrm{T} = 0$ is the equalizer of $(xf)\mathrm{T}$ with $(xg)\mathrm{T}$ which implies $(x\mathrm{T})(f\mathrm{T}) \neq (x\mathrm{T})(g\mathrm{T})$, a contradiction. Hence $E = 1$ which implies $xf = xg$ and so $f = g$. Thus T is faithful.

Also, T must preserve sums, because for each A and B in $\mathbb{C}$, the canonical map
$$A\mathrm{T} + B\mathrm{T} \longrightarrow (A + B)\mathrm{T}$$
is surjective by Axiom 7 and the fact that T is full at 1, and is injective by Axiom 7 and the fact that any functor whose domain satisfies our axiom of choice must preserve monomorphisms.

Now the lattice-completeness of $\mathbb{C}$ and $\mathbb{C}'$ together with the fact that T is full at 1 implies that every subset of $A\mathrm{T}$ is the value of T at some subset of A, for any object A in $\mathbb{C}$. For
$$(1, 2^A) \longrightarrow (1, 2^{A\mathrm{T}})$$

is a homomorphism of complete atomic Boolean algebras which induces an isomorphism

$$(1, A) \xrightarrow{\approx} (1, A\mathrm{T})$$

on the respective sets of atoms, and hence must be itself an isomorphism. But this implies that T must be full, since arbitrary maps are determined by their graphs, which are subsets.

In general, if a functor T is full and faithful and has an adjoint $\check{\mathrm{T}}$, then for each A, the canonical map $A\mathrm{T}\check{\mathrm{T}} \to A$ is an isomorphism. Hence this must hold in our case.

To complete the proof of the metatheorem we need only show that for each I in $\mathbb{C}'$, the canonical map

$$I \xrightarrow{\varphi_I} I\check{\mathrm{T}}\mathrm{T}$$

is injective, for it follows then by the previous remark that φ_I, being a subset of a value of T, must actually have its domain I isomorphic to a value of T, which immediately implies that φ_I is an isomorphism.

Because $A\mathrm{T}\check{\mathrm{T}} \cong A$, T is full and hence $1\mathrm{T} = 1$. Also, $\check{\mathrm{T}}$, being an adjoint functor, must preserve sums. Since we have already seen that T preserves 1 and sums, it follows that the particular canonical map

$$2 \xrightarrow{\varphi_2} 2\check{\mathrm{T}}\mathrm{T}$$

is an isomorphism. This, together with the fact that 2 is a cogenerator for $\mathbb{C}'$ (Proposition 1), will imply that every φ_I is injective.

For suppose $1 \overset{x}{\underset{y}{\rightrightarrows}} I$ are different. Then there is $I \xrightarrow{t} 2$ such that $xt \neq yt$. In the commutative diagram

$$\begin{tikzcd}
1 \arrow[dr, shift left, "x"] \arrow[dr, shift right, "y"'] & & \\
& I \arrow[r, "\varphi_I"] \arrow[d, "t"'] & I\check{\mathrm{T}}\mathrm{T} \arrow[d, "t\check{\mathrm{T}}\mathrm{T}"] \\
& 2 \arrow[r, "\varphi_2"'] & 2\check{\mathrm{T}}\mathrm{T}
\end{tikzcd}$$

φ_2 is an isomorphism and hence $t\check{\mathrm{T}}\mathrm{T}$ must separate $x\varphi_I$ from $y\varphi_I$. ∎

METACOROLLARY. If $\mathbb{C}$ is a complete category which satisfies our axioms then the canonical functor

$$\mathbb{C} \xrightarrow{\mathrm{H}^1} \mathcal{S}$$

is an equivalence.

PROOF. Completeness means that the categorical sum and product of any family of objects exists. However, this can easily be shown to imply lattice completeness. If we define

$$\mathcal{S} \xrightarrow{\check{\mathrm{T}}} \mathbb{C}$$

so that for each $I \in \mathcal{S}$, $I\check{\mathrm{T}}$ is the I-fold repeated sum of 1 with itself, then $\check{\mathrm{T}}$ is adjoint to $\mathrm{T} = \mathrm{H}^1$. By definition of $\mathrm{H}^1 = \mathrm{T}$, it is full at 1. ∎

REMARK 10: To give some indication of the extent to which a model of our axioms may fail to be complete, note that the set of all mappings between sets of rank less than $\omega + \omega$ is a model.

REMARK 11: Even countable repeated sums of an object with itself may fail to exist, for the equation

$$(\sum_N A, X) \cong (A, X)^N$$

shows that any non-trivial category in which such sums exist must be uncountable, whereas any first-order theory has countable models. Of course, if such sums do exist in a model for our theory, then

$$\sum_N A \cong N \times A$$

and the latter operation *can* always be performed.

REMARK 12: Of course, it is important to investigate first-order strengthenings of our axiom system in the direction of completeness, although cardinals larger than $\aleph_\omega$ are usually not needed, e.g., for the development of analysis. The existence of this and many more cardinals would be guaranteed, and hence the model of Remark 10 excluded, if we take the following as an axiom schema

$$\begin{array}{l} \forall A \forall B \forall B' \; [\, \varphi(A,B) \;\&\; \varphi(A,B') \quad \Rightarrow \quad B \cong B' \,] \\ \qquad\qquad\qquad\qquad\qquad\qquad \Downarrow \\ \qquad\qquad \forall X \exists Y \; [\, X < Y \;\&\; \forall A \forall B \; [\, A < Y \;\&\; \varphi(A,B) \quad \Rightarrow \quad B < Y]] \end{array}$$

where φ is any formula and $X < Y$ refers to cardinality.

However, it is the author's feeling that when one wishes to go substantially beyond what can be done in the theory presented here, a much more satisfactory foundation for practical purposes will be provided by a theory of the category of categories.

REFERENCES.

Eilenberg, Samuel and Saunders Mac Lane (1945), 'General Theory of Natural Equivalences', *Transactions of the American Mathematical Society* **58**, 231–94.

Lawvere, F. William (1963a), 'Functorial Semantics of Algebraic Theories', *Proceedings of the National Academy of Sciences of the U.S.A.* **50**, 869–872.

Lawvere, F. William (1963b), Functorial Semantics of Algebraic Theories, PhD thesis, Columbia University. Republished in: *Reprints in Theory and Applications of Categories*, No. 5 (2004) pp 1-121, available on-line at http://tac.mta.ca/tac/reprints/articles/5/tr5abs.html.

This article may be accessed from `http://www.tac.mta.ca/tac/reprints` or by anonymous ftp at
`ftp://ftp.tac.mta.ca/pub/tac/html/tac/reprints/articles/11/tr11.{dvi,ps}`

Reprints in Theory and Applications of Categories, No. 16, 2006, pp. 1–16.

ADJOINTNESS IN FOUNDATIONS

F. WILLIAM LAWVERE

Author's commentary

In this article we see how already in 1967 category theory had made explicit a number of conceptual advances that were entering into the everyday practice of mathematics. For example, local Galois connections (in algebraic geometry, model theory, linear algebra, etc.) are globalized into functors, such as Spec, carrying much more information. Also, "theories" (even when presented symbolically) are viewed explicitly as categories; so are the background universes of sets that serve as the recipients for models. (Models themselves are functors, hence preserve the fundamental operation of substitution/composition in terms of which the other logical operations can be characterized as local adjoints.)

My 1963 observation (referred to by Eilenberg and Kelly in La Jolla, 1965), that cartesian closed categories serve as a common abstraction of type theory and propositional logic, permits an invariant algebraic treatment of the essential problem of proof theory, though most of the later work by proof theorists still relies on presentation-dependent formulations. This article sums up a stage of the development of the relationship between category theory and proof theory. (For more details see Proceedings of the AMS Symposium on Pure Mathematics XVII (1970), pp. 1–14, and Marcel Dekker, Lecture Notes in Pure and Applied Mathematics, no. 180 (1996), pp. 181–189.)

The main problem addressed by proof theory arises from the existential quantifier in "there exists a proof…". The strategy to interpret proofs themselves as structures had been discussed by Kreisel; however, the influential "realizers" of Kleene are not yet the usual mathematical sort of structures. Inspired by Läuchli's 1967 success in finding a completeness theorem for Heyting predicate calculus lurking in the category of ordinary permutations, I presented, at the 1967 AMS Los Angeles Symposium on Set Theory, a common functorization of several geometrical structures, including such proof-theoretic structures. As Hyperdoctrines, those structures are described in the Proceedings of the

Originally published: Dialectica, **23** (1969), used by permission.
My heartfelt thanks go to Christoph Schubert who expertly rendered the original article into TEX and to the editors of TAC for their work and dedication.
Received by the editors of TAC February 22, 2006 and by Dialectica December 15, 1967.
Transmitted by R. Paré, R. Rosebrugh, R.J. Wood. Republished on 2006-05-21, revised 2006-10-30.
2000 Mathematics Subject Classification: 00A30, 14F20, 18B25, 18B30, 18B40, 55U40, 81P05.
Key words and phrases: Formal-conceptual duality, cartesian-closed categories, algebraic logic, globalized Galois connections.

AMS New York Symposium XVII, cited above.

Proof theory may be regarded as the study of the presentations of certain algebraic structures, for example, locally cartesian closed categories $\mathbf{C}$ with finite coproducts. The map category $\mathbf{C}/X$ models "proof bundles" with its morphisms playing the role of deductions. Then $\mathbf{P}X$, defined as the poset reflection of $\mathbf{C}/X$, is a Heyting algebra; any map $X \to Y$ in $\mathbf{C}$ contravariantly induces a Heyting homomorphism of substitution and covariantly induces existential and universal quantifier operations that not only satisfy the correct rules of inference but, moreover, satisfy the usual proof-theoretic slogans that

(a) a proof of an existential statement includes the specification of an element of the kind required by the statement;

(b) a proof of a universal statement is a functional giving a uniform proof of all the instances.

The poset reflection expresses the idea of "there exists a deduction $A \to B$" in $\mathbf{C}/X$, and $\mathbf{P}X$ serves as a system of "proof-theoretic propositions" about elements of X. In case $\mathbf{C}$ is actually a topos, there is a natural map

$$\mathbf{P}X \to \mathbf{P}_{\mathbf{C}}(X)$$

to the usual subobject lattice, defined by taking the image of any $A \to X$. This map will be an isomorphism for all X only if $\mathbf{C}$ satisfies the axiom of choice, but we might hope that at least $\mathbf{P}$ is "small" like $\mathbf{P}_{\mathbf{C}}$ and that the idempotent of $\mathbf{P}$ whose splitting is $\mathbf{P}_{\mathbf{C}}$ could be described in a useful way. There are two results about that.

(1) The Heyting algebras $\mathbf{P}X$ are small iff $\mathbf{C}$ itself is Boolean, as in Läuchli's original construction. The foregoing equivalence is true at least for presheaf toposes, as was shown by Matias Menni in his thesis (2000). A stronger precise condition of "smallness", expressed by the existence of a generic proof, has also been investigated by him, including the relation to local connectedness and other conditions on more general toposes in his paper "Cocomplete Toposes whose exact completions are toposes" to appear in the Journal of Pure and Applied Algebra[1].

(2) If we localize the definition of $\mathbf{P}$ in the sense that we do not ask for the existence of maps $A \to B$ in $\mathbf{C}/X$, but only for maps $A' \to B$ where A' has an epimorphism to A, then this coarsened version of $\mathbf{P}$ is actually isomorphic to $\mathbf{P}_{\mathbf{C}}$ for all toposes $\mathbf{C}$.

The second result divides the existential quantifier in the proof problem into two steps; the preliminary search for an appropriate A' which covers A illustrates that a hypothesis must sometimes be analyzed before it can be used to launch a deduction. The collapse from a type system to the corresponding system of proof-theoretic propositions is of course not an "isomorphism" (contrary to a fashionable colloquialism), even though both involve cartesian closed categories. To arrive at mathematical truth, we require the further collapse resulting from the unbounded preliminary step.

Buffalo, 22 February 2006

[1]The latter two sentences of this paragraph were added October 30, 2006

1. The Formal–Conceptual Duality in Mathematics and in its Foundations[2]

That pursuit of exact knowledge which we call mathematics seems to involve in an essential way two dual aspects, which we may call the Formal and the Conceptual. For example, we manipulate algebraically a polynomial equation and visualize geometrically the corresponding curve. Or we concentrate in one moment on the deduction of theorems from the axioms of group theory, and in the next consider the classes of actual groups to which the theorems refer. Thus the Conceptual is in a certain sense the subject matter of the Formal.

Foundations will mean here the study of what is universal in mathematics. Thus Foundations in this sense cannot be identified with any "starting-point" or "justification" for mathematics, though partial results in these directions may be among its fruits. But among the other fruits of Foundations so defined would presumably be guide-lines for passing from one branch of mathematics to another and for gauging to some extent which directions of research are likely to be relevant.

Being itself part of Mathematics, Foundations also partakes of the Formal-Conceptual duality. In its formal aspect, Foundations has often concentrated on the formal side of mathematics, giving rise to Logic. More recently, the search for universals has also taken a conceptual turn in the form of Category Theory, which began by viewing as a new mathematical object the totality of all morphisms of the mathematical objects of a given species A, and then recognizing that these new mathematical objects all belong to a common non-trivial species C which is independent of A. Naturally, the formal tendency in Foundations can also deal with the conceptual aspect of mathematics, as when the semantics of a formalized theory T is viewed itself as another formalized theory T', or in a somewhat different way, as in attempts to formalize the study of the category of categories. On the other hand, Foundations may conceptualize the formal aspect of mathematics, leading to Boolean algebras, cylindric and polyadic algebras, and to certain of the structures discussed below.

One of the aims of this paper is to give evidence for the universality of the concept of adjointness, which was first isolated and named in the conceptual sphere of category theory, but which also seems to pervade logic. Specifically, we describe in section III the notion of cartesian closed category, which appears to be the appropriate abstract structure for making explicit the known analogy (sometimes exploited in proof theory) between the theory of functionality and propositional logic. The structure of a cartesian closed category is entirely given by adjointness, as is the structure of a "hyperdoctrine", which includes quantification as well. Precisely analogous "quantifiers" occur in realms of mathematics normally considered far removed from the province of logic or proof theory.

---

[2]I inserted the following subtitles into the text for this reprinting:
1. The formal-conceptual duality in mathematics and in its foundations
2. Adjointness in the meta-category of categories
3. Cartesian-closed categories and hyperdoctrines
4. Globalized Galois connections in algebraic geometry and in foundations

As we point out, recursion (at least on the natural numbers) is also characterized entirely by an appropriate adjoint; thus it is possible to give a theory, roughly proof theory of intuitionistic higher-order number theory, in which all important axioms (logical or mathematical) express instances of the notion of adjointness.

The above-discussed notions of Conceptual, Formal, and Foundations play no mathematical role in this paper; they were included in this introductory section only to provide one possible perspective from which to view the relationship of category theory in general and of this paper in particular to other work of universal tendency. However, if one wished to take these notions seriously, it would seem to follow that an essential feature of any attempt to formalize Foundations would be a description of this claimed "duality" between the Formal and the Conceptual; indeed, both category theory and set theory succeed to some extent in providing such a description in certain cases. Concerning the latter point there is a remark at the end of section IV, which is otherwise devoted to a discussion of a class of adjoint situations differing from those of section III but likewise seeming to be of universal significance — these may be described briefly as a sort of globalized Galois connection.

2. Adjointness in the Meta-Category of Categories

The formalism of category theory is itself often presented in "geometric" terms. In fact, to give a category is to give a meaning to the word *morphism* and to the *commutativity* of diagrams like

$$A \xrightarrow{f} B \qquad \begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "h"'] & B \arrow[d, "g"] \\ & C \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "f"] \arrow[d, "a"'] & B \arrow[d, "b"] \\ A' \arrow[r, "f'"'] & B' \end{tikzcd} \qquad \text{etc.}$$

which involve morphisms, in such a way that the obvious associativity and identity conditions hold, as well as the condition that whenever

$$A \xrightarrow{f} B \qquad \text{and} \qquad B \xrightarrow{g} C$$

are commutative then there is just one h such that

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "h"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

is commutative.

To save printing space, one also says that A is the *domain*, and B the *codomain* of f when

$$A \xrightarrow{f} B$$

is commutative, and in particular that h is the *composition* $f.g$ if

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "h"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

is commutative. We regard *objects* as co-extensive with identity morphisms, or equivalently with those morphisms which appear as domains or codomains. As usual we call a morphism which has a two-sided inverse an *isomorphism.*

With any category $\mathbf{A}$ is associated another $\mathbf{A}^{\mathrm{op}}$ obtained by maintaining the interpretation of "morphism" but reversing the direction of all arrows when re-interpreting diagrams and their commutativity. Thus $A \xrightarrow{f} B$ in $\mathbf{A}^{\mathrm{op}}$ means $B \xrightarrow{f} A$ in $\mathbf{A}$, and $h = f.g$ in $\mathbf{A}^{\mathrm{op}}$ means $h = g.f$ in $\mathbf{A}$; clearly the interpretations of "domain", "codomain" and "composition" determine the interpretation of general "commutativity of diagrams". The category-theorist owes an apology to the philosophical reader for this unfortunately well-established use of the word "commutativity" in a context more general than any which could reasonably have a description in terms of "interchangeability".

A *functor* F involves a domain category $\mathbf{A}$, a codomain category $\mathbf{B}$, and a mapping assigning to every morphism x in $\mathbf{A}$ a morphism xF in $\mathbf{B}$ in a fashion which preserves the commutativity of diagrams. (Thus in particular a functor preserves objects, domains, codomains, compositions, and isomorphisms.) Utilizing the usual composition of mappings to define commutative diagrams of functors, one sees that functors are the morphisms of a super-category, usually called the *(meta-) category of categories.*

A category with exactly one morphism will be denoted by $\mathbf{1}$. It is determined uniquely up to isomorphism by the fact that for any category $\mathbf{A}$, there is exactly one functor $\mathbf{A} \longrightarrow \mathbf{1}$; the functors $\mathbf{1} \xrightarrow{A} \mathbf{A}$ correspond bijectively to the objects in $\mathbf{A}$.

A *natural transformation* φ involves a domain category $\mathbf{A}$, a codomain category $\mathbf{B}$, domain and codomain functors

$$\mathbf{A} \underset{F_1}{\overset{F_0}{\rightrightarrows}} \mathbf{B}$$

and a mapping assigning to each object A of $\mathbf{A}$ a morphism

$$AF_0 \xrightarrow{A\varphi} AF_1$$

in $\mathbf{B}$, in such a way that for every morphism

$$A \xrightarrow{a} A'$$

in $\mathbf{A}$, the diagram

$$\begin{tikzcd} AF_0 \arrow[r, "A\varphi"] \arrow[d, "aF_0"'] & AF_1 \arrow[d, "aF_1"] \\ A'F_0 \arrow[r, "A'\varphi"'] & A'F_1 \end{tikzcd}$$

is commutative in $\mathbf{B}$. For fixed $\mathbf{A}$ and $\mathbf{B}$, the natural transformations are the morphisms of a *functor category* $\mathbf{B}^{\mathbf{A}}$, where we write

$$F_0 \xrightarrow{\varphi} F_1 \quad \text{in } \mathbf{B}^{\mathbf{A}}$$

for the above described situation, and define commutativity of triangles

$$\begin{tikzcd} F_0 \arrow[r, "\varphi"] \arrow[dr, "\varphi\cdot\overline{\varphi}"'] & F_1 \arrow[d, "\overline{\varphi}"] \\ & F_2 \end{tikzcd} \quad \text{in } \mathbf{B}^{\mathbf{A}}$$

by the condition

$$A(\varphi.\overline{\varphi}) = (A\varphi).(A\overline{\varphi})$$

for all objects A in $\mathbf{A}$.

A second *Godement multiplication* is also defined for natural transformations, in a way which extends in a sense the composition of functors. Namely, if

$$F_0 \xrightarrow{\varphi} F_1 \text{ in } \mathbf{B}^{\mathbf{A}} \text{ and } G_0 \xrightarrow{\psi} G_1 \text{ in } \mathbf{C}^{\mathbf{B}}$$

then $F_0G_0 \xrightarrow{\varphi\psi} F_1G_1$ in $\mathbf{C}^{\mathbf{A}}$ is the natural transformation which assigns to each A of $\mathbf{A}$ the morphism of $\mathbf{C}$ which may be indifferently described as either composition in the commutative square

$$\begin{tikzcd} AF_0G_0 \arrow[r, "(A\varphi)G_0"] \arrow[d, "(AF_0)\psi"'] & AF_1G_0 \arrow[d, "(AF_1)\psi"] \\ AF_0G_1 \arrow[r, "(A\varphi)G_1"'] & AF_1G_1 \end{tikzcd}$$

The Godement multiplication is functorial

$$\mathbf{B}^{\mathbf{A}} \times \mathbf{C}^{\mathbf{B}} \longrightarrow \mathbf{C}^{\mathbf{A}}$$

which means in particular that

$$(\varphi.\overline{\varphi})(\psi.\overline{\psi}) = (\varphi\psi).(\overline{\varphi}\overline{\psi}),$$

and associative, meaning that if also $H_0 \xrightarrow{\vartheta} H_1$ in $\mathbf{D}^{\mathbf{C}}$, then

$$(\varphi\psi)\vartheta = \varphi(\psi\vartheta) \text{ in } \mathbf{D}^{\mathbf{A}}.$$

Other rules follow from the facts that in any product category $\mathbf{L} \times \mathbf{M}$ (obvious definition) a sort a commutativity relation

$$\langle l, M\rangle.\langle L', m\rangle = \langle L, m\rangle.\langle l, M'\rangle$$

holds for

$$L \xrightarrow{l} L' \text{ in } \mathbf{L} \text{ and } M \xrightarrow{m} M' \text{ in } \mathbf{M},$$

and that if Godement multiplication is applied to identity natural transformations, it reduces to composition of the corresponding functors. Two intermediate cases will be important in the sequel: if

$$F_0 = \varphi = F_1 = F,$$

then

$$A(F\psi) = (AF)\psi$$

while if $G_0 = \psi = G_1 = G$, then $A(\varphi G) = (A\varphi)G$. Thus each functor $\mathbf{A} \xrightarrow{F} \mathbf{B}$ induces via Godement multiplication by F a functor $\mathbf{C}^{\mathbf{B}} \xrightarrow{\mathbf{C}^F} \mathbf{C}^{\mathbf{A}}$ for any category $\mathbf{C}$, and similarly each $\mathbf{B} \xrightarrow{G} \mathbf{C}$ induces $\mathbf{B}^{\mathbf{A}} \xrightarrow{G^{\mathbf{A}}} \mathbf{C}^{\mathbf{A}}$ for any category $\mathbf{A}$.

Now we come to the central concept, of which we will presently see many examples. An *adjoint situation* involves two categories $\mathbf{A}$ and $\mathbf{B}$, two functors

$$\mathbf{A} \underset{U}{\overset{F}{\rightleftarrows}} \mathbf{B}$$

and two natural transformations

$$\begin{aligned} \mathbf{A} &\xrightarrow{\eta} FU \qquad \text{in } \mathbf{A}^{\mathbf{A}} \\ UF &\xrightarrow{\varepsilon} \mathbf{B} \qquad \text{in } \mathbf{B}^{\mathbf{B}} \end{aligned}$$

satisfying the two equations (commutative triangles)

$$\begin{aligned} \eta F . F\varepsilon &= F \qquad \text{in } \mathbf{B}^{\mathbf{A}} \\ U\eta . \varepsilon U &= U \qquad \text{in } \mathbf{A}^{\mathbf{B}}. \end{aligned}$$

We state immediately an equivalent form of the definition. Let $(F, \mathbf{B})$ denote the category whose morphisms are those quadruples of morphisms $A \xrightarrow{a} A'$ in $\mathbf{A}$, $B \xrightarrow{b} B'$, $AF \xrightarrow{h} B$, $A'F \xrightarrow{h'} B'$ in $\mathbf{B}$ for which

$$\begin{tikzcd} AF \arrow[r, "aF"] \arrow[d, "h"'] & A'F \arrow[d, "h'"] \\ B \arrow[r, "b"'] & B' \end{tikzcd}$$

is commutative in $\mathbf{B}$; commutative diagrams in $(F, \mathbf{B})$ are defined so that forgetting h, h' in the above defines a functor

$$(F, \mathbf{B}) \longrightarrow \mathbf{A} \times \mathbf{B}.$$

The objects of $(F, \mathbf{B})$ thus correspond to pairs each consisting of a morphism $AF \xrightarrow{h} B$ in $\mathbf{B}$ with a given A in $\mathbf{A}$. Similarly a category $(\mathbf{A}, U)$ is defined whose objects correspond to pairs each consisting of a morphism $A \xrightarrow{f} BU$ in $\mathbf{A}$ with a given B in $\mathbf{B}$, and which also has a forgetful functor

$$(\mathbf{A}, U) \longrightarrow \mathbf{A} \times \mathbf{B}.$$

Now it can be shown that giving η, ε satisfying the two conditions above in order to complete an adjoint situation is equivalent with giving a *functor*

$$(F, \mathbf{B}) \xrightarrow{Q} (\mathbf{A}, U)$$

which commutes with the forgetful functors to $\mathbf{A} \times \mathbf{B}$ and which has a two-sided inverse

$$(\mathbf{A}, U) \xrightarrow{Q^{-1}} (F, \mathbf{B}).$$

In particular, Q assigns to the object $AF \xrightarrow{h} B$ of $(F, \mathbf{B})$ the object

$$A \xrightarrow{(A\eta).(hU)} BU$$

of $(\mathbf{A}, U)$, while Q^{-1} assigns

$$AF \xrightarrow{(fF).(B\varepsilon)} B$$

to $A \xrightarrow{f} BU$. Sometimes Q is called an adjointness, η an adjunction, and ε a co-adjunction for F and U; if such exists, F is said to be *left adjoint* to U and U is said to be *right adjoint* to F, and we write $F \dashv U$. Briefly, F and U may be placed in an adjoint situation iff bijections

$$(AF, B) \cong (A, BU)$$

can be given in a way which is "natural" or "functorial" when A and B vary; here the left-hand side denotes the set of all $AF \xrightarrow{h} B$ in $\mathbf{B}$ and the right-hand side denotes the set of all $A \xrightarrow{f} BU$ in $\mathbf{A}$.

Several important properties hold for adjoint situations in general. One is that adjoints are unique (if they exist). That is, any two adjoint situations with a given $\mathbf{A} \xrightarrow{F} \mathbf{B}$ are uniquely isomorphic, and similarly for a given $\mathbf{B} \xrightarrow{U} \mathbf{A}$. More generally, given two adjoint situations with fixed $\mathbf{A}$ and $\mathbf{B}$ (with and without primes), then any natural transformation $F' \longrightarrow F$ induces a unique $U \longrightarrow U'$, and conversely. If $\mathbf{A} \xrightarrow{F} \mathbf{B}$ with $F \dashv U$, and $\mathbf{B} \xrightarrow{\tilde{F}} \mathbf{C}$, with $\tilde{F} \dashv \tilde{U}$, then $F\tilde{F} \dashv \tilde{U}U$, and also $F^{\mathbf{D}} \dashv U^{\mathbf{D}}$ for any category $\mathbf{D}$. Thus a necessary condition for a functor U to have a left adjoint is that U commutes (up to isomorphism of functors) with all generalized limits which exist in $\mathbf{B}$ and $\mathbf{A}$, and dually a functor F can have a right adjoint only if F commutes up to isomorphism with all generalized colimits which exist in $\mathbf{A}$ and $\mathbf{B}$. Here, for any functor $\mathbf{D}' \xrightarrow{L} \mathbf{D}$, generalized limits in $\mathbf{A}$ along L and generalized colimits in $\mathbf{A}$ along L are, when they exist, respectively right and left adjoint to the induced functor $\mathbf{A}^L$. These generalized colimits include direct limits in the usual sense as well as free products and pushouts and evaluation adjoints; for many $\mathbf{A}$ of importance, they exist for all L with $\mathbf{D}'$ and $\mathbf{D}$ smaller than some fixed regular cardinal, but they will exist for all L iff $\mathbf{A}$ is a category in which the morphisms reduce to nothing more than a preordering on the objects, with respect to which the latter form a complete lattice.

The general adjoint functor theorem asserts that if a functor $\mathbf{B} \xrightarrow{U} \mathbf{A}$ satisfies the necessary continuity condition mentioned above, then it will have a left adjoint provided that $\mathbf{B}$ is suitably complete and U is suitably bounded. Freyd's Special Adjoint Functor Theorem shows that certain categories $\mathbf{B}$ (such as the category of sets or of abelian groups, but *not* the category of groups) are "compact" in the sense that every continuous U to a standard $\mathbf{A}$ (say the category of sets) is bounded. A schema formalizing these adjoint functor theorems for the case where $\mathbf{A}$ and $\mathbf{B}$ are replaced by metacategories comparable to the conceptual universe would, on the one hand, justify by itself a large portion of the existential richness of that universe, as follows at least to the extent of the higher types and of infinite objects from the following section, and on the other hand, provide a common rationale for the inevitability and basic properties of a large number of mathematical constructions, as the following examples indicate.

If Top denotes the category of continuous mappings of topological spaces, the diagonal functor

$$\mathrm{Top} \longrightarrow \mathrm{Top} \times \mathrm{Top}$$

has a right adjoint, which forces the definition of product topology. If A denotes, say, the unit interval, then

$$\mathrm{Top} \xrightarrow{A \times (\;)} \mathrm{Top}$$

also has a right adjoint, yielding the construction of the compact-open topology on path spaces. The inclusion functor Comp $\longrightarrow$ Top from the category of continuous mappings between compact spaces has a *left* adjoint, giving the Stone-Čech compactification construction.

Turning to algebra, the forgetful functor Gps $\longrightarrow$ Sets, which views every group-homomorphism as a mapping of the carrier sets, has a left adjoint, yielding the notion of free groups. The "commutator bracket" functor Assoc $\longrightarrow$ Lie, which interprets every homomorphism of associative linear algebras as a homomorphism of the same underlying vector spaces viewed only as Lie algebras, has also a left adjoint, yielding universal enveloping algebras. These two adjoint situations belong to a special large class, which also includes abelianization of groups, monoid rings, symmetric algebras, etc.

There are also adjoint situations in analysis, leading for example to l^1 spaces and almost-periodic functions. Needless to say, viewing all these constructions explicitly as adjoint situations seems to have certain formal and conceptual utility apart from any philosophical attempt to unify their necessity.

3. Cartesian-closed Categories and Hyperdoctrines

A *cartesian closed category* is a category $\mathbf{C}$ equipped with adjoint situations of the following three sorts:

(1) A *terminal object* $\mathbf{1} \xrightarrow{1} \mathbf{C}$, meaning a right adjoint to the unique $\mathbf{C} \longrightarrow \mathbf{1}$; if η denotes the adjunction, then $C \xrightarrow{C\eta} 1$ is the unique morphism in $\mathbf{C}$ with domain C and codomain 1, for each object C.

(2) A *product* $\mathbf{C} \times \mathbf{C} \xrightarrow{\times} \mathbf{C}$, meaning a right adjoint to the diagonal functor $\mathbf{C} \longrightarrow \mathbf{C} \times \mathbf{C}$. Denoting the adjunction and co-adjunction by δ and π respectively, we have $X\delta$ as a *diagonal* morphism $X \longrightarrow X \times X$ and $(Y_0, Y_1)\pi$ as a *projection* pair usually denoted for short by

$$Y_0 \times Y_1 \xrightarrow{\pi_0} Y_0, \qquad Y_0 \times Y_1 \xrightarrow{\pi_1} Y_1.$$

Given morphisms $X \xrightarrow{f_0} Y_0$, $X \xrightarrow{f_1} Y_1$, then $\langle f_0, f_1 \rangle = X\delta \cdot (f_0 \times f_1)$ is the unique $X \longrightarrow Y_0 \times Y_1$ whose compositions with the π_i are the f_i. The projection $A \times 1 \longrightarrow A$ is an isomorphism for any object A.

(3) For each object A, *exponentiation by* A, $\mathbf{C} \xrightarrow{(\ )^A} \mathbf{C}$, meaning a right adjoint to the functor $\mathbf{C} \xrightarrow{A \times (\ )} \mathbf{C}$. Denote by λ_A and ε_A the adjunction and co-adjunction respectively. Then

$$X \xrightarrow{X\lambda_A} (A \times X)^A \qquad \text{and} \qquad A \times Y^A \xrightarrow{Y\varepsilon_A} Y,$$

and for any $A \times X \xrightarrow{h} Y$, one has that

$$X \xrightarrow{X\lambda_A . h^A} Y^A$$

is the unique morphism g for which $(A \times g).Y\varepsilon_A = h$. For any morphism $A \xrightarrow{f} Y$, one may consider in the above process the case $X = 1$, $h = (\langle A, 1 \rangle \pi_1)^{-1}.f$, obtaining a morphism $1 \longrightarrow Y^A$ denoted for short by $\ulcorner f \urcorner$; every morphism $1 \longrightarrow Y^A$ is of form $\ulcorner f \urcorner$ for a unique $A \xrightarrow{f} Y$. Now for any $1 \xrightarrow{a} A$, $A \xrightarrow{f} Y$, one has that

$$\langle a \ulcorner f \urcorner \rangle . Y\varepsilon_A = a.f,$$

which justifies calling $Y\varepsilon_A$ the *evaluation* morphism, though in most cartesian closed $\mathbf{C}$, a paucity of morphisms with domain 1 prevents the last equation from determining ε.

From the uniqueness of adjoints it is clear that any two cartesian closed categories with a given category $\mathbf{C}$ are uniquely isomorphic. Most categories $\mathbf{C}$ cannot be made into cartesian closed categories at all. The category of all mappings between finite sets can be made cartesian closed, as can many larger categories of mappings between sets. Also, order-preserving mappings between partially ordered sets can be made cartesian closed, Y^A then being necessarily isomorphic to the set of order-preserving mappings $A \longrightarrow Y$ equipped with the usual pointwise partial ordering. Another kind of example is provided by a Brouwerian semi-lattice, these being essentially co-extensive with the "trivial" cartesian closed categories (those in which there is at most one morphism $X \longrightarrow Y$ for any given objects X and Y, or equivalently those in which all diagonal morphisms $X\delta$ are isomorphisms $X \longrightarrow X \times X$). More examples will appear presently.

A *hyperdoctrine* shall consist of at least the following four data

1. A cartesian closed category $\mathbf{T}$. We will refer sometimes to the objects of $\mathbf{T}$ as types, to the morphisms in $\mathbf{T}$ as terms, and to the morphisms $1 \longrightarrow X$ in particular as constant terms of type X.

2. For each type X, a corresponding cartesian closed category $P(X)$ called the category of attributes of type X. Morphisms of attributes will be called deductions over X, entailments, or inclusions as is appropriate. The basic functors giving $P(X)$ its closed structure will be denoted by

$$1_X, \quad \wedge_X, \quad \alpha_X \Rightarrow (\ )$$

to distinguish them from the analogous

$$1, \quad \times, \quad (\ )^A$$

in $\mathbf{T}$ (the subscripts X may be omitted when no confusion is likely). Thus the evaluation morphisms in $P(X)$ for each pair of objects α, ψ in $P(X)$

$$\alpha \wedge_X (\alpha \Rightarrow \psi) \longrightarrow \psi$$

may be sometimes more appropriately referred to as the modus ponens deductions.

3. For each term $X \xrightarrow{f} Y$ (morphism in $\mathbf{T}$) a corresponding functor

$$P(Y) \xrightarrow{f\cdot(\ )} P(X).$$

Then for any attribute ψ of type Y, $f \cdot \psi$ will be called the attribute of type X resulting from substituting f in ψ. We assume that for $Y \xrightarrow{g} Z$, $(f.g)\cdot\zeta = f\cdot(g\cdot\zeta)$ for all attributes ζ of type Z (at least up to coherent isomorphism) and similarly for deductions over Z.

4. For each $X \xrightarrow{f} Y$ in $\mathbf{T}$ two further functors

$$P(X) \xrightarrow{(\ )\Sigma f} P(Y) \quad \text{and} \quad P(X) \xrightarrow{(\ )\Pi f} P(Y)$$

equipped with adjunctions that make them respectively left and right adjoint to substitution f. For any object φ in $P(X)$ we call $\varphi\Sigma f$ (respectively $\varphi\Pi f$) the existential (respectively universal) quantification of φ along f. For ψ in $P(Y)$, the adjointnesses reduce to two (natural) bijections, one between deductions $\varphi\Sigma f \longrightarrow \psi$ over Y and deductions $\varphi \longrightarrow f \cdot \psi$ over X, and the other between deductions $\psi \longrightarrow \varphi\Pi f$ over Y and deductions $f \cdot \psi \longrightarrow \varphi$ over X. From the assumed functorality of substitution we have for a further term $Y \xrightarrow{g} Z$ that $\varphi\Sigma(f.g) = (\varphi\Sigma f)\Sigma g$ and $\varphi\Pi(f.g) = (\varphi\Pi f)\Pi g$, at least up to an invertible deduction in $P(Z)$.

Now among general features of hyperdoctrines we note for example that the existence of existential quantification implies that substitution commutes with conjunction; usually substitution does not commute with implication, as one of our examples below shows. There is at least one interesting realm that has roughly all the features of a hyperdoctrine *except* existential quantification, namely sheaf theory, wherein $\mathbf{T}$ is the category of continuous mappings between Kelley spaces[3] and $P(X)$ is the category of (morphisms between) set-valued sheaves on X.

Also, in a general hyperdoctrine we can construct for each term $X \xrightarrow{f} Y$ the attribute $1_X\Sigma\langle X, f\rangle$ of type $X \times Y$ which plays the role of the graph of f; in particular the graph of the identity morphism on X is an object in $P(X \times X)$ having the rudimentary properties of equality. With help of equality, graphs, and the behavior of quantification under composition of terms, one can in certain hyperdoctrines reduce the general quantification to quantification along very special terms, namely projections $Y \times Z \longrightarrow Y$.

Any (single-sorted) theory formalized in higher-order logic yields a hyperdoctrine in which types are just all expressions 1, V, V^V, $V \times V$, $(V \times V)^{(V\times V\times V)^V}$, etc. and in which terms are identified if they are in the theory provably equal (we assume that "higher-order logic" does involve at least terms corresponding to the δ, π, λ, ε necessary to yield a cartesian closed category by this procedure). In this hyperdoctrine, we let the objects of $P(X)$ be those formulas of the theory whose free variables are of type X (for example if $X = V \times V^V$, $P(X)$ consists of the formulas with two free variables, one an "element" variable and the other a "function" variable, but with bound variables of arbitrary type) and we let the morphisms of $P(X)$ be simply entailments deducible in the theory. Thus the diagonal $\varphi \longrightarrow \varphi \wedge_X \varphi$ is an isomorphism in this example. Also each $P(X)$ has a *coterminal* object 0_X (falsity) with a deduction $\varphi \longrightarrow (\varphi \Rightarrow_X 0_X) \Rightarrow_X 0_X$ which is an isomorphism if the logic is classical. If the theory is number theory, then $\mathbf{T}$ participates in another adjoint situation (setting $\omega = V$)

$$\begin{tikzcd} \mathbf{T} \arrow[loop above] \arrow[r, shift left] & \mathbf{T} \arrow[l, shift left, "\omega\times(\ )"] \end{tikzcd}$$

[3]Intended here are the k-spaces of Hurewicz, which are well-described in the book General Topology by J. L. Kelley.

where the unlabeled right adjoint functor is the forgetful functor from the category whose objects correspond to endomorphisms $X \xrightarrow{t} X$ in $\mathbf{T}$ and whose morphisms are diagrams

$$\begin{tikzcd} X \arrow[loop above, "t"] \arrow[r, "h"] & Y \arrow[loop above, "s"] \end{tikzcd}$$

in $\mathbf{T}$ with $t.h = h.s$. Conversely, any cartesian closed category in which this forgetful functor has a left adjoint will contain a definite morphism corresponding to each higher-type primitive recursive function, and if the category contains somewhere at least one non-identity endomorphism, then there will be an infinite number of morphisms $1 \longrightarrow \omega$ in the category.

We also obtain a hyperdoctrine if we consider as types arbitrary sets, as terms arbitrary mappings, and as attributes arbitrary subsets, (i.e. $P(X) = 2^X$) and take substitution as the inverse image operator, which forces existential quantification to be the direct image operator. With the appropriate definition of morphism between hyperdoctrines, such a morphism into this one from a higher order theory (as in the previous paragraph) involves nothing more nor less than a *model* of the theory in question.

A different hyperdoctrine with the same category of types and terms (namely arbitrary sets and mappings) is obtained by defining $P(X) = (\mathbf{T}, X) =$ the category of "sets over X" whose morphisms are arbitrary commutative triangles

$$\begin{tikzcd} A \arrow[rr, "d"] \arrow[dr, "\varphi"'] & & A' \arrow[dl, "\varphi'"] \\ & X & \end{tikzcd}$$

of mappings of sets, and taking for $f \cdot \psi$ (when $X \xrightarrow{f} Y$ and $B \xrightarrow{\psi} Y$) the projection to X of the set of pairs $\langle x, b\rangle$ with $xf = b\psi$. Then existential quantification is simply composition $\varphi\Sigma f = \varphi.f$. Since $P(1) = \mathbf{T}$, for any element $1 \xrightarrow{x} X$ and "property" φ in $P(X)$, $x \cdot \varphi$ is simply a set, the "fiber" of φ over x. Calling a deduction $1_X \longrightarrow \varphi$ over X a "proof" of φ, and noting that in the present example 1_X is just the identity mapping, we see that a proof of φ is a mapping that assigns to each x a proof of $x \cdot \varphi$. Then a proof of $\varphi\Sigma f$ is a mapping that assigns to each y a specific x such that $xf = y$, together with a proof of $x \cdot \varphi$. Since the fiber over y of a universal quantification $y.(\varphi\Pi f) = \Pi x \cdot \varphi$, a proof of $\varphi\Pi f$ involves a mapping that assigns to each y a mapping that assigns to each x with $xf = y$ a proof that $x \cdot \varphi$, and more generally deductions with premises other than 1_X have similar descriptions. The validity of the axiom of choice in $\mathbf{T}$ then implies that

$$(\mathbf{T}, X) \xrightarrow{\text{direct image}} 2^X$$

commutes with all the "propositional" and quantificational operations and hence defines the strictest sort of morphism from the present hyperdoctrine to the previous one. Since in this case every ψ in $P(Y)$ is of the form $1_X\Sigma f = \psi$ for a unique X and f, it is possible to view conjunction and implication as particular cases of substitution and universal quantification. In particular a deduction over X of a conjunction $\varphi_1 \wedge_X \varphi_2$ involves an ordered pair of mappings, each of which may be an arbitrary deduction of φ_1, or φ_2 respectively; this language makes clear the reasonableness of the non-idempotence of conjunction.

As a final example of a hyperdoctrine, we mention the one in which types are finite categories and terms arbitrary functors between them, while $P(\mathbf{A}) = \mathbf{S}^{\mathbf{A}}$, where $\mathbf{S}$ is the category of finite sets and mappings, with substitution as the special Godement multiplication. Quantification must then consist of generalized limits and colimits, while implication works like this: $\alpha \Rightarrow \psi$ is a functor whose value at an object A in $\mathbf{A}$ is the set of natural transformations $h_A \times \alpha \longrightarrow \psi$, where $h_A : \mathbf{A} \longrightarrow \mathbf{S}$ is the "representable" functor assigning to A' the set of morphisms $\mathbf{A}(A, A')$. By focusing on those $\mathbf{A}$ having one object and all morphisms invertible, one sees that this hyperdoctrine includes the theory of permutation groups; in fact, such $\mathbf{A}$ *are* groups and a "property" of $\mathbf{A}$ is nothing but a representation of $\mathbf{A}$ by permutations. Quantification yields "induced representations" and implication gives a kind of "intertwining representation". Deductions are of course equivariant maps.

4. Globalized Galois Connections in Algebraic Geometry and in Foundations

If $\mathbf{O}_1$ and $\mathbf{O}_2$ are partially ordered sets the traditional notion of a *Galois connection* between them is easily seen to be equivalent to an adjoint situation

$$\mathbf{O}_1^{\mathrm{op}} \rightleftarrows \mathbf{O}_2$$

between the corresponding trivial categories (one of them turned opposite). Recently it has begun to appear that the basic examples of Galois connections are really just fragments of more global adjoints which involve non-trivial categories and which carry more information.

For example, Galois theory itself has profitably been treated as the study of adjoint situations

$$\mathrm{Alg}_k^{\mathrm{op}} \rightleftarrows \mathrm{Sets}^{\mathrm{Gal}(\overline{k},k)}$$

where the left-hand category has as objects all commutative and associative linear algebras over a field k and the right-hand all continuous permutation representations of the compact Galois group of the separable closure $\overline{k}$ of k, while the left-to-right right adjoint functor assigns to every k-algebra A the representation by permutations of the set $\mathrm{Alg}_k(A, \overline{k})$ of all algebra homomorphisms $A \longrightarrow \overline{k}$. Restricting to subalgebras of $\overline{k}$ on the one hand and to quotients of the regular representation on the other retrieves in effect the usual Galois connection. Similarly, in algebraic geometry the Galois connection between ideals in a polynomial ring $k[x_1, \ldots, x_n]$ and subsets of k^n has been globalized to an adjoint situation

$$\mathrm{Alg}_k^{\mathrm{op}} \underset{\Gamma}{\overset{\mathrm{spec}}{\rightleftarrows}} \mathrm{Schema}_k$$

where Γ assigns the algebra of global functions to each scheme.

Finally, in Foundations there is the familiar Galois connection between sets of axioms and classes of models, for a fixed set of relation variable R_i. Globalizing to an adjoint pair allows making precise the semantical effects, not only of increasing the axioms, but also of omitting some relation symbols or reinterpreting them in a unified way. The *categories* of models then reciprocally determine their own full sets of natural relation variables, thus giving definability theory a new significance outside the realm of axiomatic classes. To do this for a given species — equational, elementary, higher-order, etc. — of, say, I-sorted theories, one defines an adjoint situation

$$\mathrm{Theories}^{\mathrm{op}} \underset{\mathrm{structure}}{\overset{\mathrm{semantics}}{\rightleftarrows}} (\mathrm{Cat}, [\mathrm{Sets}^I])$$

in which the right hand side denotes a category whose morphisms are commutative triangles

$$\begin{tikzcd} \mathbf{C} \arrow[rr] \arrow[dr] & & \mathbf{C}' \arrow[dl] \\ & \mathrm{Sets}^I & \end{tikzcd}$$

of functors with $\mathbf{C}$ and $\mathbf{C}'$ more or less arbitrary categories. The invariant notion of theory here appropriate has, in all cases considered by the author, been expressed most naturally by identifying a theory T itself with a category of a certain sort, in which case the semantics (categories of models) of T is a certain subcategory of the category of functors $T \longrightarrow \mathrm{Sets}^I$. There is then a further adjoint situation

$$\mathrm{Formal} \rightleftarrows \mathrm{Theories}$$

describing the presentation of the invariant theories by means of the formalized languages appropriate to the species. Composing these two adjoint situations, and tentatively identifying the Conceptual with categories of the general sort $(\mathrm{Cat}, [\mathrm{Sets}^I])$, we arrive at a family of adjoint situations

$$\mathrm{Formal}^{\mathrm{op}} \rightleftarrows \mathrm{Conceptual}$$

(one for each species of theory) which one may reasonably hope constitute the fragments of a precise description of the duality with which we began our discussion.

Bibliography

EILENBERG, S., and KELLY, G.M., Closed Categories, Proceedings of the Conference on Categorical Algebra (La Jolla 1965), Springer Verlag 1966.

EILENBERG, S., and MAC LANE, S., General Theory of Natural Equivalences, Trans. Amer. Math. Soc. **58**, 231–294 (1945).

KAN, D. M., Adjoint Functors, Trans. Amer. Math. Soc. **87**, 294–329 (1958).

LAWVERE, F. W., Functorial Semantics of Algebraic Theories, Proc. Nat. Acad. Sc. U.S.A. **50**, 869–872 (1963).

LAWVERE, F. W., The Category of Categories as a Foundation for Mathematics, Proceedings of the Conference on Categorical Algebra (La Jolla 1965), Springer Verlag 1966 [See also Review 7332 by J. Isbell, Dec. 1967, Math. Reviews].

MAC LANE, S., Categorical Algebra, Bull. Amer. Math. Soc. **71**, 40–106 (1965).

This article may be accessed via WWW at `http://www.tac.mta.ca/tac/reprints` or by anonymous ftp at

`ftp://ftp.tac.mta.ca/pub/tac/html/reprints/articles/16/tr16.{dvi,ps,pdf}`

Reprints in Theory and Applications of Categories, No. 8, 2005, pp. 1–24.

TAKING CATEGORIES SERIOUSLY

F. WILLIAM LAWVERE

ABSTRACT. The relation between teaching and research is partly embodied in simple general concepts which can guide the elaboration of examples in both. Notions and constructions, such as the spectral analysis of dynamical systems, have important aspects that can be understood and pursued without the complication of limiting the models to specific classical categories. Pursuing that idea leads to a dynamical objectification of Dedekind reals which is particularly suited to the simple identification of metric spaces as enriched categories over a special closed category. Rejecting the complacent description of that identification as a mere analogy or amusement, its relentless pursuit [8] is continued, revealing convexity and geodesics as concepts having a definite meaning over any closed category. Along the way various hopefully enlightening exercises for students (and possible directions for research) are inevitably encountered: (1) an explicit treatment of the contrast between multiplication and divisibility that, in inexorable functorial fashion, mutates into the adjoint relation between autonomous and non-autonomous dynamical systems; (2) the role of commutation relations in the contrast between equilibria and orbits, as well as in qualitative distinctions between extensions of Heyting logic; (3) the functorial contrast between translations and rotations (as appropriately defined) in an arbitrary non symmetric metric space.

The theory of categories originated [1] with the need to guide complicated calculations involving passage to the limit in the study of the qualitative leap from spaces to homotopical/homological objects. Since then it is still actively used for those problems but also in algebraic geometry [2], logic and set theory [3], model theory [4], functional analysis [5], continuum physics [6], combinatorics [7], etc. In all these the categorical concept of adjoint functor has come to play a key role.

Such a universal instrument for guiding the learning, development, and use of advanced mathematics does not fail to have its indications also in areas of school and college mathematics, in the most basic relationships of space and quantity and the calculations based on those relationships. In saying "take categories seriously", I advocate noticing, cultivating,

---

After the original manuscript was stolen, I was able to reconstruct, in time for inclusion in the proceedings of the 1983 Bogotá Workshop, this"sketch" (as Saunders Mac Lane's review was to describe it). Interested teachers can elaborate this and similar material into texts for beginning students. For the present opportunity, I am grateful to Xavier Caicedo who gave permission to reprint, to Christoph Schubert for his expert transcription into TEX, and to the editors of *TAC*.

Received by the editors 2004-11-02.

Transmitted by M. Barr, R. Rosebrugh and R. F. C. Walters. Reprint published on 2005-03-23.

2000 Mathematics Subject Classification: 06D20, 18D23, 37B55, 34L05, 52A99, 53C22, 54E25, 54E40.

Key words and phrases: General spectral theory, Nonautonomous dynamical systems, Symmetric monoidal categories, Semimetric spaces, General convexity, Geodesics, Heyting algebras, Special maps on metric spaces.

The article originally appeared in Revista Colombiana de Matemáticas, XX (1986) 147-178.

and teaching of helpful examples of an elementary nature.

1. Elementary mutability of dynamical systems

Already in [1] it was pointed out that a preordered set is just a category with at most one *morphism* between any given pair of objects, and that functors between two such categories are just order-preserving maps; at the opposite extreme, a monoid is just a category with exactly one *object*, and functors between two such categories are just homomorphisms of monoids. But category theory does not rest content with mere classification in the spirit of Wolffian metaphysics (although a few of its practitioners may do so); rather it is the *mutability* of mathematically precise structures (by morphisms) which is the essential content of category theory. If the structures are themselves categories, this mutability is expressed by functors, while if the structures are functors, the mutability is expressed by natural transformations. Thus if Λ is a preordered set and $\mathcal{X}$ is any category (for example the category of sets and mappings, the category of topological spaces and continuous mappings, the category of linear spaces and linear transformations, or the category of bornological linear spaces and bounded linear transformations) then there are functors

$$\Lambda \longrightarrow \mathcal{X}$$

sometimes called "direct systems" in $\mathcal{X}$, and the natural transformations

$$\begin{tikzcd} \Lambda \arrow[r, bend left] \arrow[r, bend right] \arrow[r, phantom, "\Downarrow"] & \mathcal{X} \end{tikzcd}$$

between two such functors are the appropriate morphisms for the study of such direct systems as objects.

An important special case is that where $\Lambda = \boxed{0 \to 1}$, the ordinal number $\mathbf{2}$, then the functors $\mathbf{2} \longrightarrow \mathcal{X}$ may be identified with the morphisms in the category $\mathcal{X}$ itself; likewise if $\Lambda = \boxed{0 \to 1 \to 2 \to \cdots}$ is the ordinal number ω, functors

$$\omega \xrightarrow{X} \mathcal{X}$$

are just sequences of objects and morphisms

$$X_0 \longrightarrow X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow \ldots$$

in $\mathcal{X}$, and a natural transformation $X \xrightarrow{f} Y$ between two such is a sequence $X_n \xrightarrow{f_n} Y_n$ of morphisms in $\mathcal{X}$ for which all squares

$$\begin{tikzcd} X_n \arrow[r, "f_n"] \arrow[d] & Y_n \arrow[d] \\ X_{n+1} \arrow[r, "f_{n+1}"'] & Y_{n+1} \end{tikzcd}$$

commute in $\mathcal{X}$ (here the vertical maps are the ones given as part of the structure of X and Y).

Similarly, if M is a monoid then the functors $M \longrightarrow \mathcal{X}$ are extremely important mathematical objects often known as *actions* of M on objects of $\mathcal{X}$ (or *representations* of M by $\mathcal{X}$-endomorphisms, or ...) and the natural transformations between such actions are known variously as M-equivariant maps, intertwining operators, homogeneous functions, etc. depending on the traditions of various contexts. Historically the notion of monoid (or of group in particular) was abstracted from the actions, a pivotally important abstraction since as soon as a particular action is constructed or noticed, the demands of learning, development, and use mutate it into: 1) other actions on the same object, 2) actions on other related objects, and 3) actions of related monoids. For if $M \xrightarrow{X} \mathcal{X}$ is an action and $M' \xrightarrow{h} M$ is a homomorphism, then (composition of functors!) Xh is an action of M', while if $\mathcal{X} \xrightarrow{C} \mathcal{Y}$ is a functor, then CX is an action of M on objects of $\mathcal{Y}$. To exemplify, if M is the additive group of time-translations, then a functor $M \longrightarrow \mathcal{X}$ is often called a dynamical system (continuous-time and autonomous) in $\mathcal{X}$, but if we are interested in observing the system only on a daily basis we could consider a homomorphism $M' \xrightarrow{h} M$ where $M' = \mathbb{N}$ is the additive monoid of natural numbers, and concentrate attention on the predictions of the discrete-time, autonomous, future-directed dynamical system Xh. In other applications we might have $M' = M =$ the multiplicative monoid of real numbers, but consider the homomorphism $M \xrightarrow{(\ )^p} M$ of raising to the pth power; then if we are given two actions $M \rightrightarrows \mathcal{X}$ on objects of $\mathcal{X}$, a natural transformation $X \xrightarrow{f} (Y)^p$ is just a morphism of the underlying $\mathcal{X}$-objects which satisfies

$$f(\lambda x) = \lambda^p f(x)$$

for all λ in M and all $T \xrightarrow{x} X$ in $\mathcal{X}$, i.e. a function homogeneous of degree p. An extremely important example of the second mutation of action mentioned above is that in which $\mathcal{Y}$ is the opposite of an appropriate category of algebras and the functor C assigns to each object (domain of variation) of $\mathcal{X}$ an algebra of functions (= intensive quantities) on it. Then the induced action CX of M describes the evolution of intensive quantities which results from the evolution of "states" as described by the action X. A frequently-occurring example of the third type of mutation of action arises from the surjective homomorphisms $M' \longrightarrow M$ from the additive monoid of time-translations M' to the circle group M. Then a dynamical system $M' \xrightarrow{X} \mathcal{X}$ is said to be "periodic of period h" if there exists a commutative diagram of functors as follows:

$$\begin{tikzcd}
M' \arrow[rr, "X"] \arrow[dr, "h"'] & & X \\
& M \arrow[ur, dotted] &
\end{tikzcd}$$

2. Kan quantifiers in spectral analysis

Most dynamical systems are only partly periodic, and such an analysis can conveniently be expressed by "Kan-extensions" as follows (we do not assume that M, M' are monoids):

For a functor $M' \xrightarrow{h} M$ and a category $\mathcal{X}$, the induced functor $\mathcal{X}^M \longrightarrow \mathcal{X}^{M'}$ will often have a left adjoint $X \longmapsto h\sum X$ and a right adjoint $X \longmapsto h\prod X$. These Kan adjoints vastly generalize the existential and universal quantifiers of logic, special cases arising when all the objects of $\mathcal{X}$ are truth-values. (Usually this just means that the objects of $\mathcal{X}$ are canonically idempotent with respect to cartesian product or coproduct.) Kan adjoints also generalize the induced representations frequently considered when $\mathcal{X}$ is a *linear* category (i.e. finite coproduct is canonically isomorphic to finite product) and when $M' \longrightarrow M$ is of "finite index", in which case there is a strong tendency for $h\sum(\ )$ and $h\prod(\ )$ to coincide. The defining property of these as adjoints are the natural bijections

$$\frac{h\sum X \longrightarrow Y}{X \longrightarrow Yh} \qquad\qquad \frac{T \longrightarrow h\prod X}{Th \longrightarrow X}$$

between M'-natural, respectively M-natural transformations, where Y, T are functors $M \longrightarrow \mathcal{X}$ (that is objects of the category $\mathcal{X}^M$ whose morphisms are the M-natural transformations) and X is a functor $M' \longrightarrow \mathcal{X}$ (that is an object of the category $\mathcal{X}^{M'}$ whose morphisms are M'-natural transformations). Since these refined "rules of inference" uniquely characterize the adjoints up to unique natural isomorphism, if $M'' \xrightarrow{k} M' \xrightarrow{h} M$ are two functors for each of which the two Kan adjoints exist, then from the associativity of substitution, $Y(hk) = (Yh)k$, follow the two rules

$$(hk)\sum Z \cong h\sum(k\sum Z)$$
$$(hk)\prod Z \cong h\prod(k\prod Z).$$

If M' is a discrete category $\mathcal{I}$ with I objects (and no morphisms except the identity morphisms) and if M is the single morphism category $\mathbf{1}$, then there is a unique functor $\mathcal{I} \longrightarrow \mathbf{1}$, often also called I and the Kan adjoints are just the coproduct and product functors respectively:

$$I\sum X = \textstyle\sum_{i\in I} X_i$$
$$I\prod X = \textstyle\prod_{i\in I} X_i$$

where a functor $I \xrightarrow{X} \mathcal{X}$ is just a family of objects. It is chiefly in regard to the existence of Kan extensions that questions of "largeness" and "smallness" enter category theory. The class of all categories $\mathcal{X}$ for which $h\sum(\ )$ and $h\prod(\ )$ exist in $\mathcal{X}$ can be called the "smallness" of $M' \xrightarrow{h} M$, while dually (in the sense of Galois connections) the class of all functors h for which these exist over a given $\mathcal{X}$ can be called the degree of "(bi) completeness" of $\mathcal{X}$, with obvious refinements for left completeness where only $\sum$ is considered and for right completeness where only $\prod$ is considered. Informally we may just say that $M' \xrightarrow{h} M$ is sufficiently small for $\mathcal{X}$ or that $\mathcal{X}$ is sufficiently complete for h when these constructions can be carried out.

Returning to the example of a "period", i.e. a surjective homomorphism $M' \xrightarrow{h} M$ from the additive group of time-translations to the circle group, the induced functor

$$\mathcal{X}^M \hookrightarrow \mathcal{X}^{M'}$$

is just the full inclusion, into the category of all $\mathcal{X}$-dynamical systems (continuous, autonomous), of the subcategory of those that happen to have period h. Then the construction $h \prod X$ just gives the part of X consisting of the h-periodic states. More precisely the following adjunction morphism (derived from the rule of inference for $\prod$)

$$(h \textstyle\prod X)h \longrightarrow X$$

will typically be the inclusion of the h-periodic part.

[Of course there is also

$$X \longrightarrow (h \textstyle\sum X)h$$

obtained by forcing the arbitrary dynamical system X into the h-periodic mold, with an accompanying collapse of states, whose detailed understanding depends on a detailed understanding of the "collapsing" or quotient process in $\mathcal{X}$. The quotient process is just in general $\mathbf{\Delta}_1^{\mathrm{op}} \sum(\ )$. Here $\mathbf{\Delta}_1^{\mathrm{op}}$ is the finite category $\boxed{E \;\substack{\longrightarrow\\ \longleftarrow\\ \longrightarrow}\; V}$ in which the two composites at V are both the identity (implying that the two composites at E are idempotents which absorb one another in a non-commutative way) and functors $\mathbf{\Delta}_1^{\mathrm{op}} \longrightarrow \mathcal{X}$ are often referred to as (reflexive) graphs in $\mathcal{X}$; $\mathbf{\Delta}_1$ itself can be concretely represented as the full category of the category of categories consisting of the two objects $V = \mathbf{1}$ and $E = \mathbf{2} = \boxed{0 \to 1}$ in which representation the two arrows $V \;\substack{\longrightarrow\\ \longleftarrow\\ \longrightarrow}\; E$ in $\mathbf{\Delta}_1$ are the two *adjoints* of the unique $E \longrightarrow V$. The reflexive graph in $\mathcal{X}$ arising from a period h (homomorphism of monoids) and a particular dynamical system X is just $\check{X}$ given by

$$\begin{tikzcd} E_h \cdot X \arrow[r, shift left] \arrow[r, shift right] & X \arrow[l, bend left] \end{tikzcd}$$

where E_h is the set of all pairs m'_1, m'_2 for which $h(m'_1) = h(m'_2)$ and $E_h \cdot X$ is the coproduct of E_h copies of X. The detailed properties of $\mathbf{\Delta}_1^{\mathrm{op}} \sum \check{X}$ depend sensitively on the nature of the category $\mathcal{X}$, usually in concrete examples more so than do the detailed properties of the dual construction $\mathbf{\Delta}_1 \prod \hat{X}$, where $\hat{X}$ is the $\mathbf{\Delta}_1 \longrightarrow X$ given by

$$\begin{tikzcd} X \arrow[r, shift left] \arrow[r, shift right] & X^{E_h} \arrow[l, bend left] \end{tikzcd}$$

(where X^{E_h} denotes the product of E_h copies of X, and where, as throughout this bracket, we have followed the usual abuse of notation of using the same letter X to denote also the object $X(0)$ of $\mathcal{X}$ that underlies the action of the monoid M'); thus $\mathbf{\Delta}_1 \prod \hat{X}$ is the $(h \prod X)h$ outside this bracket.]

The full period spectrum of a dynamical system $M' \xrightarrow{X} \mathcal{X}$ can be regarded as a single functor as follows. Say that a period h_2 divides a period h_1 if there exists an endomorphism q of the circle group M such that

$$\begin{tikzcd} & M' \arrow[dl, "h_1"'] \arrow[dr, "h_2"] & \\ M \arrow[rr, "q"] & & M \end{tikzcd} \qquad h_2 = qh_1$$

(then q is unique because h_1 is assumed surjective and q itself is surjective because h_2 is assumed to be). Denote by $\mathcal{Q}$ the category (actually a pre-ordered set) whose objects are the periods and whose morphisms are the q as indicated. By the definitions, if h_2 divides h_1 and if X is (part of) a dynamical system of period h_2, then it is (part of) a dynamical system of period h_1 as well:

$$\begin{tikzcd} & M' \arrow[r, "X"] \arrow[dl, "h_1"'] \arrow[dr, "h_2"] & \mathcal{X} \\ M \arrow[rr] \arrow[urr, dotted, "Y_1"] & & M \arrow[u, dotted, "Y_2"'] \end{tikzcd} \qquad \exists Y_2 \Longrightarrow \exists Y_1$$

Similar reasoning shows that for any dynamical system X, q induces in a functorial way an inclusion

$$(h_2 \textstyle\prod X)h_2 \longrightarrow (h_1 \textstyle\prod X)h_1$$

of the h_2-periodic part of X into the h_1-periodic part of X, whenever q is the reason for h_2 dividing h_1. Thus we get a functor $\bar{X} : \mathcal{Q}^{\mathrm{op}} \longrightarrow \mathcal{X}$ (where $\bar{X}(h) = (h \prod X)h$) that in turn depends functorially on X so that $X \longmapsto \bar{X}$ defines the "periodic pre-spectrum" functor

$$\mathcal{X}^{M'} \xrightarrow{\bar{(\,)}} \mathcal{X}^{\mathcal{Q}^{\mathrm{op}}}$$

from dynamical systems in any sufficiently $\prod$-complete category $\mathcal{X}$ into the category of direct systems in $\mathcal{X}$ indexed by the poset $\mathcal{Q}$ of periods.

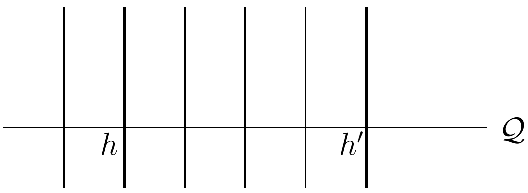
More closely corresponding to the usual notion of spectrum is the following: the weight attached to a given period h is not so much the space $(h \prod X)h$ of states having that period as it is the smaller space of orbits of such states, where in general the notion of orbit space is the left adjoint

$$\mathcal{X}^{M'} \xrightarrow{M' \sum(\,)} \mathcal{X}$$

to the functor induced by the unique $M' \longrightarrow \mathbf{1}$. Since for any h, the space $(h \prod X)h$ of h-periodic points is itself a dynamical system in its own right, hence combining these as h varies through $\mathcal{Q}$ we get a lifted pre-spectrum functor indicated by the dotted arrow below; the latter can be composed with the functor induced by the orbit space functor (upon parameterizing by $\mathcal{Q}$ its inputs and outputs):

$$\begin{tikzcd}
\mathcal{X}^{M'} \arrow[r, dotted] \arrow[dr, "\bar{(\;)}"'] \arrow[rr, bend left, "\bar{\bar{(\;)}}"] & (\mathcal{X}^{M'})^{\mathcal{Q}^{\mathrm{op}}} \arrow[r, "(M'\sum(\;))^{\mathcal{Q}}"] \arrow[d] & \mathcal{X}^{\mathcal{Q}^{\mathrm{op}}} \\
 & \mathcal{X}^{\mathcal{Q}^{\mathrm{op}}} &
\end{tikzcd}$$

to yield what could be called the periodic spectrum $X \longmapsto \bar{\bar{X}}$. The periodic spectrum $\bar{\bar{X}}$ of a dynamical system X can in many cases be pictured as



where the darkness of the line at a period $h \in \mathcal{Q}$ is proportional to the size of the space $M\sum((h\prod X)h)$ of equivalence classes of states of period h, where two states are equivalent if the dynamical action moves one through the other.

There is one other point (not in $\mathcal{Q}$) which may also be considered part of the spectrum, namely the map $M' \longrightarrow \mathbf{1}$ whose corresponding $M'\prod X$ is the space of fixed states of the dynamical system X. If M' is a group, then the fixed point space is usually a subspace of the orbit space, for example if $\mathcal{X}$ is the category of sets and mappings. The same conclusion follows if M' is any commutative monoid. However, for the three element monoid (essentially equivalent, insofar as actions are concerned, to the category $\mathbf{\Delta}_1$ mentioned above) consisting of the morphisms 1, δ_0, δ_1 with the multiplication table $\delta_i\delta_j = \delta_i$ one can find examples of (right) actions $\mathbf{\Delta}_1^{\mathrm{op}} \xrightarrow{X} \mathcal{S}$ on sets (i.e. reflexive graphs) having any given number of fixed points (i.e. vertices) but only one orbit so that the map $\mathbf{\Delta}_1^{\mathrm{op}}\prod X \longrightarrow \mathbf{\Delta}_1^{\mathrm{op}}\sum X$ is not at all a monomorphism.

3. Enhanced algebraic structure in dynamics and logic

Incidentally, the above remark that both groups and commutative monoids share a property not true for all monoids can be made more explicitly algebraic by the following exercise. If $\mathcal{C}$ is a category that is *either* a group, i.e. every morphism in $\mathcal{C}$ has a two-sided inverse, *or* a commutative monoid, then $\mathcal{C}$ acts on its endomorphisms in the following way: for any morphism $X' \xrightarrow{a} X$ in $\mathcal{C}$, and for any endomorphism x of X, we can define an endomorphism x^a of X' such that

$$\boxed{xa = ax^a} \qquad \begin{tikzcd} X' \arrow[loop above, "x^a"] \arrow[r, "a"] & X \arrow[loop above, "x"] \end{tikzcd}$$

and moreover this is an action in the sense that $\boxed{1^a = 1}$ and

$$\boxed{x^{ab} = (x^a)^b} \qquad \begin{tikzcd} X'' \arrow[loop above, dotted] \arrow[r, "b"] & X' \arrow[loop above] \arrow[r, "a"] & X \arrow[loop above, "x"] \end{tikzcd}$$

and even an action by monoid homomorphisms in that

$$\boxed{(xy)^a = x^a y^a}$$
$$\boxed{1^a = 1}$$

for any two endomorphisms x, y of the codomain of a. Of course if $\mathcal{C}$ is itself a monoid then all its morphisms are endomorphisms, and if all morphisms in $\mathcal{C}$ are monomorphisms (a cancellation law) then there could be at most one operation $x, a \longmapsto x^a$ with the crucial property $xa = ax^a$. In the intersection of the two claimed cases, (i.e. for abelian groups) the two formulas for x^a reduce to the same (trivial) thing. If we restrict consideration to the full subcategory of monoids determined by groups and commutative monoids (i.e. the union of the two kinds of objects but containing all four types of homomorphisms between them) we get an example of a *natural operation* on the underlying-set functor that does not extend (from the full subcategory) to all monoids; note that the "algebraic structure" of a full subcategory of an equationally defined algebraic category may have additional operations as well as additional identities between the given operations. In our example the subcategory is the union (made full) of two full subcategories which are themselves equationally defined (in the sense that each consists of *all* algebras satisfying all the identities on all its natural operations). If we take the algebraic category equationally defined by the identities listed above, we get

$$\begin{tikzcd} & \text{Monoids with commutation rule} \arrow[dr, dotted] & \\ \mathrm{Gr} \cup \mathrm{Comm} \arrow[ur, dotted] \arrow[rr, hook, "\text{full}"] \arrow[dr, "\substack{\text{underlying} \\ \text{set functor}}"'] & & \text{Monoids} \arrow[dl] \\ & \text{Sets} & \end{tikzcd}$$

where the descending dotted arrow is a faithful functor which however does not reflect isomorphisms. That is, there exists a monoid (necessarily *not* satisfying the monomorphic cancellation rule) on which there exist two *different* self actions satisfying the commutation rule $xa = ax^a$. Of course, one interest for operations of this sort on a monoid $\mathcal{C}$ is the strong properties it implies for the category $\mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$ of right actions on sets (or any Boolean topos $\mathcal{S}$) in particular with regard to the properties of the intrinsically defined "intuitionistic" negation operator defined on the sub-actions A of any action X by

$$\neg A = \{x \in X \,|\, \forall r \in \mathcal{C}[xr \notin A]\}.$$

Namely, if the monoid $\mathcal{C}$ admits a self-action with a commutation rule as above, then any non-empty A contained in $X = T$ (= the action of $\mathcal{C}$ on itself by right multiplication) satisfies $\neg\neg A = T$. By contrast, if $\mathcal{C}$ is so non-grouplike and non-commutative as the free monoid on two generators, then every "principal" $A = w\mathcal{C} \subset T$ satisfies $\neg\neg A = A$.

4. Functors from school mathematics mutate monoids into ordered sets, and back

Before passing to the discussion of non-autonomous dynamical systems let us point out a crucial example of a functor which occurs in school mathematics: suppose $x \le y \le z$ are non-negative integers or non-negative reals, then the differences $y - x$, $z - y$, $z - x$ are also non-negative and satisfy

$$\boxed{z - x = (z - y) + (y - x)}$$
$$\boxed{0 = x - x}$$

Even though this theorem is a very familiar and useful identity, it cannot be explained by either the monotonicity property nor as a homomorphism property in the usual sense, for in fact it is a structure-preserving property of a process (namely difference) that goes from a pre-ordered set to a monoid. As we have already pointed out, posets and monoids are on the face of it very "opposite" kinds of categories, thus it appears that once we have recognized the necessity for giving a rational status to something as basic as the difference operation discussed above, we are nearly compelled to accept the category of categories, since it is the only reasonable category broad enough to include objects as disparate as posets and monoids *and* hence to include the above difference operator as one of the concomitant structural mutations. To be perfectly explicit, let us denote the relation $x \le y$ (in the poset of quantities in question) by f, and similarly $y \le z$ by g. Define

$$\phi(f) = y - x$$

and similarly $\phi(g) = z - y$ and $\phi(1_x) = 1_0$ for any x where 1_x denotes $x \le x$. 0 may be identified with (the identity morphism of) the unique object of the monoid of quantities where composition is addition. Then ϕ satisfies

$$\phi(gf) = \phi(g)\phi(f)$$
$$\phi(1_x) = 1_0$$

and hence is precisely a functor from a poset to a monoid. We can be still more explicit. In our example, what does $f : x \le y$ mean? We could identify f with the proof that $x \le y$ holds, that is with the non-negative quantity f such that $x + f = y$, or in other words with the morphism f in the monoid for which

$$\begin{tikzcd} & 0 \arrow[dl, "x"'] \arrow[dr, "y"] & \\ 0 \arrow[rr, "f"'] & & 0 \end{tikzcd}$$

commutes. That is f is a morphism in the so called "comma category" $0/\mathcal{C}$ where 0 is the unique object of the monoid $\mathcal{C}$. Of course, f is just the difference, so that ϕ is identified as the forgetful functor

$$\begin{tikzcd} 0/\mathcal{C} \arrow[d, "\phi"'] \\ \mathcal{C} \end{tikzcd}$$

well-defined for any comma category. Note that the comma category will be a poset only in case $\mathcal{C}$ satisfies a cancellation law.

This construction can also be applied to $\mathcal{P}$ = multiplicative monoid whose morphisms are positive whole numbers. Then $1/\mathcal{P}$ is isomorphic to the poset whose *objects* are positive whole numbers ordered by divisibility, and under that identification, the functoriality of the forgetful functor back to the monoid is expressed by

$$n|m \;\&\; m|r \implies \frac{r}{n} = \frac{r}{m} \cdot \frac{m}{n}.$$

There are non-trivial consequences of these observations, for any forgetful functor (such as $1/\mathcal{P} \longrightarrow \mathcal{P}$) on a comma category satisfies the "unique lifting of factorizations" property: $\phi(f) = vu$ implies there are unique $\bar{v}$, $\bar{u}$ such that $f = \bar{v}\bar{u}$, $\phi(\bar{v}) = v$, $\phi(\bar{u}) = u$. If $\mathcal{P}$ is a category satisfying a suitable local finiteness condition then we can define an algebra structure on the set $a(\mathcal{P})$ of all complex-valued functions on the set of morphisms of $\mathcal{P}$ by the convolution formula

$$(\beta * \alpha)(f) = \sum_{ba=f} \beta(b)\alpha(a).$$

Then the "unique lifting of factorizations" property of a functor $\mathcal{P}' \xrightarrow{\phi} \mathcal{P}$ is just what is needed to induce a convolution-preserving homomorphism $a(\mathcal{P}) \longrightarrow a(\mathcal{P}')$. In case $\mathcal{P}$ has cancellation, we thus get an inclusion of the Dirichlet algebra $a(\mathcal{P})$ into the algebra $a(1/\mathcal{P})$ associated to a *poset*; in particular the μ-function (defined as the $*$-inverse, when it exists, of the constantly 1 function) of P becomes the μ-function of the poset $1/\mathcal{P}$. Since the ordering in $1/\mathcal{P}$ is by divisibility, one thus sees how the functions μ, μ^2, etc. in $a(\mathcal{P})$ can be related to counting primes.

5. Non-autonomous systems, monoidal categories, Dedekind completeness, and the construction of the real numbers as an abstraction of dynamical waiting-times

Because right actions (= contravariant functors) of $\mathcal{C}$ are more directly related to the analysis of $\mathcal{C}$ itself (for by the Cayley-Dedekind-Grothendieck-Yoneda lemma, there is a canonical full embedding $\mathcal{C} \hookrightarrow \mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$) we are led to repeat the above discussion also for the comma categories $\mathcal{C}/0 \longrightarrow \mathcal{C}$ whose morphisms are commutative triangles

$$\begin{tikzcd} \bullet \arrow[rr, dotted] \arrow[dr, "x"'] & & \bullet \arrow[dl, "y"] \\ & 0 & \end{tikzcd}$$

In case $\mathcal{C}$ is an additive monoid with cancellation, we would thus naturally write $x \geq y$ to denote the existence of a morphism $x \to y$ in $\mathcal{C}/0$. Since for any object X of a category of the form $\mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$ (where $\mathcal{S}$ is the category of sets and $\mathcal{C}$ is any small category) there is an equivalence of categories

$$\mathcal{S}^{\mathcal{C}^{\mathrm{op}}}/X \cong \mathcal{S}^{(\mathcal{C}/X)^{\mathrm{op}}}$$

where $\mathcal{C}/X$ is the "category of elements" of X, we get in particular for a commutative and cancellative monoid $\mathcal{C}$ that

$$\boxed{\mathcal{S}^{\mathcal{C}^{\mathrm{op}}}/T \xrightarrow{\sim} \mathcal{S}^{(\mathcal{C}/0)^{\mathrm{op}}}}$$

where T is $\mathcal{C}$ acting on itself on the right and $\mathcal{C}/0$ is the poset described above. How is the equation in the box to be interpreted in terms of dynamical systems? Well, T is the simple autonomous dynamical system whose "states" reduce to just the instants of time themselves, and an object of the left hand category is just an arbitrary autonomous dynamical system X equipped with an equivariant morphism $X \longrightarrow T$; (of course many X, for example any periodic one, will not admit any such further structure $X \longrightarrow T$). The nature of the functor from left to right is just to consider the family of fibers (another instance of a $\prod$ construction) X_t of the given map $X \longrightarrow T$ as t varies through T, and whenever $t' \geq t$ the global dynamics of X induces a map $X_t \longrightarrow X_{t'}$, which completes the specification of a functor $(\mathcal{C}/0)^{\mathrm{op}} \longrightarrow \mathcal{S}$ corresponding to X. A natural interpretation of the objects of $\mathcal{S}^{(\mathcal{C}/0)^{\mathrm{op}}}$ is that they are *non-autonomous* dynamical systems, such as arise from the solution of ordinary differential equations which contain "forcing" terms or whose inertial or frictional terms depend on time in some manner (such as usury or heating) external to the self-interaction modeled by the differential equation itself. In general for a non-autonomous system the space of states X_t available at time t may itself depend on t. The (left) adjoint functor is, as already remarked, actually an equivalence of categories; it assigns to any non-autonomous system the single state space

$$\sum_{t \in T} X_t$$

with the single autonomous dynamics naturally induced by the non-autonomous dynamics given as $X_t \longrightarrow X_{t'}$ for $t' \geq t$. In case all the instantaneous state spaces are identifiable as a single Y

$$X_t \cong Y \quad \text{all} \quad t$$

then we see that the associated autonomous system is identifiable with

$$T \times Y$$

and this is just the universally-used construction for making a system autonomous: augment the state space by adding one dimension.

In case $\mathcal{C}$ is a commutative monoid, the category $\mathcal{C}/0$ becomes a "symmetric monoidal" category in the sense that there is a "tensor" *functor*

$$\mathcal{C}/0 \times \mathcal{C}/0 \longrightarrow \mathcal{C}/0$$

induced by the composition (which we will also write as $+$) and having the terminal object $\overset{0}{\underset{0}{\downarrow}}$ of $\mathcal{C}/0$ as "unit" object. In many cases this "tensor" has a right adjoint "Hom" $(\mathcal{C}/0)^{\mathrm{op}} \times (\mathcal{C}/0) \longrightarrow \mathcal{C}/0$ that can be naturally denoted by "subtraction"; it is characterized (assuming $\mathcal{C}$ has cancellation) by the logical equivalence

$$\frac{t + a \geq s}{t \geq s - a}.$$

In the fundamental example where $\mathcal{C}$ is the monoid of non-negative real (or rational) numbers, the meaning of "subtraction" is forced by this adjointness to be *truncated* subtraction. This adjointness persists after *completing*, as discussed below.

For a poset (using $\geq$ for $\rightarrow$) to be complete means that for any functor $M' \xrightarrow{h} M$, $h\sum(\ )$ exists in the poset (these are essentially *infima*) and also that $h\prod(\ )$ exists in the poset (these are essentially arbitrary *suprema*). To "complete" the poset of non-negative reals (or rationals) means roughly to adjoin (reals and) ∞, but since the precise meaning of this in terms of one-sided Dedekind cuts is sensitive to the precise nature of the internal cohesiveness/variation of the "sets" in $\mathcal{S}$, it is fortunate that there is a precise analysis of this process that goes back to the *monoid* $\mathcal{C}$. Namely, define $\mathrm{Pos} \hookrightarrow T$ to be the intersection of all the subdynamical systems P of T which are large enough so that, given any family $f(p) \in T$ indexed by $p \in P$ and satisfying $f(p) + t = f(p + t)$ for all $p \in P$, $t \in T$, there exists a unique $s \in T$ such that $f(p) = s + p$ for all $p \in P$. Then in favorable examples $\mathcal{C}$ is "continuous" (not necessarily complete) in the sense that

$$p \in \mathrm{Pos} \implies \exists p_1, p_2 \in \mathrm{Pos}[p = p_1 + p_2]$$

and Pos itself is the smallest such P. It is then reasonable to consider the subcategory $\mathcal{A} \xrightarrow{i_*} \mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$ of "semicontinuous dynamical systems" defined to consist of those X for which every "possible future" $\mathrm{Pos} \xrightarrow{f} X$ comes from a unique present state, or in other words for which the inclusion $\mathrm{Pos} \hookrightarrow T$ induces a bijection $(T, X) \longrightarrow (\mathrm{Pos}, X)$ between the indicated sets of $\mathcal{C}$-equivariant (= natural) morphisms. Then the inclusion i_* has a left adjoint i^* that in turn has a left adjoint $\mathcal{A} \xrightarrow{i_!} \mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$ including the "identical" category $\mathcal{A}$ as an "opposite" (to i_*) subcategory of $\mathcal{S}^{\mathcal{C}^{\mathrm{op}}}$. This implies that $\mathcal{A}$ is a topos whose truth value object $\Omega_{\mathcal{A}}$ has as elements all the $\mathcal{A}$-subobjects of T; essentially the same is true of $\mathcal{R} \underset{\mathrm{def}}{=} \mathcal{A}/i^*T$, the category of semicontinuous non-autonomous dynamical systems. But its truth values are also the subobjects of the terminal object 1_T because $\mathcal{R}$ is the topos of sheaves on a topological space $[0, \infty]$ topologized in such a way that there are as many open sets as points, with ∞ corresponding to the empty set (or truth value "false"). A dynamical system X in $\mathcal{R}$ has a real number $|X|$ as its support, namely $\inf\{t |\, X_t \neq 0\}$, and this construction can also be viewed as a functor

$$\mathcal{R} \longrightarrow \mathcal{V}$$

from the topos $\mathcal{R}$ to the poset $\mathcal{V} = [0, \infty]$ (the latter having $\geq$ as arrows) and both $\mathcal{R}$, $\mathcal{V}$ have tensor and Hom operations related to addition and truncated subtraction, that are compared by $\mathcal{R} \longrightarrow \mathcal{V}$. Because this discussion can be viewed as a deep objective version of Dedekind's constructions of $\mathcal{V}$, it appears that $\mathcal{R}$ as well as the $\mathcal{A}$ that gives rise to it as a comma category, should be taken seriously.

6. Metric spaces as enriched categories clarify rotations as "homotopy invariants"

It was extensively discussed in a 1973 seminar in Milan [8] that categories enriched in $\mathcal{V}$ are just metric spaces and hence that a detailed mutual clarification of enriched category theory [9] and metric space theory can be exploited. Continuing to take that remark seriously between 1973 and the Bogotá meeting of 1983 led me to several additional points of mutual clarification, that I will now explain. For a metric space A we have

$$\begin{array}{c} 0 = A(a, a) \\ A(a, b) + A(b, c) \geq A(a, c) \end{array} \qquad \text{in } \mathcal{V}$$

for any triple a, b, c of objects (points) of $\mathcal{A}$; in general as explained in the cited article, it is better, both for the theory and for the examples, not to insist on further axioms of finiteness or symmetry. $\mathcal{V}$-functors $A \longrightarrow B$ turn out to be just distance-non-increasing maps, and the $\mathcal{V}$-object of $\mathcal{V}$-natural transformations between two such f, g is easily proved to be

$$B^A(f, g) = \sup_{a \in A} B(fa, ga).$$

More general than $\mathcal{V}$-functors are the $\mathcal{V}$-modules (= $\mathcal{V}$-relations = profunctors) $A \xrightarrow{\phi} B$ which may be viewed as $\mathcal{V}$-functors $B^{\mathrm{op}} \times A \xrightarrow{\phi} \mathcal{V}$; the composition of such arising from the enriched category notion of Trace (= "coend" = tensor product of modules) can be shown in this case to reduce to

$$(\psi \circ \phi)(c, a) = \inf_{b \in B} [\psi(c, b) + \phi(b, a)]$$

for $A \xrightarrow{\phi} B \xrightarrow{\psi} C$.

Note that if ϕ, ψ happened to have only ∞ = false and 0 = true as values, then $\psi \circ \phi$ would reduce to the usual $\exists$, $\wedge$ composition of relations as a special case of the above "least-cost" composition that arises when all of $\mathcal{V} = [0, \infty]$ is admitted. This relationship can be made even more explicit as follows:

Let $\mathcal{V}_0$ be the two-object closed category $\boxed{\text{false} \rightarrow \text{true}}$ with conjunction as tensor and logical implication as internal Hom. Then the inclusion $\mathcal{V}_0 \hookrightarrow \mathcal{V}$ defined in the previous paragraph is actually a closed functor that has a right adjoint $\mathcal{V} \longrightarrow \mathcal{V}_0$ inducing the poset structure on any metric space; that exemplifies the kind of de-enrichment process which is a universal possibility in enriched category theory, giving the underlying ordinary category for categories enriched in any closed category $\mathcal{V}$. However, in our example there is moreover also a left adjoint

$$\mathcal{V} \xrightarrow{\pi_0} \mathcal{V}_0$$

to the inclusion that is also a closed functor. [I have called it π_0 because of the close analogy with other graphs of adjoint functors

$$\begin{array}{ccc} & \xrightarrow{\text{components}} & \\ \mathcal{E} & \xleftarrow{\text{discrete}} & \mathcal{E}_0 \\ & \xrightarrow{\text{points}} & \end{array}$$

that occur, such as $\mathcal{E}$ = simplicial sets, $\mathcal{E}_0$ = sets, and because of the tradition in topology of calling the components functor π_0 because it is sometimes part of a sequence in which the next term is the Poincaré groupoid π_1]. Because π_0 is closed, if we consider for any metric space A the relation defined by

$$\pi_0 A(a, b)$$

we get another (usually very coarse) preordering. This trivial construction is the key to a pedagogical problem as follows:

I wanted to give, for a beginning course in abstract algebra, the basic example of a normal subgroup and a quotient group.

$$\text{Translations} \hookrightarrow \text{Motions} \twoheadrightarrow \text{Rotations}$$

(where each point of the underlying space should give rise to a splitting and hence to a concrete representation of the abstract rotation group as a subgroup of motions, namely those motions which are rotations about the point in question). However, it is desirable to be able to make this basic construction before assuming detailed axioms on the structure of the underlying metric space A. Motions should of course be defined as invertible $\mathcal{V}$-functors f; these will then in particular be distance-preserving

$$A(fa, fb) = A(a, b).$$

But how to define "translations"? At first is seems reasonable to say that they are motions t which moreover satisfy $A(ta, a) =$ constant, for it is not hard to show that the set of such t is "normal", i.e.

$$A(ftf^{-1}a, a) = A(tf^{-1}a, f^{-1}a) = \text{same constant}.$$

But they may *not* form a sub*group* of the motions, for (theoretically) knowing nothing about the structure of A, how could we know which constant should result from composing t_1, t_2 having constants c_1, c_2? In this case the theoretical worry is substantiated by the practical fact that we can construct an example of a five point metric space, indeed embeddable in three-dimensional space as three equidistant points on the rim of a wheel and two judiciously placed points on the axis through the center of the wheel, such that the "translations" as defined are *not* closed under composition. But we shouldn't give up.

Define a translation of an arbitrary metric space A to be an automorphism t such that

$$\sup_{a \in A} A(ta, a) < \infty.$$

Now it is easy (assuming the metric is symmetric) to prove *both* parts of the statement that the translations form a normal subgroup of the motions, as desired. Even better, there are many examples of metric spaces A, such as the ordinary Euclidean plane, which have the property that every translation does in fact move all points through the same distance due to the "searchlight effect": if $A(ta, a) \neq A(tb, b)$ then $A(tx, x)$ can be arbitrarily large, for if there are enough strict translations we can assume that $ta = a$ and construct

$$\begin{tikzcd}
 & b \arrow[r, dash, dashed] \arrow[dd] & x \arrow[dd] \\
a \arrow[ur, dash, dashed] \arrow[dr, dash, dashed] & & \\
 & tb \arrow[r, dash, dashed] & tx
\end{tikzcd}$$

Why is the left adjoint to the inclusion

$$\{\text{false}, \text{true}\} \hookrightarrow [0, \infty]$$

the key to this problem? Because it (in contrast to the right adjoint, which seems to admit as truly possible only those projects that cost no effort) is given by

$$\pi_0(s) = \text{true} \quad \text{iff} \quad s < \infty$$

as is easily verified. Thus it appears we should take seriously the idea that the homotopy theory of metric spaces, that is the 2-functor

$$\mathcal{V}\text{-cat} \longrightarrow \mathcal{V}_0\text{-cat}$$

induced by π_0, is in large part the theory of rotations.

7. Convexity, Isbell conjugacy, closed balls, and radii in enriched categories

The Cayley-Dedekind-Grothendieck-Yoneda embedding

$$A \longrightarrow \mathcal{V}^{A^{\mathrm{op}}}$$

for $\mathcal{V}$-enriched categories A, reduces in the case $\mathcal{V} = [0, \infty]$ under discussion to the fact that $A(-, a)$ is a distance-non-increasing real function for any point of any metric space A, and that the sup-distance between two such functions is equal to the distance between the given points or, more generally, that if $A^{\mathrm{op}} \xrightarrow{f} \mathcal{V}$ is any distance-non-decreasing nonnegative real function (not necessarily of the special form indicated) on the opposite of the metric space A, then

$$\begin{array}{rl}
A(a', a) + fa \geq fa' & \quad \text{all } a' \\
\hline
fa \geq fa' - A(a', a) & \quad \text{all } a' \\
\hline
fa \geq \mathcal{V}(A(a', a), fa') & \quad \text{all } a' \\
\hline
fa \geq \sup\limits_{a'} \mathcal{V}(A(a', a), fa') & \\
\hline
fa \geq \mathcal{V}^{A^{\mathrm{op}}}(A(-, a), f) &
\end{array}$$

but the last inequality is *actually an equality* because the sup is achieved at $a' = a$.

Now in between we can insert the space of *closed subsets* of A

$$A \hookrightarrow \mathcal{F}(A) \hookrightarrow \mathcal{V}^{A^{\mathrm{op}}}$$

by assigning to each $F \hookrightarrow A$ the function

$$F(a') = \inf_{a \in F} A(a', a)$$

which vanishes on (by definition) the closure of F. The sup metric on $\mathcal{V}^{A^{\mathrm{op}}}$, restricted to $\mathcal{F}(A)$ is a refined non-symmetric version of the Hausdorff metric; that is, its symmetrization is the Hausdorff metric, but it itself reduces via $\mathcal{V} \longrightarrow \mathcal{V}_0$ to an ordering which reflects the *inclusion* of the closed sets

$$F_1 \subseteq F_2 \quad \text{as closed sets iff} \quad F_1 \geq F_2 \quad \text{in } \mathcal{V}^{A^{\mathrm{op}}}.$$

Since any $f \in \mathcal{V}^{A^{\mathrm{op}}}$ does have a zero-set, therefore a *right adjoint* to the inclusion

$$\begin{tikzcd} \mathcal{F}(A) \arrow[r, hook] & \mathcal{V}^{A^{\mathrm{op}}} \arrow[l, bend right, dotted, "Z"'] \end{tikzcd}$$

can be constructed. This leads to the idea that the objects in $\mathcal{V}^{A^{\mathrm{op}}}$ might be considered as "refined" closed sets; a point of view already forced upon researchers in constructive analysis and variational calculus by the stringent requirements of their proofs and calculations thus receives also a conceptual support from enriched category theory.

Now an extremely fundamental construction in enriched category theory is the adjoint pair known as *Isbell conjugation*

$$\mathcal{V}^{A^{\mathrm{op}}} \underset{(\ )^{\#}}{\overset{(\ )^{*}}{\rightleftarrows}} (\mathcal{V}^{A})^{\mathrm{op}}$$

which is defined in both directions by a similar formula

$$\begin{aligned} \xi^*(a) &= \mathcal{V}^{A^{\mathrm{op}}}(\xi, a) \\ \alpha^{\#}(a) &= \mathcal{V}^{A}(\alpha, \bar{a}) \end{aligned}$$

where we have followed the usual practice of letting the Yoneda lemma justify the abuses of notation

$$\begin{aligned} A(-, a) &= a \text{ in } \mathcal{V}^{A^{\mathrm{op}}} \\ A(a, -) &= \bar{a} \text{ in } \mathcal{V}^{A}. \end{aligned}$$

The general significance of this construction is somewhat as follows: if $\mathcal{V}$ is the category of sets, or simplicial sets, or (properly construed) topological spaces, or bornological linear spaces, and if the $\mathcal{V}$-category A is construed as a category of basic geometrical figures, then $\mathcal{V}^{A^{\mathrm{op}}}$ is a large category which includes very general geometrical objects that can be probed with help of A, but that would inevitably come up in a thorough study of A itself. On the other hand, $\mathcal{V}^{A}$ includes very general algebras of quantities whose operations (à la Descartes) mirror the geometric constructions and incidence relations in A itself. Then the conjugacies are the first step toward expressing the duality between space and quantity fundamental to mathematics: $(\ )^*$ assigns to each general space the algebra of functions on it, whereas $(\ )^{\#}$ assigns to each algebra its "spectrum" which is a general space. Of course neither of the conjugacies is usually surjective; the second step in expressing the fundamental duality is to find subcategories with reasonable properties that still include the images of the conjugacies:

$$\begin{tikzcd}
 & A \arrow[dl] \arrow[dr] & \\
\mathcal{X} \arrow[rr, shift left] \arrow[d, hook] & & \mathcal{A}^{\mathrm{op}} \arrow[ll, shift left] \arrow[d, hook] \\
\mathcal{V}^{A^{\mathrm{op}}} \arrow[rr, shift left] \arrow[urr, dotted] & & (\mathcal{V}^{A})^{\mathrm{op}} \arrow[ll, shift left] \arrow[ull, dotted]
\end{tikzcd}$$

For example if $\mathcal{V}$ is the category of sets and A is a small category with finite products, we could take $\mathcal{X}$ to be the *topos* of "canonical sheaves" on A, and $\mathcal{A}$ to be the *algebraic category* of all finite-product-preserving functors.

Our poset $\mathcal{V} = [0, \infty]$ appears rather puny compared to the grand examples mentioned in the previous paragraph, but seriousness eventually leads us to try the Isbell conjugacy on it as well, and in particular to ask which closed sets $F \in \mathcal{F}(A)$ are fixed by the composed Isbell conjugacies for a metric space A.

$$\begin{tikzcd} \mathcal{F}(A) \arrow[r, hook] & \mathcal{V}^{A^{\mathrm{op}}} \arrow[l, bend right, "Z"'] \arrow[loop above, "(\,)^{*\#}"] \arrow[r, shift left, "(\,)^{*}"] & (\mathcal{V}^{A})^{\mathrm{op}}. \arrow[l, shift left, "(\,)^{\#}"] \end{tikzcd}$$

Note that from the adjointness, $\xi \geq \xi^{*\#}$ for all ξ in $\mathcal{V}^{A^{\mathrm{op}}}$, so that the idempotent operation $(\,)^{*\#}$ gives a kind of lower envelope for functions and hence a kind of hull $Z(\xi) \subseteq Z(\xi^{*\#})$ for the corresponding closed zero-sets; the question is what kind of hull?

To answer this it is relevant to explicitly introduce the following parameterized family of special elements of $\mathcal{V}^{A^{\mathrm{op}}}$:

$$\mathcal{V}^{\mathrm{op}} \otimes A \xrightarrow{B} \mathcal{V}^{A^{\mathrm{op}}}$$

defined by

$$B(r,c)(a') = \mathcal{V}(r, A(a',c)),$$

where $\mathcal{V}(x,y) = y - x$ is the truncated subtraction in our example $\mathcal{V} = [0,\infty]$; but for any closed category $\mathcal{V}$ denotes its internal Hom, and where the tensor product of $\mathcal{V}$-categories (defined in our case as the metric which is the *sum* of the coordinate distances) is used instead of the cartesian product because it guarantees that B itself is also a $\mathcal{V}$-functor. The letter B stands for *closed ball* of a given radius and center since

$$0 \geq B(r,c)(a') \quad \text{iff} \quad r \geq A(a',c).$$

An amazing example of the seriously-pursued study of the mutual relationship of a key example with general philosophy is that these "closed balls" occur and are useful over many apparently quite diverse closed categories $\mathcal{V}$, for example in homological algebra. Moreover, let us denote by $\bigcap$ the supremum operation on $\mathcal{V}^{A^{\mathrm{op}}}$

$$\left(\bigcap \xi_i\right)(a') = \sup_i \xi_i(a')$$

since that is what it corresponds to under the operation Z of taking zero-sets.

With the above-introduced notation, we see that $\xi = \xi^{*\#}$ iff $\xi^{*\#} \geq \xi$, and also that

$$\begin{aligned} \alpha^{\#}(a'') &= \mathcal{V}^{A}(\alpha, \bar{a}'') \\ &= \sup_{a'} \mathcal{V}(\alpha(a'), A(a'', a')) \end{aligned}$$

so that

$$\alpha^{\#} = \bigcap_{a'} B(\alpha(a'), a')$$

is the intersection of all the closed balls, centered at all points a', of specified radius $\alpha(a')$. [Naturally this construction, under the name of "end" or "center", also comes up for general $\mathcal{V}$]. Now what if α is of the form ξ^*? The number $\xi^*(a')$ is a radius (about an arbitrary center a') that ξ somehow prefers:

$$\xi^*(a') = \sup_a \mathcal{V}(\xi(a), A(a, a'))$$

so that

$$\begin{aligned} r \geq \xi^*(a') \quad &\text{iff} \quad r + \xi(a) \geq A(a, a') \quad \text{for all } a \\ &\text{iff} \quad \xi(a) \geq A(a, a') - r \quad \text{for all } a \\ &\text{iff} \quad \xi \geq B(r, a'). \end{aligned}$$

Thus (since throwing some larger balls into the family won't change the intersection)

$$\xi^{*\#} = \bigcap \{B(r, a') |\, \xi \geq B(r, a')\}$$

is a geometrical description of the double-conjugate lower envelope of ξ. In the case where ξ is fixed under the other idempotent operation on $\mathcal{V}^{A^{\mathrm{op}}}$ coming from Z, i.e. if ξ is determined by its closed zero-set F as explained before

$$\xi(a'') = \inf_{a \in F} A(a'', a)$$

then the condition $\xi \geq B(r, a')$ reduces to

$$\begin{aligned} \forall a \in F\ \forall a'' \quad A(a'', a) &\geq B(r, a')(a'') \\ &= A(a'', a') - r \end{aligned}$$

i.e. to

$$A(a'', a) + r \geq A(a'', a') \quad \text{for} \quad a \in F,\ \text{arbitrary } a''.$$

In particular this means that $F \subseteq$ the ball ($=$ zero set of) $B(r, a')$, so we see that

$$F \subseteq F^{*\#} \subseteq \text{the intersection of all closed balls that contain } F.$$

But in fact that intersection of balls is equal to $F^{*\#}$, since the right adjointness of Z says in particular that for any closed set F and any ball we have the equivalence

$$\frac{F \subseteq ZB(r, a')}{F \geq B(r, a')}.$$

Putting it differently, since we have in general that

$$A(a'', a) + A(a, a') \geq A(a'', a'),$$

if $a \in F$ implies $0 \geq B(r, a')(a)$, i.e. $r \geq A(a, a')$, then for any a'' and any $a \in F$

$$A(a'', a) + r \geq A(a'', a) + A(a, a') \geq A(a'', a')$$

so that

$$A(a'', a) \geq B(r, a')(a'')$$

meaning that

$$F(a'') \geq B(r, a')(a'') \quad \text{as functions of } a'',$$

because the right hand side is independent of a and the function F was defined as the *infimum* (adjointness of "$\sum$ along a diagonal"). Thus

$$\begin{aligned} F^{*\#} &= \text{intersection of all closed balls which contain } F \\ &= \text{closed convex hull of } F, \end{aligned}$$

at least in many metric spaces of geometric importance, and for the others the proposed use of the term "convex" for those closed sets F satisfying $F = F^{*\#}$ should be as good or better than other proposals because of the apparent importance of adjointness in calculations.

As noticed above, the numbers $\xi^*(a')$ have in many connections the concrete significance of radii for balls about a'. Using the obvious "direct limit" functor $\mathcal{V}^A \xrightarrow{\inf} \mathcal{V}$ (which exists because $A \longrightarrow 1$ is a $\mathcal{V}$-functor), we can define a single "radius" for ξ itself by applying the composite

$$\begin{tikzcd} \mathcal{V}^{A^{\mathrm{op}}} \arrow[r, "(\,)^*"] \arrow[rr, bend right, "\mathrm{rad}"'] & (\mathcal{V}^A)^{\mathrm{op}} \arrow[r, "\inf^{\mathrm{op}}"] & \mathcal{V}^{\mathrm{op}} \end{tikzcd}$$

In particular, the radius of a closed set $F \in \mathcal{F}(A) \subseteq \mathcal{V}^{A^{\mathrm{op}}}$ is

$$\mathrm{rad}(F) = \inf\{r \mid \exists a' \, [F \subseteq ZB(r, a')]\}$$

where the candidate centers a' are not themselves necessarily in F. (Note that to say we have a functor $\mathcal{F}(A) \longrightarrow \mathcal{V}^{\mathrm{op}}$ from a poset to $\mathcal{V}^{\mathrm{op}}$ is to say that the values *increase* as the objects in $\mathcal{F}(A)$ increase). The radius is a more functorial quantity than the habitually-used "diameter"; for a symmetric metric space there is the estimate diam ≤ 2 rad whereas for certain reasonably well behaved spaces one may also have a converse estimate rad $\leq$ diam. There is a strong tendency for the radius to be realized at a *unique* center a'. Note that $\mathrm{rad}(F^{*\#}) = \mathrm{rad}(F)$.

8. Geodesic remetrization as an adequacy comonad

Now let us say a few words about the important role of *paths* in metric spaces. The comma categories $d/\mathcal{V} = [0, d]$ with their canonical $d/\mathcal{V} \hookrightarrow \mathcal{V}$ have a retraction given again by "double dualization":

$$x \longmapsto \mathcal{V}(\mathcal{V}(x, d), d) = d - (d - x)$$

is always in the interval; moreover single d-dualization is an invertible duality

$$(d/\mathcal{V}) \rightleftarrows (d/\mathcal{V})^{\mathrm{op}}$$

when restricted, provided $d < \infty$. Denote by $\mathcal{V}' \subset \mathcal{V}$ all those d such that $\pi_0(d) = \text{true}$, i.e. for which $d < \infty$. Let $\mathcal{V}(d)$ be the *symmetric* metric space determined by $d/\mathcal{V}$, so that in particular $d - (\ )$ becomes (for $d < \infty$) a self-motion of the interval $\mathcal{V}(d)$, because symmetrizing is a functor. Indeed, $d \longmapsto \mathcal{V}(d)$ defines a functor

$$\mathcal{V}^{\mathrm{op}} \longrightarrow \mathcal{V}\text{-cat}$$

using $d' \geq d \Longrightarrow \mathcal{V}(d) \subseteq \mathcal{V}(d')$, but also a functor $\mathcal{V}' \longrightarrow \mathcal{V}$-cat by invoking the retractions. The essential question we want to understand is: to what extent is the structure of an arbitrary metric space $A \in \mathcal{V}$-cat analyzable in terms of the paths (= $\mathcal{V}$-functors)

$$\mathcal{V}(d) \longrightarrow A$$

of duration d, with d variable? Since all constants are paths, such analysis easily maintains the points of A. If there is such a path, passing a_0 at time 0 and a_1 at time d, then

$$d \geq A(a_0, a_1).$$

This leads to the idea of *geodesic distance*:

$$(\Gamma A)(a_0, a_1) = \inf\{d \in \mathcal{V}' \,|\, \exists\, \sigma : V(d) \longrightarrow A \text{ with } \sigma(0) = a_0, \sigma(d) = a_1\}$$

which can be seen to be a new metric since

$$\begin{tikzcd}
\mathcal{V}(0) \arrow[r] \arrow[d] & \mathcal{V}(d') \arrow[d] \\
\mathcal{V}(d) \arrow[r, hook] & \mathcal{V}(d+d')
\end{tikzcd}$$

is a pushout in $\mathcal{V}$-cat. That is, if σ is a path in A of duration d, and σ' of duration d', and if

$$\sigma(d) = \sigma'(0) = a_1$$

we must show that if $d \geq t$, $d + d' \geq s \geq d$, then

$$A(\sigma(t), \sigma'(s-d)) \overset{?}{\leq} s - t.$$

But we have

$$\begin{aligned}
A(\sigma(t), \sigma'(s-d)) &\leq A(\sigma(t), a_1) + A(a_1, \sigma'(s-d)) \\
&= A(\sigma(t), \sigma(d)) + A(\sigma'(0), \sigma'(s-d)) \\
&\leq d - t + s - d = s - t
\end{aligned}$$

because $d \geq t$ and $s \geq d$. The other cases of the $\mathcal{V}$-functorality of $(d + d') \xrightarrow{\sigma' * \sigma} A$ are obvious. Thus (using the logicians' symbol for "proves")

$$\left.\begin{array}{r} \sigma \Vdash d \geq \Gamma A(a_0, a_1) \\ \sigma' \Vdash d' \geq \Gamma A(a_1, a_2) \end{array}\right\} \Longrightarrow \sigma' * \sigma \Vdash d + d' \ \geq \ \Gamma A(a_0, a_2).$$

Because direct limits in metric spaces are essentially computable in terms of direct limits of sets and infima of distances, it can be seen that

$$\Gamma A = \varinjlim_{\sigma \in \mathcal{V}'/A} \mathcal{V}(\operatorname{dom} \sigma)$$

is an endofunctor of $\mathcal{V}$-cat having a natural distance-non-increasing map

$$\begin{array}{c} \Gamma A \\ \downarrow \\ A \end{array}$$

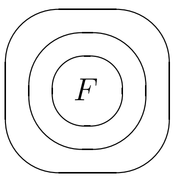
back to the identity functor; it is in fact the "adequacy comonad" of $\mathcal{V}' \longrightarrow \mathcal{V}$-cat, a notion defined for any small subcategory of any category having direct limits. Note for example that path-connectedness of A becomes π_0-connectedness of ΓA, for if two points of A are not connectable by a path, then (empty inf) their geodesic distance in ΓA is infinite.

9. History still has much to teach, and raises fresh questions

Finally, I believe that we should take seriously the historical precursors of category theory, such as Grassmann, whose works contain much clarity, contrary to his reputation for obscurity. For example, I read there a statement of the sort "diversity can be added" whereas "unity can be multiplied" together with quite convincing geometrical and algebraic substantiation of these principles. The first of them suggests the following:

If $\mathcal{F}(A)$ is a *poset* of parts of a space and $\mathcal{C}$ is a suitable additive *monoid*, then the amount by which F must be extended to achieve a diverse $G \supseteq F$ might be given by $\mu(F, G) \in \mathcal{C}$; that should again give a functor $\mathcal{F}(A) \longrightarrow \mathcal{C}$ in that

$$F \subseteq G \subseteq H \Longrightarrow \mu(F, H) = \mu(F, G) + \mu(G, H).$$



Of course, if $0 \in \mathcal{F}(A)$ and if $\mathcal{C}$ has cancellation then $\mu(G, H)$ will be determined by $\mu H \underset{\text{def}}{=} \mu(0, H)$ and μF. To further express the "quantitative" nature of such a measurement of extension μ, we can consider the further condition that μ "only depends on the difference". Since the crucial property of difference is again that it is a Hom, adjoint this time to union as $\otimes$,

$$\frac{S \supseteq H \setminus G}{G \cup S \supseteq H}$$

we can express this invariance of the functor μ using only the union structure on $F(A)$:

$$\{S|\, G \cup S \supseteq H\} = \{S|\, G_1 \cup S \supseteq H_1\} \Longrightarrow \mu(G, H) = \mu(G_1, H_1).$$

It would be interesting to determine for which upper semilattices $\mathcal{F}$, 0, $\cup$ there exists a commutative monoid $\mathcal{C}$ and a functor μ with this invariance property which moreover has the "unique-lifting-of-factorizations" property previously discussed, i.e., for which the object X in a process of intermediate expansion $F \subseteq X \subseteq H$ is uniquely determined by sufficiently many quantitative measurements of its size. For example, area alone is not sufficient but $\mathcal{C}$ can be a cartesian product of many different kinds of quantities. Again adjointness makes at least an initial contribution to the problem: for each $\mathcal{F}$ there is a well-defined universal $\mathcal{C}$ and μ, so that one need only study the lifting question for that.

10. Dialectical relation between teaching and research can be exemplified by metrical development of enriched category theory

We have seen that the application of some simple general concepts from category theory leads from a clarification of basic constructions on dynamical systems to a construction of the real number system with its structure as a closed category; applied to that particular closed category, the general enriched category theory leads inexorably to embedding theorems and to notions of Cauchy completeness, rotation, convex hull, radius, and geodesic distance for arbitrary metric spaces. In fact, the latter notions present themselves in such a form that the calculations in elementary analysis and geometry can be explicitly guided by the experience that is concentrated in adjointness. It seems certain that this approach, combined with a sober appreciation of the historical origin of all notions, will apply to many more examples, thus unifying our efforts in the teaching, research, and application of mathematics.

References

[1] Eilenberg, S. and Mac Lane, S., *General Theory of Natural Equivalences.* Trans. Amer. Math. Soc. **58**(1945), 231–294

[2] Grothendieck, et. al., *Theorie des Topos et Cohomologie Etalé des Schemas.* Springer Lecture Notes in Mathematics 269, 270, 1972.

[3] Lawvere, F.W. (editor), *Toposes, Algebraic Geometry and Logic*. Springer Lecture Notes in Mathematics **274**, 1972.

[4] Lawvere, Maurer, and Wraith (editors), *Model Theory and Topoi*. Springer Lecture Notes in Mathematics **445**, 1975.

[5] Banaschewski (editor), *Categorical Aspects of Topology and Analysis*. Springer Lecture Notes in Mathematics **915**, 1982.

[6] Lawvere and Schanuel (editors), *Categories in Continuum Physics*. Springer Lecture Notes in Mathematics **1174**, 1986.

[7] Joyal, A., *Une théorie combinatoire de séries formelles*, Advances in Mathematics, **42**(1981), 1–81.

[8] Lawvere, F.W., *Metric Spaces, Generalized Logic, and Closed Categories*, Seminario Matematico e Fisico di Milano **43**(1973), 135–166, and Reprints in Theory and Applications of Categories, no. 1(2002), 1–37.

[9] Kelly, G.M., *Basic Concepts of Enriched Category Theory*. London Mathematical Society Lecture Notes Series **64**, Cambridge University Press, 1982 and Reprints in Theory and Applications of Categories, (2005).

University at Buffalo, Department of Mathematics
244 Mathematics Building, Buffalo, New York 14260-2900
Email: `wlawvere@buffalo.edu`

This article may be accessed from `http://www.tac.mta.ca/tac/reprints` or by anonymous ftp at
`ftp://ftp.tac.mta.ca/pub/tac/html/tac/reprints/articles/8/tr8.{dvi,ps}`

Reprints in Theory and Applications of Categories, No. 9, 2005, pp. 1–7.

CATEGORIES OF SPACES MAY NOT BE GENERALIZED SPACES AS EXEMPLIFIED BY DIRECTED GRAPHS

F. WILLIAM LAWVERE

Author commentary: When this paper was distributed at the 1986 international category theory meeting in Cambridge, its reception was mixed. So when Xavier Caicedo, the academic editor of the "Revista Colombiana de Matemáticas" proposed to publish it together with the proceedings of the 1983 Bogotà Workshop, I was pleased to accept; thanks to his continued generosity in granting copyright permission, it can now be reprinted in *TAC*.

The simple idea at the core of this paper has not yet been much pursued by workers in topos theory, even though I have tried in several later publications to point out its importance to various branches of mathematics, where those colleagues with greater knowledge and ability could, I believe, contribute.

Already in SGA4, Grothendieck had made a major advance on this problem, in a series of ten exercises for which he quite justly awarded himself "une médaille de chocolat". His construction, generalizing Giraud's gros topos of a topological space, is in terms of sites and has apparently not yet been assimilated well enough to suggest a corresponding invariant description.

In a broad sense, any topos over a base S can be conceived as a "generalized space"; even the basic facts that it may have a proper class of points, or that these points may form a category that does not reduce to a poset, do not prevent this imagination from being useful. For example, not only terminology such as "connected", taken from geometry, but even far-reaching constructions such as distributions, the "space" of distributions studied by Bunge and Carboni, and supports of distributions seen as singular coverings, studied by Bunge and Funk, are partly motivated by that exuberant generality. More precise results depend on limiting the generality, even on taking into explicit account some opposite qualitative distinctions within the generality.

The completeness theorems of Barr, Deligne, Diaconescu, Freyd, Joyal, Makkai, Reyes, and others, summed up by Johnstone in 1983 [J], in particular involve constructions showing that any S-topos (for example conceived as the infinitary positive theory of some kind of structure) has "enough" morphisms (= points, models) from other S -toposes of a very special kind, localic (i.e. with poset site) or sometimes groupoidal. Extending those methods, Tierney and Moerdijk, in collaboration with Joyal, showed by 1990 that every

---

Thanks to Christoph Schubert for his expert transcription into TEX.
Received by the editors 2004-11-02.
Transmitted by M. Barr, R. Rosebrugh and R. F. C. Walters. Reprint published on 2005-03-23.
2000 Mathematics Subject Classification: 00A30, 14F20, 18B25, 18B30, 18B40, 55U40, 81P05.
Key words and phrases: topos of spaces, directed graphs, sheaves, idempotents, Mengen.
The article originally appeared in Revista Colombiana de Matemáticas, XX (1986) 179-186.

S-topos agrees with respect to certain cohomological invariants (not only with respect to internal logic) with special toposes deserving of being called generalized spaces in a more concrete sense.

That concrete sense was already much alive before category theory was made explicit: the invention of the cohomology of groups stemmed from discoveries by Hopf, Hurewicz, Steenrod, and others. Today those discoveries could be summed up by noting that both classical spaces and abstract groups are fully embedded in a single larger category (on which cohomology is defined) and participate there in mixed exact sequences embodying fundamental groups, universal covers, and even more particular geometric information. This single larger category could be taken to be the one of all S-toposes; however, a much smaller category suffices for that, namely the category of étendu (i.e. of those toposes having a site consisting entirely of monomorphisms). Grothendieck established the étendu as generalized spaces in a concrete sense, by qualitatively extending the classical work through the construction of more informative quotients of spaces.

But then, around 1960, Grothendieck made a remarkable pair of constructions. Workers in topos theory have yet to come to grips with the specific content of those two constructions, in spite of the 35 year development of a simplified general methodology. The petit étale topos of a scheme, brilliantly overcoming the lack of an inverse function theorem, is clearly a generalized space in a very concrete sense; yet it is not an étendu, and so it's specifically topos-theoretic particularity begs for clarification. On the other hand, the techniques of construction for analytic spaces [G2] involve embedding them in a single large topos qualitatively different from the sheaves on any concrete kind of generalized space.

A general topos incorporating all spaces of a given kind is not a new idea either. Since 1950 the topos of simplicial sets has been widely used, and in fact this combinatorial example served as a kind of model for the discussion in the present paper. That it has a connected object with distinct points concretely implies the axioms 1 and 2. The gros Zariski topos and other environments for algebraic and smooth geometry also enjoy that feature.

Johnstone's topological topos and my bornological topos (sheaves for finite coverings of countable sets) are intuitively also "general" in content (as opposed to "particular") and yet do not satisfy my axioms; but I share with Grothendieck the belief that a suitable development of tame topology will avoid phenomena such as Peano's space filling curves and the non-discreteness of anti-connected spaces. (According to a Joyal-Johnstone result in the 1992 book by Mac Lane and Moerdijk, the topos-theoretic match between continuous and combinatorial topology requires that the continuous interval be totally ordered, and it seems that a tame interval might satisfy that even though the classical one, taken as a site in its own right, does not.)

It was fortunate that the simple enterprise of clarifying the role of two kinds of graphs provided sufficient illustration of the basic distinction here discussed. A feature apparent in this example, namely that the "sheaves" on a particular space B in a general topos are contained in a "particular" topos (obtained by collapsing idempotents in a site), I later found in many other examples in geometry.

Abstract. Axioms are proposed for the distinctive internal connectedness of a topos that models all spaces of a "general" combinatorial, algebraic, or smooth kind. It is shown that sheaves on any particular space, like representations of any particular group, do not satisfy these axioms. For each object B in a general topos, a topos of the "opposite" or "particular" kind (containing sheaves and unramified coverings of B) is constructed. All these features are illustrated by the simple example of reflexive directed graphs.

It has long been recognized [G1], [L] that even within geometry (that is, even apart from their algebraic/logical role) toposes come in (at least) two varieties: as spaces (possibly generalized, treated via the category of sheaves of discrete sets), or as categories of spaces (analytic [G2], topological [J], combinatorial, etc.). The success of theorems [J'] which approximate toposes by generalized spaces has perhaps obscured the role of the second class of toposes, though some explicit knowledge of it is surely necessary for a reasonable axiomatic understanding of toposes of C^∞ spaces or of the topos of simplicial sets. Perhaps some of the confusion is due to the lack of a stabilized definition of morphism appropriate to categories of spaces in the way that "geometric morphisms" are appropriate to generalized spaces.

There are certain properties which a topos of spaces often has; a wise selection of these should serve as an axiomatic definition of the subject. While we have not achieved that goal yet, we list some important properties and show that these properties cannot be true for a "generalized space" of the localic or groupoid kind.

We consider a topos $\mathcal{E}$ defined over another topos $\mathcal{S}$. The latter need not be the category of abstract sets, though it will often be Boolean. In many cases it is instructive to think of $\mathcal{S}$ as *derived from* $\mathcal{E}$ (rather than the other way around), as Cantor derived "cardinal numbers" (= abstract sets) from "Mengen" (= sets with topological or similar structure, as they arise in geometry and analysis). Indeed $\mathcal{S}$ can be viewed as a sheaf topos in $\mathcal{E}$, for an essential topology:

Axiom 0. $\mathcal{E} \longrightarrow \mathcal{S}$ is local; $\Gamma^* \dashv \Gamma_* \dashv \Gamma^!$.

The Γ^* may be considered as the inclusion of *discrete* spaces $\mathcal{S}$ into "all" spaces $\mathcal{E}$, whereas the sheaf inclusion $\Gamma^!$ may be considered as the inclusion of *codiscrete* or chaotic spaces into $\mathcal{E}$; that these inclusions have the same domain category $\mathcal{S}$ may be summed up in Hegelian fashion by "pure becoming is identical with non-becoming".[1]

Of course, there are some spatial toposes which satisfy axiom 0, although they are extremely special since Γ_* is the fiber-functor for a canonically-defined extremal point of $\mathcal{E}$; for example, the Zariski spectrum of a *local* ring does admit such a point $\Gamma^!$. On the other hand, the topos of G-sets for a groupoid G cannot satisfy axiom 0.

Our further axioms will be stated in terms of a further left adjoint $\Pi_0 = \Gamma_!$ assigning to each space a discrete space of components.

[1] That is, completely random motion, as a category in itself, is indistinguishable from immobility, as a category in itself, even though they are of course completely different (except for 0, 1) as subcategories of the category of spaces (= frames for continuous motion).

Axiom 1. $\mathcal{E} \longrightarrow \mathcal{S}$ is *essential*, that is $\Gamma_! \dashv \Gamma^*$ exists, but moreover we require that it preserves finite products

$$\Gamma_!(X \times Y) \xrightarrow{\sim} \Gamma_!(X) \times \Gamma_!(Y)$$
$$\Gamma_!(1) \xrightarrow{\sim} 1$$

for all X, Y in $\mathcal{E}$.

The axiom is necessary for the naive construction of the homotopic passage from quantity to quality; namely, it insures that (not only Γ_* but also) $\Gamma_!$ is a closed functor, thus inducing a second way of associating an $\mathcal{S}$-enriched category to each $\mathcal{E}$-enriched category

$$\mathcal{E}\text{-cat} \xrightarrow{[\ ]} \mathcal{S}\text{-cat}.$$

For example, $\mathcal{E}$ itself as an $\mathcal{E}$-enriched category gives rise to a homotopy category in which

$$[\mathcal{E}](X, Y) = \Gamma_!(Y^X).$$

This product-preserving property of $\Gamma_!$ is well-known to be false in the group case, where $\Gamma_!(G \times G) = n$, where $n = \#G$, whereas $\Gamma_!(G) = 1$. Again, it *can* hold for some (extremely special) spaces: For a topos $\mathcal{E}$ localic over $\mathcal{S}$, $\Gamma_!$ is left exact if only it preserves products, and hence there is again a canonically defined point, at the opposite extreme; for example, the Zariski spectrum of an integral domain admits a product preserving $\Gamma_!$. If $\mathcal{S}$ is an "exponential variety" in $\mathcal{E}$, then $\Gamma_!(\Gamma^*(A) \times Y) \xrightarrow{\sim} A \times \Gamma_!(Y)$ which is, however, only a fragment of our axiom 1. It is at this point that the constructions of generalized spaces which "cover" a given topos insofar as the "internal logic" is concerned, fail to preserve the structure of a "topos of spaces". (For covering as an "exponential variety" would preserve our axiom 2).

Axiom 2. $\Gamma_!(\Omega) = 1$, where Ω is the truth-value object in the topos $\mathcal{E}$ of spaces.

Since Ω has the structure of a monoid with zero, in the presence of axiom 1 its being connected (axiom 2) implies its being contractible in that

$$[\mathcal{E}](X, \Omega) = 1$$

for all X in $\mathcal{E}$, and hence that $X \longrightarrow \Omega^X$ is a natural embedding of every space into a contractible space; moreover, any retract, such as Ω_j for a topology j, (for example the Boolean algebra $\Omega_{\neg\neg}$) is also contractible. Of course axiom 2 cannot be true of a Boolean topos since $\Gamma_!$ preserves any sum such as $1 + 1$.

PROPOSITION 1. Axioms 1 and 2 cannot both be true for a localic topos $\mathcal{E}$ over sets $\mathcal{S}$.

Proof. Axiom 1 implies that $\Gamma_!$ preserves pullbacks in the localic case. In any case there is a pullback diagram

$$\begin{array}{ccc} 2 & \longrightarrow & 1 \\ \downarrow & & \downarrow \\ \Omega & \longrightarrow & \Omega \end{array}$$

in $\mathcal{E}$ where $2 = 1 + 1$. Thus applying $\Gamma_!$ we get an impossible pullback diagram

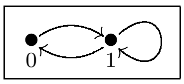
$$\begin{array}{ccc} 2 & \longrightarrow & 1 \\ \downarrow & & \downarrow \\ 1 & \longrightarrow & 1 \end{array}$$

in $\mathcal{S}$. ∎

The above axioms (incomplete though they may be) enable us to make some rather sharp distinctions. For example, there are (at least) two distinct toposes commonly referred to as "the category of directed graphs" and even commonly considered to be more or less of the same value since, for example, the notion of "free category" generated by either kind of graph makes sense. The two are

$$\mathcal{S}^{\Delta_1^{\mathrm{op}}} \qquad \mathcal{S}^{\cdot\rightrightarrows\cdot}$$

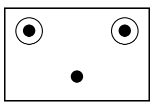
where Δ_1 is the three-element monoid of all order-preserving endomaps of the two-element linearly ordered set [1]; splitting the idempotents shows that Γ_* is essentially representable and hence $\Gamma^!$, the notion of codiscrete graph, exists for $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$, though not for $\mathcal{S}^{\cdot\rightrightarrows\cdot}$. However, the one-dimensional simplicial sets $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$ and the "irreflexive" graphs $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ differ already in regard to axiom 1: the functor $\Gamma_!$ is in either case just the coequalizer of the structural maps, but, as is well-known, reflexive coequalizers preserve products, whereas irreflexive coequalizers do not. The subobject classifier for $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$ has five elements



and is obviously connected. A similar statement is true for $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ but the foregoing remarks are sufficient to show the following.

PROPOSITION 2. The topos $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$ satisfies the axioms 0, 1, 2 for a "topos of spaces", whereas the topos $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ of diagram schemes does not satisfy 0 or 1.

In fact, at least two arguments can be given to show that $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ definitely belongs to the other variety of toposes, namely that it is in fact a simple example of a generalized space. For one thing, the category $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ of irreflexive graphs is an étendue; in fact, it is locally localic in an illuminating manner: Consider the space



which has three points and five open sets. A sheaf on this space consists of a set E of global sections, two sets V_0, V_1 of sections over the two open points, and two restriction maps $E \longrightarrow V_0$, $E \longrightarrow V_1$.

If we consider the two-element group acting on the space by interchanging the two open points, we can take the "quotient" (descent) in the 2-category of toposes by the equivalence relation associated to this action; this has the effect of forcing $V_0 = V_1$, but allowing the two restrictions $E \rightrightarrows V$ to remain different. Conversely, there is an object A in $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ such that $\mathcal{S}^{\cdot\rightrightarrows\cdot}/A$ is (the topos of sheaves on) the three point space, showing explicitly the local homeomorphism of the two toposes.

Another aspect of the status of $\mathcal{S}^{\cdot\rightrightarrows\cdot}$ as a generalized space is revealed by its relationship to the category of spaces $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$. If $\mathcal{E}$ over $\mathcal{S}$ is a topos of "spaces", then each object B of $\mathcal{E}$ should be capable of serving as a domain of variation in its own right; in particular it should have sense to speak of abstract sets varying over B, giving rise to a topos $\mathcal{S}(B)$ (usually a subcategory of $\mathcal{E}/B$), which should be an example of a generalized space ("should be" since we don't yet have axioms strong enough to capture the special nature of generalized spaces, yet general enough to include the classical petit étale example!). In case $B \in \mathcal{E} = \mathcal{S}^{\Delta_1^{\mathrm{op}}}$ is a graph, one reasonable definition of

$$\mathcal{E}/B \supset \mathcal{S}(B)$$

is simply to take all $E \longrightarrow B$ which have discrete fibers in the sense that

$$\begin{tikzcd} \Gamma^*\Gamma_* E \arrow[r] \arrow[d] & E \arrow[d] \\ \Gamma^*\Gamma_* B \arrow[r] & B \end{tikzcd}$$

is a pullback. These might be called "B-partite graphs" generalizing the bipartite graphs which arise as the special case where

$$B = \boxed{\bullet \rightleftarrows \bullet}$$

The toposes $\mathcal{S}(B)$ are *all* étendues, and behave with excellent functorial comportment under morphisms $B \longrightarrow B'$ in the topos of spaces $\mathcal{E} = \mathcal{S}^{\Delta_1^{\mathrm{op}}}$; thus they seem to embody well one idea of the generalized spaces associated to objects of $\mathcal{E}$.

PROPOSITION 3. If $L = \boxed{\bullet \circlearrowleft}$ is the object of $\mathcal{S}^{\Delta_1^{\mathrm{op}}}$ obtained by identifying the two points of the representable object $\Delta[1]$, then irreflexive graphs may be identified with L-partite graphs:

$$\mathcal{S}^{\cdot\rightrightarrows\cdot} \simeq \mathcal{S}(L).$$

References

[G1] Grothendieck, A., *SGA IV*, Springer Verlag LNM 269(1972), pp. 358–365.

[G2] Grothendieck, A., *Methods of Construction in Analytic Geometry*, Cartan Seminar 1960.

[L] Lawvere, F. W., Introduction to Springer Verlag LNM 274(1972), pp. 1–12.

[J] Johnstone, P., *Topological Topos*, Proc. London Math. Soc. (3) 38(1979), pp. 237–271.

[J'] Johnstone, P., *How general is a generalized space*, in Aspects of Topology, in memory of Hugh Dowker, London Math. Soc. Lecture Notes Series 93, 1985, 77–111.

University at Buffalo, Department of Mathematics
244 Mathematics Building, Buffalo, N.Y. 14260-2900, USA
Email: `wlawvere@buffalo.edu`

This article may be accessed from `http://www.tac.mta.ca/tac/reprints` or by anonymous ftp at
`ftp://ftp.tac.mta.ca/pub/tac/html/tac/reprints/articles/9/tr9.{dvi,ps}`

FUNCTORIAL SEMANTICS OF ALGEBRAIC THEORIES

AND

SOME ALGEBRAIC PROBLEMS IN THE CONTEXT OF FUNCTORIAL SEMANTICS OF ALGEBRAIC THEORIES

F. WILLIAM LAWVERE

Contents

Part A

Author's comments

The 40th anniversary of my doctoral thesis was a theme at the November 2003 Florence meeting on the "Ramifications of Category Theory". Earlier in 2003 the editors of *TAC* had determined that the thesis and accompanying problem list should be made available through *TAC* Reprints. This record delay in the publication of a thesis (and with it a burden of guilt) is finally coming to an end. The saga began when in January 1960, having made some initial discoveries (based on reading Kelley and Godement) such as adjoints to inclusions (which I called "inductive improvements") and fibered categories (which I called "galactic clusters" in an extension of Kelley's colorful terminology), I bade farewell to Professor Truesdell in Bloomington and traveled to New York. My dream, that direct axiomatization of the category of categories would help in overcoming alleged set-theoretic difficulties, was naturally met with skepticism by Professor Eilenberg when I arrived (and also by Professor Mac Lane when he visited Columbia). However, the continuing patience of those and other professors such as Dold and Mendelsohn, and instructors such as Bass, Freyd, and Gray allowed me to deepen my knowledge and love for algebra and logic. Professor Eilenberg even agreed to an informal leave which turned out to mean that I spent more of my graduate student years in Berkeley and Los Angeles than in New York. My stay in Berkeley tempered the naive presumption that an important preparation for work in the foundations of continuum mechanics would be to join the community whose stated goal was the foundations of mathematics. But apart from a few inappropriate notational habits, my main acquisition from the Berkeley sojourn was a more profound acquaintance with the problems and accomplishments of 20th century logic, thanks again to the remarkable patience and tolerance of professors such as Craig, Feferman, Scott, Tarski, and Vaught. Patience began to run out when in February 1963, wanting very much to get out of my Los Angeles job in a Vietnam war "think" tank to take up a teaching position at Reed College, I asked Professor Eilenberg for a letter of recommendation. His very brief reply was that the request from Reed would go into his waste basket unless my series of abstracts be terminated post haste and replaced by an actual thesis. This tough love had the desired effect within a few weeks, turning the tables, for it was now he who had the obligation of reading a 120-page paper of baroque notation and writing style. (Saunders Mac Lane, the outside reader, gave the initial approval and the defence took place in Hamilton Hall in May 1963.) The hasty preparation had made adequate proofreading difficult; indeed a couple of lines (dealing with the relation between expressible and definable constants) were omitted from the circulated version, causing consternation and disgust among universal algebraists who tried to read the work. Only in the new millennium did I discover in my mother's attic the original handwritten draft, so that now those lines can finally be restored. Hopefully other obscure points will be clarified by this actual publication, for which I express my gratitude to Mike Barr, Bob Rosebrugh, and all the other editors of TAC, as well as to Springer-Verlag who kindly consented to the republication of the 1968 article.

1. Seven ideas introduced in the 1963 thesis

(1) The category of categories is an accurate and useful framework for algebra, geometry, analysis, and logic, therefore its key features need to be made explicit.

(2) The construction of the category whose objects are maps from a value of one given functor to a value of another given functor makes possible an elementary treatment of adjointness free of smallness concerns and also helps to make explicit both the existence theorem for adjoints and the calculation of the specific class of adjoints known as Kan extensions.

(3) Algebras (and other structures, models, etc.) are actually functors to a background category from a category which abstractly concentrates the essence of a certain general concept of algebra, and indeed homomorphisms are nothing but natural transformations between such functors. Categories of algebras are very special, and explicit axiomatic characterizations of them can be found, thus providing a general guide to the special features of construction in algebra.

(4) The Kan extensions themselves are the key ingredient in the unification of a large class of universal constructions in algebra (as in [Chevalley, 1956]).

(5) The dialectical contrast between presentations of abstract concepts and the abstract concepts themselves, as also the contrast between word problems and groups, polynomial calculations and rings, etc. can be expressed as an explicit construction of a new adjoint functor out of any given adjoint functor. Since in practice many abstract concepts (and algebras) arise by means other than presentations, it is more accurate to apply the term "theory", not to the presentations as had become traditional in formalist logic, but rather to the more invariant abstract concepts themselves which serve a pivotal role, both in their connection with the syntax of presentations, as well as with the semantics of representations.

(6) The leap from particular phenomenon to general concept, as in the leap from cohomology functors on spaces to the concept of cohomology operations, can be analyzed as a procedure meaningful in a great variety of contexts and involving functorality and naturality, a procedure actually determined as the adjoint to semantics and called extraction of "structure" (in the general rather than the particular sense of the word).

(7) The tools implicit in (1)–(6) constitute a "universal algebra" which should not only be polished for its own sake but more importantly should be applied both to constructing more pedagogically effective unifications of ongoing developments of classical algebra, and to guiding of future mathematical research.

In 1968 the idea summarized in (7) was elaborated in a list of solved and unsolved problems, which is also being reproduced here.

2. Delays and Developments

The 1963 acceptance of my Columbia University doctoral dissertation included the condition that it not be published until certain revisions were made. I never learned what exactly those revisions were supposed to be. Four years later, at the 1967 AMS Summer meeting in Toronto, Sammy had thoroughly assimilated the concepts and results of *Functorial Semantics of Algebraic Theories* and had carried them much further; one of his four colloquium lectures at that meeting was devoted to new results in that area found in collaboration with [Eilenberg & Wright, 1967]. In that period of intense advance, not only Eilenberg and Wright, but also [Beck, 1967], [Bénabou, 1968], [Freyd, 1966], [Isbell, 1964], [Linton, 1965], and others, had made significant contributions. Thus by 1968 it seemed that any publication (beyond my announcements of results [Lawvere, 1963, 1965]) should not only correct my complicated proofs, but should also reflect the state of the art, as well as indicate more systematically the intended applications to classical algebra, algebraic topology, and analysis. A book adequate to that description still has not appeared, but *Categories and Functors* [Pareigis, 1970] included an elegant first exposition. Ernie Manes' book called *Algebraic Theories*, treats mainly the striking advances initiated by Jon Beck, concerning the Godement-Huber-Kleisli notion of standard construction (triple or monad) which at the hands of Beck, [Eilenberg & Moore, 1965], Linton, and Manes himself, had been shown to be intimately related to algebraic theories, at least when the background is the category of abstract sets. Manes' title reflects the belief, which was current for a few years, that the two doctrines are essentially identical; however, in the less abstract background categories of topology and analysis, both monads and algebraic theories have applications which are complementary, but not identical.

Already in spring 1967, at Chicago, I had identified some of the sought-for links between continuum mechanics and category theory. Developing those would require some concepts from algebraic theories in particular, but moreover, much work on topos theory would be needed. These preoccupations in physics and toposes made it clear, however, that the needed book on algebraic theories would have to be deferred; only a partial summary was presented as an introduction to the 1968 list of generic problems.

The complicated proofs in my thesis of the lemmas and main theorems have been much simplified and streamlined over the past forty years in text and reference books, the most recent [Pedicchio & Rovati, 2004]. This has been possible due to the discovery and employment since 1970 of certain decisive abstract general relations expressed in notions such as regular category, Barr exactness [Barr, 1971], and factorization systems based on the "orthogonality" of epis and monos. However, an excessive reliance on projectives has meant that some general results of this "universal" algebra have remained confined to the abstract-set background where very special features such as the axiom of choice can even trivialize key concepts that would need to be explicit for the full understanding of algebra in more cohesive backgrounds.

Specifically, there is the decisive abstract general relation expressed by the commutativity of reflexive coequalizers and finite products, or in other words, by the fact that the connected components of the product of finitely many reflexive graphs form the product of the corresponding component sets. Only in recent years has it become widely known that this property is essentially characteristic of universal algebra (distinguishing it from the more general finite-limit doctrine treated by [Gabriel & Ulmer, 1971] and also presumably by the legendary lost manuscript of Chevalley). But the relevance of reflexive coequalizers was already pointed out in 1968 by [Linton, 1969], exploited in topos theory by [Johnstone, 1977], attributed a philosophical (i.e. geometrical) role by me [1986], and finally made part of a characterization theorem by [Lair, 1996]. It is the failure of the property for infinite products that complicates the construction of coequalizers in categories of infinitary algebras (even those where free algebras exist). On the other hand, the property holds for algebra in a topos, even a topos which has no projectives and is not "coherent" (finitary). A corollary is that algebraic functors (those induced by morphisms of theories) not only have left adjoints (as proved in this thesis and improved later), but also themselves preserve reflexive coequalizers. The cause for the delay of the general recognition of such a fundamental relationship was not only the reliance on projectives; also playing a role was the fact that several of the categories traditionally considered in algebra have the Mal'cev property (every reflexive subalgebra of a squared algebra is already a congruence relation) and preservation of coequalizers of congruence relations may have seemed a more natural question.

3. Comments on the chapters of the 1963 Thesis

3.1. Chapter I. There are obvious motivations for making explicit the particular features of the category of categories and for considering the result as a guide or framework for developing mathematics. Apart from the contributions of homological algebra and sheaf theory to algebraic topology, algebraic geometry, and functional analysis, and even apart from the obvious remark that category theory is much closer to the common content of all these, than is, say, the iterated membership conception of the von Neumann hierarchical representation of Cantor's theory, there is the following motivation coming from logical considerations (in the general philosophical sense). Much of mathematics consists in calculating in various abstract theories, specifically interpreting one abstract theory into another, interpreting an abstract theory into a background to obtain a concrete category of structures, and transforming these structures in and among these categories. Now, for one thing, the use of the term "category" of structures of a certain kind had already become obvious in the 1950's and for another thing the idea of theories themselves as structures whose mutual interpretations would form a category was also evidently possible if one cared to carry it out, and indeed Hall, Halmos, Henkin, Tarski, and possibly others had already made significant moves in that direction. But what of the relation itself between abstract theory and concrete background? To conceptually relate any two things, it is necessary that they belong to a common category; that is, speaking more mathematically, it is first necessary to functorially transport them into a common third category (if indeed they were initially conceived as belonging to different categories); but then if the attempt to relate them is successful, the clearest exposition of the whole matter may involve presenting the two things themselves as citizens of that third category. In the case of "backgrounds" such as a universe of sets or spaces it was already clear that they form categories, wherein the basic structure is composition. In the case of logical theories of all sorts the most basic structure they support is an operation of substitution, which is most effectively viewed as a form of composition. Thus, if we construe theories as categories, models are functors! Miraculously, the Eilenberg-Mac Lane notion of natural transformation between functors specializes exactly to the morphisms of models which had previously been considered for various doctrines of theories. But only for the simplest theories are all functors models, because something more than substitution needs to be preserved; again, miraculously, the additional features of background categories which were often expressible in terms of composition alone via universal mapping properties, turned out to have precise analogs: the operations of disjunction, existential quantification, etc. on a theory are all uniquely determined by the behavior of substitution. Roughly, any collection K of universal properties of the category of sets specifies a doctrine: the theories in the doctrine are all the categories having the properties K; the mutual interpretations and models in the doctrine are just all functors preserving the properties K. The simplest non-trivial doctrine seems to be that of finite categorical products, and the natural setting for the study of it is clear. Thus Chapter I tries to make that setting explicit, with details being left to a later publication [Lawvere, 1966].

Since (any model of a theory of) the category of categories consists of arrows called functors, how are we to get inside the individual categories themselves (and indeed how can we correctly justify calling the arrows functors)? There has been for a long time the persistent myth that objects in a category are "opaque", that there are only "indirect" ways of "getting inside" them, that for example the objects of a category of sets are "sets without elements", and so on. The myth seems to be associated with an inherited belief that the only "direct" way to deal with whole/part relations is to write an unexplained epsilon or horseshoe symbol between A and B and to say that A is then "inside" B, even though in any model of such a discourse A and B are distinct elements on an equal footing. In fact, the theory of categories is the most advanced and refined instrument for getting inside objects, because it does provide explanations (existence of factorizations of inclusion maps) and also makes the sort of distinctions that Volterra and others had noted were necessary for the elements of a space (because the elements are morphisms whose domains are various figure-types that are also objects of the category). But there is also a restriction on the wholesale meaningfulness of membership and inclusion, namely that they are meaningless unless both figures A and B under consideration are morphisms with the same codomain. It was the lack of such restriction in the Frege conception that forced Peano to introduce the "singleton" operation and the attendant rigid distinction between membership and inclusion. (A kind of singleton operation does reappear in category theory, but with quite different conception and properties, namely, as a natural transformation from an identity functor to a covariant power-set functor).

The construction in Chapter I (and in [Lawvere, 1966]) of a formula with one more free variable A from any given formula of the basic theory of categories, was a refutation of the above myth. Based on nailed-down descriptions of special objects 1, 2, 3 which serve as domains for the arrows that are the objects, maps, and commutativities in any codomain, this figure-and-incidence analysis is typical of what is possible for many categories of interest. But much more is involved. The analysis of objects as structures of such a kind in a background category of more abstract "underlying" objects is revealed as a possible construction within a category itself (here the category of categories), which is thus "autonomous". The method is to single out certain trivial or "discrete" objects by the requirement that all figures (whose shape belongs to the designated kind) be constant, and establish an adjunction between the "background" category so defined and the whole ambiance. The assumed exponentiation operation then allows one to associate an interpretation of the whole ambient category into finite diagrams in the background category, with the arrows in the diagrams being induced by the incidence relations, such as, for example, the composition by functors between 3 and 2 in the case of the category of categories.

At any rate, since the whole figure and incidence scheme here is finite (count the order-preserving maps between the ordinals 1,2,3), the proof-theoretic aspect of the problem of axiomatizing a category of categories is at all levels equivalent to that of axiomatizing a category of discrete sets.

The exponentiation operation mentioned above, that is, the ubiquity of functor categories, characterized by adjointness, was perhaps new in the thesis. Kan had defined adjoints and proved their uniqueness, and of course he knew that function spaces, for example in simplicial sets, are right adjoint to binary product. But the method here, and indeed in the whole ensuing categorical logic, is to exploit the uniqueness by using adjointness itself as an axiom [Lawvere, 1969]. The left adjoint characterization (p.19) of the "natural" numbers is another instance of the same principle (I later learned that it is more accurate to attribute it to Dedekind rather than to Peano).

The construction denoted by (,) was here introduced for the purpose of a foundational clarification, namely to show that the notion of adjointness is of an elementary character, independent of complications such as the existence, for the codomain categories of given functors, of an actual category (as opposed to a mere metacategory) of sets into which both are enriched (as is often needlessly assumed); of course, enrichment is important when available, but note that notions like monoidal closed category, into which enrichments are considered, are themselves to be described in terms of the elementary adjointness.

The general calculus of adjoints and limits could be presented as based on the (,) operation. Note that it may also be helpful to people in other fields who sometimes find the concept of adjointness difficult to swallow. From the simpler idempotent cases of a full reflective subcategory and the dual notion of a full coreflective subcategory, the general case can always be assembled by composing. The third, larger category thus mediating a given adjointness can be universally chosen to be (two equivalent) (,) categories, although in particular examples another mediating category may exist.

The (,) operation then turned out to be fundamental in computing Kan extensions (i.e. adjoints of induced functors). Unfortunately, I did not suggest a name for the operation, so due to the need for reading it somehow or other, it rather distressingly came to be known by the subjective name "comma category", even when it came to be also denoted by a vertical arrow in place of the comma. Originally, it had been common to write (A, B) for the set of maps in a given category $\mathcal{C}$ from an object A to an object B; since objects are just functors from the category 1 to $\mathcal{C}$, the notation was extended to the case where A and B are arbitrary functors whose domain categories are not necessarily 1 and may also be different. Since it is well justified by naturality to name a category for its objects, the notation $\mathrm{Map}(A, B)$ might be appropriate and read "the category of maps from (a value of) A to (a value of) B"; of course, a morphism between two such maps is a pair of morphisms, one in the domain of A and one in the domain of B, which satisfy a commutative square in $\mathcal{C}$. The word "map" is actually sometimes used in this more structured way (i.e. not necessarily just a mere morphism), for example, drawing one-dimensional pictures on two-dimensional surfaces must take place in a category $\mathcal{C}$ where both kinds of ingredients can be interpreted.

In the Introduction to the thesis I informally remarked that "no theorem is lost" by replacing the metacategory of all sets by an actual category $\mathcal{S}$ of small sets. Of course, (even if $\mathcal{S}$ is not taken as the "smallest" inaccessible) many theorems will be gained; some of those theorems might be considered as undue restriction on generality. However, we can always arrange that $\mathcal{S}$ not contain any Ulam cardinals (independently of whether such exist in the meta-universe at large) and that yields many theorems of a definitely positive character for mathematics. Often those mathematically positive theorems involve "dualities" between $\mathcal{S}$-valued algebra and $\mathcal{S}$-valued topology (or bornology). For example, consider the contravariant adjoint functor C from topological spaces to real algebras; the "duality" problem is to describe in topologically meaningful ways which spaces are fixed under the adjunction. That all metric spaces X (or even all discrete spaces) should be so fixed (in the sense that $X \to \mathrm{spec}(C(X))$ is a homeomorphism) is equivalent to the fact that $\mathcal{S}$ contains no Ulam cardinals. The existence of Ulam cardinals is equivalent to the failure of such duality in the simple case which opposes discrete spaces (i.e. $\mathcal{S}$ itself) to the unary algebraic category of left M-sets where M is the monoid of endomaps of any fixed countable set [Isbell, 1964].

3.2. Chapter II. There are several possible doctrines of algebraic theories, even within the very particular conception of "general" algebra that is anchored in the notion of finite categorical product. The most basic doctrine simply admits that any small category with products can serve as a theory, for example any chosen such small subcategory of a given large geometric or algebraic category. This realization required a certain conceptual leap, because such theories do not come equipped with a syntactical presentation, although finding and using presentations for them can be a useful auxiliary to the study of representations (algebras). Here a presentation would involve a directed graph with specified classes of diagrams and cones destined to become commutative diagrams and product cones in the intended interpretations in other theories or interpretations as algebras; this is an instance of the flexible notion of sketch due to Ehresmann. The algebras according to this general doctrine have no preferred underlying sets (which does not prevent them from enjoying nearly all properties and subtleties needed or commonly considered in algebra) and hence no preferred notion of free algebra, although the opposite of the theory itself provides via Yoneda a small adequate subcategory of regular projectives in the category of algebras.

Intermediate between the sketches and the theories themselves is another kind of presentation (but this one lacks the fully syntactical flavor usually associated with a notion of presentation): actual small categories equipped with a class of cones destined to become product cones. This rests on the fact that the free category with finite products generated by any given small category exists. Due to that remark the inclusion of algebras into prealgebras (IV.1.1) is itself an example of an algebraic functor, so that its left adjoint is a special case of a general result concerning algebraic adjoints.

The general doctrine of products permits a kind of flexibility in exposition whereby not only can free algebras always be studied in terms of free theories (by adjoining constants as in V.1) and free theories can be reciprocally reduced to the consideration of initial algebras in suitable categories, but moreover all S-sorted theories (for fixed S) are themselves algebras for one fixed theory. Apart from such generalities, perhaps the most useful feature of this doctrine is that it involves the broadest notion of algebraic functor (since a morphism of theories is any product-preserving functor) and that all these still have left adjoints (while themselves preserving reflexive coequalizers). Examples of these are treated in this basically single-sorted thesis as "algebraic functors of higher degree" (IV.2).

For a doctrine of sorted theories and of algebras with underlying "sets", one needs the further structure consisting of a fixed theory S and a given interpretation of S into the theory being studied; when the theory changes, the change is required to preserve these given interpretations. Then any algebraic category according to this doctrine will be equipped automatically with an underlying (i.e. evaluation) functor to the category of S-algebras. The left adjoint to this underlying functor provides the notion of free algebras. Also the whole semantical process has an adjoint which to almost any functor from almost any category $\mathcal{X}$ to the category of S-algebras assigns its unique "structure", the best approximation to $\mathcal{X}$ in the abstract world of S-sorted theories; the adjunction $\mathcal{X} \to \mathrm{sem}(\mathrm{str}\mathcal{X})$ is a sort of closure: "the particular included in the induced concrete general".

Note that there is a slip in Proposition II.1.1 because the underlying functor from S-sorted theories to theories does not preserve coproducts; of course it does permit their construction in terms of pushouts ("comeets") so that all intended applications are correct (compare V.1).

With the general notion of S-sorted theory discussed above, the free algebras will not be adequate nor will the underlying functor be faithful without some normalization, usually taken to be that the functor A from S to the theory be bijective on objects.

In fact, the theory "is" the functor A and not merely its codomain category. This remains true even in the single-sorted case, on which I chose to concentrate in an attempt to make contact with the work of the universal algebra community. Here S is taken as the opposite of the category of finite sets, and $A(n)$ is the n-fold product of $A(1)$, with the maps $\pi_i = A(i)$ for $1 \to n$ as projections, often abbreviated to $A(n) = A^n$. The further "abuse of notation" (II.1.Def) which omits the functor A entirely has been followed in most subsequent expositions even though it can lead to confusion when (as in a famous example studied by Jonsson, Tarski and Freyd) the specific theory involves operations which make $A(2)$ isomorphic to $A(1)$. Actually, it seems appropriate in the sorted doctrine (as opposed to the unsorted one) to have a uniform source not only for the meaning of $A(i)$, but also of $A(s)$ where s is the involution of a 2-element set.

The discussion of presentations of theories with fixed sorting is entirely parallel to that for presentations of algebras for a fixed theory, because both are special cases of a general construction that applies to any given adjoint pair of functors. In the present case the underlying functor from single-sorted theories to sequences of sets, called "signatures", has a left adjoint yielding the free theory generated by any given signature (of course the underlying "signature" of that resulting theory is much larger than the given signature). Given such an adjoint pair, the associated category of presentations has as objects quadruples G, E, l, r in the lower category (codomain of the right adjoint) where l and r are morphisms from E to $T(G)$, where T is the monad resulting from composing the adjoints. Note that in the case at hand, E is also a signature, whose elements serve as names for laws rather than for basic operations as in G. Each such (name for a) basic equation has a left hand side and a right hand side, specified by l and r. The act of presentation itself is a functor (when the upper category has coequalizers) from the category of presentations: it consists of first applying the given adjointness to transform the pair l, r into a coequalizer datum in the upper category, and then forming that coequalizer. If the given adjoint pair is monadic, every object will have presentations, for example the "standard" one obtained by iterating the comonad. The full standard presentation is usually considered too unwieldy for practical (as opposed to theoretical) calculation, but it does have one feature sometimes used in smaller presentations: every equation name e is mapped by l to a generating symbol in G (only a small part of $T(G)$) and the equation merely defines that symbol as some polynomial $r(e)$ in the generating symbols; in other words, all information resides in the fact that those special generators may have more than one definition. That asymmetry contrasts with the traditional presentations [Duskin, 1969] in the spirit that Hilbert associated with syzygies; there one might as well assume that E is equipped with an involution interchanging l and r, because E typically arises as the kernel pair of a free covering by G of some algebra; the resolution can always be continued, further analyzing the presentation, by choosing a free covering (perhaps minimal) of the free algebra on E itself, then taking the kernel pair of the composites, etc. But another mild additional structure which could be assumed for presentations (and resolutions) is reflexivity, in the sense that there is a given map $G \to E$ which followed by l, r assures that there be for each generating symbol an explicit proof that it is equal to itself. That seeming banality may be important in view of the now-understood role of reflexivity, not only in combinatorial topology [Lawvere, 1986], but especially in key distinguishing properties of universal algebra itself [Lair, 1996, Pedicchio & Rovati, 2004].

3.3. Chapter III. As remarked already, the main "mathematical" theorem characterizing algebraic categories now has much more streamlined proofs [Pedicchio & Rovati, 2004], embedding it in a system of decisive concepts. It is to be hoped that future expositions will not only make more explicit the key role of reflexive coequalizers, but also relativize or eliminate the dominating role of projectives in order to fully address as "algebraic" the algebraic structures in sheaf toposes, as [Grothendieck, 1957] already did 45 years ago for the linear case.

But the other main result, that semantics has an adjoint called "structure" with a very general domain, appears as a kind of philosophical theorem in a soft mathematical guise. Indeed, that is one aspect, but note that the motivating example was cohomology operations, and that highly non-tautological examples continue to be discovered. Of course if a functor is representable, Yoneda's lemma reduces the calculation of its structure to an internal calculation. In general the calculation of the structure of a non-representable functor seems to be hopelessly difficult, however the brilliant construction by Eilenberg and Mac Lane of representing spaces for cohomology in the Hurewicz homotopy category (it is not representable, of course, in the original continuous category) not only paved the way for calculating cohomology operations but illustrated the importance of changing categories. Another more recent discovery by [Schanuel, 1982] provides an astonishing example of the concreteness of the algebraic structure of a non-representable functor, which however has not yet been exploited sufficiently in analysis and "noncommutative geometry". Namely the underlying-bornological-set functor on the category of finite dimensional noncommutative complex algebras has as its unary structure precisely the monoid of entire holomorphic functions. The higher part of the structure is a natural notion of analytic function of several noncommuting variables, and holomorphy on a given domain is also an example of this natural structure. Moreover, a parallel example involving special finite-dimensional noncommutative real algebras leads in the same way to C^∞ functions.

3.4. Chapter IV. The calculus of algebraic functors and their adjoints is at least as important in practice as the algebraic categories themselves. Thus it is unfortunate that there was no indication of the fundamental fact that these functors preserve reflexive coequalizers (of course their preservation of filtered colimits has always been implicit). As remarked above, "the algebra engendered by a prealgebra" is a special case of an adjoint to a (generalized) algebraic functor. However, contrary to what might be suggested by the treatment in this chapter, the use of that reflection is not a necessary supplement to the use of Kan extensions in proving the general existence of algebraic adjoints: as remarked only later by Michel André, Jean Bénabou, Hugo Volger, and others, the special exactness properties of the background category of sets imply that the left Kan extension, along any morphism of algebraic theories, of any algebra in sets, is already itself again product-preserving. Clearly, the same sort of thing holds, for example, with any topos as background. On the other hand, this chapter could be viewed as an outline of a proof that such adjoints to induced functors should at least exist even for algebras in backgrounds which are complete but poor in exactness properties,

4. Some developments related to the problem list in the 1968 Article

4.1. 1968 Section 4.

Problem 2. [Wraith, 1970] made considerable progress on the understanding of which algebraic functors have right adjoints by pointing out that they all involve extensions by new unary operators only. He also gave a general framework for understanding which kinds of identities could be imposed on such unary operations, while ensuring the existence of the right adjoint. These identities can be of a more general kind than the requirement that the new operators act by endomorphisms of the old ones (as treated in Chapter V of the thesis); an important example of such an identity is the Leibniz product rule, the right adjoint functor from commutative algebras to differential algebras being the formation of formal power series in one variable. Such twisted actions in the sense of Wraith are related to what I have called "Galilean monoids" as exemplified by second-order differential equations [http://www.buffalo.edu/~wlawvere]. Because they always preserve reflexive coequalizers, algebraic functors in our narrow sense will have right adjoints if only they preserve finite coproducts.

Problem 3. There are certainly non-linear examples of Frobenius extensions of theories; for example with the theory of a single idempotent, the act of splitting the idempotent serves as both left and right adjoint to the obvious interpretation into the initial theory. See [Kock, 2004] for some interesting applications of Frobenius extensions, which can be viewed dually as spaces carrying a distribution of global support.

Problem 4. There has been striking progress by [Zelmanov, 1997] on the restricted Burnside problem. He proved what had been conjectured for over 50 years, in effect that for the theory $B(n)$ of groups of exponent n, the underlying set functor from finite $B(n)$ algebras has a left adjoint, giving rise to a quotient theory $R(n)$ which is usually different from $B(n)$. This raises the question of presenting $R(n)$, i.e. what are the additional identities? As pointed out in the next problem, the structure of finite algebras is typically profinite if the (hom sets of) the theory are not already finite, so this Burnside-Zelmanov phenomenon seems exceptional; but a hope that unary tameness implies tameness for all arities m is realizable in some other contexts, such as modern extensions of real algebraic geometry [Van den Dries, 1998].

Problem 5. Embedding theorems of the type sought here have a long history in algebra and elsewhere. In earlier work it was not always distinguished that there are two separate issues, the existence of the relevant adjoint and the injectivity of the adjunction map to it. Here in this narrowly algebraic context, the answer to the first question is always affirmative, but the second question remains very much dependent on the particular case. It is remarkable that a kind of general solution has been found by [MacDonald & Sobral, 2004]. Naturally, considerable insight is required in order to apply their criterion, thus I still do not know the answer to the Wronskian question posed at the end of Section 2. Geometrically, that question is: given a Lie algebra L, whether a sufficiently complicated vector field on a sufficiently high-dimensional variety can be found so that L can be faithfully represented by commutators of those very special vector fields which are in the module (over the ring of functions) generated by that particular one. Schanuel has pointed out that the question depends on the ground field, because with complex scalars the rotation group in three dimensions can indeed be so represented.

Problem 6. Although the use of the bar notation suggests averaging, naturality of it was omitted from mention here. If naturality were included, possibly a suitable Maschke-like class of examples of this analog of Artinian semi-simplicity could be characterized in terms of a central idempotent operation.

4.2. 1968 Section 5. While problems involving the combination of the structure-semantics adjoint pair with the restriction to the finite may indeed be of interest in connection with the general problems of the Burnside-Kurosh type, few works in that sense have appeared, although for example the fact that one thus encounters theories enriched in the category of compact spaces has been remarked. But quite striking, as remarked before, is Schanuel's result that by combining the finite dimensionality with non-commutativity and bornology, the naturality construction, which here achieved merely the formal Taylor fields, captures in its amended form precisely the hoped-for ring of entire functions.

5. Concerning Notation and Terminology

In order to ease the burden of those 21st century readers who may try to read these documents in detail, let me point out some of the more frequent terminological and notational anachronisms. The order of writing compositions, which we learned from the stalwart Birkhoff-Mac Lane text, was seriously championed for most of the 1960's by Freyd, Beck and me (at least) because of the belief that it was more harmonious with the reading of diagrams. More recently it has also been urged by some computer scientists. However, the experience of teaching large numbers of students, most of whose courses are with mathematical scientists following the opposite convention, gradually persuaded us that there are other points within the great weight of mathematical tradition where reforms might be more efficiently advocated. (I personally take great comfort in Steve Schanuel's remark that there is a rational way to read the formulas that is compatible with the diagrams: gf is read "g following f".)

Some anachronisms are due to the fact the thesis was written before the partial standardization of the "co-" terminology at the 1965 La Jolla meeting: because we could not agree there on whether it is adjoints or co-adjoints that are the left adjoints, the "co-" prefix largely disappeared (from the word adjoint) and Kan's original left vs. right has been retained in that particular context.

Not exactly anachronistic, but not yet standard either, is my use of the term "congruence relation" for a graph which is the kernel pair of at least one map, in other words, for an "effective equivalence relation"; this definition raises the question of characterizing congruence relations in other terms, that is, in terms of maps in, rather than maps out, and of course the expected four clauses MRST suffice in the simple abstract backgrounds considered here, as well as in others; these are some of the relationships made partially explicit by Barr exactness.

The musical notation "flat" for the algebraic functor induced by a given morphism of theories occurs in print, as far as I know, only in the 1968 paper reprinted here. It was taken from Eilenberg's 1967 AMS Colloquium Lectures.

There was a traditional fuzziness in logic between presentations of theories and the sort of structures that I call theories. Unfortunately, in my attempt to sharpen that distinction in order to clarify the mutual transformation and rules of its aspects, I fell notationally afoul of one of the relics of the fuzziness. That is, while affirming that arities are sets, not ordinal numbers, and in particular, that the role of the "variables" in those sets is to parameterize projection maps, I inconsistently followed the logicians' tradition of writing such parameterizations as though the domains were ordinals. Such orderings are often indispensable in working with presentations of theories (or of algebras, for example using Gröbner bases) but they are spurious structure relative to the algebras or theories themselves. (Not only arities, but sets of operation symbols occurring in a presentation, have no intrinsic ordering: for example, the notion of a homomorphism between two rings is well defined because the same symbols, for plus and for times, are interpreted in both domain and codomain; this has nothing to do with any idea that plus or times comes "first".)

Finally, there was the choice, which I now view as anachronistic, of considering that an algebraic theory is a category with coproducts rather than with products. The "coproduct" convention, which involves defining algebras themselves as contravariant functors from the theory into the background, indeed did permit viewing the theory itself as a subcategory of the category of models. However, for logics more general than the equational one considered here, such a direct inclusion of a theory into its category of models cannot be expected. The "product" convention permits the concrete definition of models as covariant functors from the theory; thus the theory appears itself as a generic model. Moreover, the "product" convention seems to be more compatible with the way in which algebra of quantities and geometry of figures are opposed. In algebraic geometry, C^∞ geometry, group invariance geometry, etc. it became clear that there were many potential applications of universal algebra in contexts that do not fit the view that signatures are more basic than clones (i.e. that no examples of algebra exist beyond those whose algebraic theories are given by presentations). In these geometric (equally concrete) examples, the algebraic theories typically arise with products, not coproducts. For example, [see also IV. 1 Example 5] the algebraic theory whose operations are just all smooth maps between Euclidean spaces is not likely to ever find a useful presentation by signatures, yet its category of algebras contains many examples (the function algebras of arbitrary manifolds, formal power series, infinitesimals) that need to be related to each other by the appropriate homomorphisms.

6. Outlook

Let me close these remarks by expressing the wish that the results of universal algebra be more widely and effectively used by algebraists in such fields as operator theory and algebraic geometry (and vice-versa). For example, the Birkhoff theorem on subdirectly irreducible algebras is very helpful in understanding various Nullstellensätze in algebraic geometry (and vice-versa), and this relationship could be made more explicit in the study of Gorenstein algebras or Frobenius algebras. That theorem would also seem to be of use in connection with Lie algebra. Another striking example is the so-called "commutator" theory developed in recent decades, which would seem to have applications in specific categories of mathematical interest such as that of commutative algebra, where a direct geometrical description of the basic operation of forming the product of ideals, without invoking the auxiliary categories of modules, has proven to be elusive.

August 30, 2004

References

M. Barr (1971), *Exact Categories*, Springer LNM **236**, 1–120.

J.M. Beck (1967), *Triples, algebras and cohomology*, Columbia University, Ph.D. Thesis, first published in Reprints in Theory and Applications of Categories, **2**, 2003.

J. Bénabou (1968), Structures algébriques dans les catégories, Cahier de Top. et Géom. Diff., **10**, 1–126.

C. Chevalley (1956), *Fundamental concepts of algebra.* Academic Press, NY.

J. Duskin (1969), Variations on Beck's tripleability criterion, Springer LNM, **106**, 74–129.

S. Eilenberg and J.C. Moore (1965), Adjoint Functors and Triples, Ill. J. Math **9**, 381–965.

S. Eilenberg and J. Wright (1967), Automata in General Algebras. Notes distributed in conjunction with the Colloquium Lectures given at the 72nd AMS Summer Meeting.

P. Freyd (1966), Algebra-valued functors in general categories and tensor products in particular, Colloq. Math. **14**, 89–106.

P. Gabriel and F. Ulmer (1971), *Lokal präsentierbare Kategorien*, Lecture Notes in Mathematics 221, Springer-Verlag, Berlin.

A. Grothendieck (1957), Sur quelques points d'algbre homologique. Tohôku Math. J. **9**, 119–221.

J. Isbell (1964), Subobjects, adequacy, completeness and categories of algebras, Rozprawy Mat. Soc., **36**, 1–32.

P.T. Johnstone (1977), *Topos Theory*, Academic Press, NY.

J. Kock (2004), *Frobenius Algebras and 2D Topological Quantum Field Theories*, London Mathematical Society, Student Texts 59.

C. Lair (1996), Sur le genre d'esquissibilité des catégories modelables (accessibles) possédant les produits de deux. Diagrammes **35**, 25–52.

F.W. Lawvere (1963), Functorial semantics of algebraic theories, (summary of Ph.D. thesis, Columbia) Proc. National Acad. of Sciences **50**, 869–872.

F.W. Lawvere (1965), Algebraic theories, algebraic categories, and algebraic functors, Theory of Models, North-Holland, Amsterdam, 413–418.

F.W. Lawvere (1966), The category of categories as a foundation for mathematics, La Jolla Conference on Categorical Algebra, Springer-Verlag, 1–20.

F.W. Lawvere (1969), Adjointness in Foundations, Dialectica **23**, 281–296.

F.W. Lawvere (1982), Review of P.M. Cohn's Universal Algebra, 2nd Edition, American Scientist (May-June) 329.

F.W. Lawvere (1986), Categories of spaces may not be generalized spaces, as exemplified by directed graphs, Revista Colombiana de matemticas **XX**, 147–178.

F.E.J. Linton(1965), Some aspects of equational categories, Proceedings of the Conference on Categorical Algebra, La Jolla, 84–94.

F.E.J. Linton (1969), Coequalizers and categories of algebras, Springer LNM **80**, 75–90.

J. MacDonald and M. Sobral (2004), Aspects of Monads, Categorical Foundations, ed. M.C. Pedicchio and W. Tholen, 213–268.

B. Mitchell (1965), *Theory of Categories*, Academic Press.

B. Pareigis (1970), *Categories and Functors*, Academic Press.

M.C. Pedicchio and F. Rovati (2004), Algebraic Categories, Categorical Foundations. M.C. Pedicchio and W. Tholen, eds. Cambridge University Press, 269–310.

S. Schanuel (1982), Continuous extrapolation to triangular matrices characterizes smooth functions, JPAA **24**, 59–71.

L. Van den Dries (1998), *Tame Topology and O-minimal Structures*, London Math Society LNS 248, Cambridge University Press.

G.C. Wraith (1970), *Algebraic Theories*, Aarhus Lecture Note Series **22**.

E. Zelmanov (1997), On the restricted Burnside problem, Fields Medalists' Lectures, World Scientific Series **5**, 623–632.

Part B

Functorial Semantics of Algebraic Theories

Introduction

In this paper we attempt to achieve a further unification of general algebra by replacing 'equationally definable classes of algebras' with 'algebraic categories'. This explicit consideration of the categorical structure has several advantages. For example, from the category (or more precisely from an underlying-set functor) we can recover, not only the identities which hold between given operations in a class of algebras, but also the operations themselves (Theorem III.1.1).

We also give attention to certain functors between algebraic categories, called 'algebraic functors', which are induced by maps between algebraic theories, and show that all such functors have adjoints (Theorem IV.2.1). Thus free algebras, tensor algebras, monoid algebras, enveloping algebras, the extension of a distributive lattice to a Boolean ring, covariant extension of rings for modules, and many other algebraic constructions are viewed in a unified way as functors adjoint to algebraic functors.

There is a certain analogy with sheaf theory here. Namely, our 'prealgebras' of a given type form a category of unrestricted functors, whereas 'algebras' are prealgebras which commute with a specified class of inverse limits. The analogy with sheaf theory is further seen in the theorem of IV.1.1 which results.

Algebraic functors and algebraic categories are actually themselves values of a certain functor $\mathfrak{S}$ which we call semantics. Semantics itself has an adjoint $\hat{\mathfrak{S}}$ which we call algebraic structure. In addition to suggesting a possible principle of philosophy (namely a generalization of our Theorem III.1.2), these functors serve as a tool which enables us to give a characterization of algebraic categories (Theorem III.2.1). As a consequence we deduce that if $\mathcal{X}$ is an algebraic category and $\mathbb{C}$ a small category, then the category $\mathcal{X}^{\mathbb{C}}$ of functors $\mathbb{C} \longrightarrow \mathcal{X}$ and natural transformations is also algebraic if $|\mathbb{C}|$ (the set of objects in $\mathbb{C}$) is finite (Theorem III.2.2—the last condition is also in a sense necessary) if C is strongly connected or the theory of $\mathcal{X}$ has constants.

Basic to these results is Theorem I.2.5 and its corollaries, which give explicit formulas for the adjoint of an induced functor between functor categories in terms of a direct limit over a small category defined with the aid of an operation (,) which we also find useful in other contexts (this operation is defined in I.1).

Chapters II and V cannot be said to contain profound results. However, an acquaintance with Chapter II is necessary for the reading of III and IV, as part of our basic language, that of 'algebraic theories', is developed there. Essentially, algebraic theories are an invariant notion of which the usual formalism with operations and equations may be regarded as a 'presentation' (II.2.) Chapter V serves mainly to clarify the rest of the paper by showing how the usual concepts of polynomial algebra, monoid of operators, and module may be studied using the tools of II, III, IV.

Now a word on foundations. In I.1 we have outlined, as a vehicle for the introduction of notation, a proposed first-order theory of the category of categories, intended to serve as a non-membership-theoretic foundation for mathematics. However, since this work is still incomplete, we have not insisted on a formal exposition. One so inclined could of course view all mathematical assertions of Chapter I as axioms. The significance of 'small' and 'large', however, needs to be explained; in particular, why do we regard the category of all small algebras of a given type as an adequate version of the category of 'all' algebras of that type? In ordinary Zermelo-Fraenkel-Skolem set theory ZF_1, where the existence of one inaccessible ordinal $\theta_0 = \omega$ is assumed, it is known that the existence of a second inaccessible θ_1 is an independent assumption, and also that if this assumption is made (obtaining ZF_2), the set $R(\theta_1)$ of all sets of rank less then θ_1 is a model for ZF_1. Thus no theorem of ZF_1 about the category of 'all' algebras of a given type is lost by considering the category of small (rank less then θ_1) algebras of that type in ZF_2; but the latter has the advantage of being a legitimate object (as well as a notion, or meta-object), amenable to the usual operations of product, exponentiation, etc. Our introduction of $\mathcal{C}_1$ is just the membership-free categorical analogue to the assumption of θ_1. In order to have a reasonable codomain for our semantics functor $\mathfrak{S}$, we find it convenient to go one step further and introduce a category $\mathcal{C}_2$ of 'large' categories. However, if one wished to deal with only finitely many algebraic categories and functors at a time, $\mathcal{C}_2$ could be dispensed with.

Chapter I

The category of categories and adjoint functors

1. The category of categories

Our notion of category is that of [Eilenberg & Mac Lane, 1945]. We identify objects with their identity maps and we regard a diagram

$$A \xrightarrow{f} B$$

as a formula which asserts that A is the (identity map of the) domain of f and that B is the (identity map of the) codomain of f. Thus, for example, the following is a universally valid formula

$$A \xrightarrow{f} B \Rightarrow A \xrightarrow{A} A \ \wedge A \xrightarrow{f} B \ \wedge B \xrightarrow{B} B \ \wedge Af = f = fB.$$

Similarly, an isomorphism is defined as a map f for which there exist A, B, g such that

$$A \xrightarrow{f} B \ \wedge B \xrightarrow{g} A \ \wedge fg = A \ \wedge gf = B.$$

Note that we choose to write the order of compositions in the fashion consistent with left to right following of diagrams.

By the **category of categories** we understand the category whose maps are 'all' possible functors, and whose objects are 'all' possible (identity functors of) categories. Of course such universality needs to be tempered somewhat, and this can be done as follows. We specify a finite number of finitary operations which we always want to be able to perform on categories and functors. We also specify a finite number of special categories and functors which we want to include as objects and maps in the category of categories. Since all these notions turn out to have first-order characterizations (i.e. characterizations solely in terms of the domain, codomain, and composition predicates and the logical constants $=, \forall, \exists, \Rightarrow, \wedge, \vee, \neg$), it becomes possible to adjoin these characterizations as new axioms together with certain other axioms, such as the axiom of choice, to the usual first-order theory of categories (i.e. the one whose only axioms are associativity, etc.) to obtain the **first-order theory of the category of categories**. Apparently a great deal of mathematics (for example this paper) can be derived within the latter theory. We content ourselves here with an intuitively adequate description of the basic operations and special objects in the category of categories, leaving the full formal axioms to a later paper. We assert that all that we do can be interpreted in the theory ZF_3, and hence is consistent if ZF_3 is consistent. By ZF_3 we mean the theory obtained by adjoining to ordinary Zermelo-Fraenkel set theory (see e.g. [Suppes, 1960]) axioms which insure the existence of two inaccessible ordinals θ_1, θ_2 beyond the usual $\theta_0 = \omega$.

The first three special objects which we discuss are the categories $\mathtt{0}, \mathtt{1}, \mathtt{2}$. The empty category $\mathtt{0}$ is determined up to unique isomorphism by the property that for every category $\mathbb{A}$, there is a unique functor $\mathtt{0} \longrightarrow \mathbb{A}$. The singleton category $\mathtt{1}$ is defined dually; the objects in any category $\mathbb{A}$ are in one-to-one correspondence with functors $\mathtt{1} \longrightarrow \mathbb{A}$. A **constant** functor is any which factors through $\mathtt{1}$. The arrow category $\mathtt{2}$ is, intuitively, the category

$$\underset{0}{\bullet} \longrightarrow \underset{1}{\bullet}$$

with three maps, two of which are objects, denoted by 0,1 (not to be confused with categories $\mathtt{0}$, $\mathtt{1}$), the third map having 0 as domain and 1 as codomain. The category $\mathtt{2}$ is a **generator** for the category of categories, i.e.

$$\forall \mathbb{A}, \forall \mathbb{B}, \forall f, \forall g, [\mathbb{A} \xrightarrow{f} \mathbb{B} \wedge \mathbb{A} \xrightarrow{g} \mathbb{B} \wedge \forall u[\mathtt{2} \xrightarrow{u} \mathbb{A} \Rightarrow uf = ug] \Rightarrow f = g].$$

Furthermore $\mathtt{2}$ is a retract of every generator; i.e.

$$\forall \mathbb{G}[\mathbb{G} \text{ is a generator} \Rightarrow \exists f, \exists g[\mathtt{2} \xrightarrow{f} \mathbb{G} \wedge \mathbb{G} \xrightarrow{g} \mathtt{2} \wedge fg = \mathtt{2}]].$$

These two properties, together with the obvious fact that $\mathtt{2}$ has precisely three endofunctors, two of which are constant, characterize $\mathtt{2}$ up to unique isomorphism. There are exactly two functors $\mathtt{1} \longrightarrow \mathtt{2}$, which we denote by $0, 1$. For any category $\mathbb{A}$, we define $u \in \mathbb{A}$ to mean $\mathtt{2} \xrightarrow{u} \mathbb{A}$, and if the latter is true we say u is a member of $\mathbb{A}$ or a map in $\mathbb{A}$. Although this 'membership' has almost no formal properties in common with that of set theory, the intuitive meaning seems close enough to justify the notation.
We do have the following proposition for every functor $\mathbb{A} \xrightarrow{f} \mathbb{B}$:

$$\forall x[x \in \mathbb{A} \Rightarrow \exists! y[y \in \mathbb{B} \wedge y = xf]].$$

Thus 'evaluation' is a special case of composition.

The first five operations on categories and functors which we mention are product, sum, equalizer, coequalizer and exponentiation. For any two categories $\mathbb{A}$, $\mathbb{B}$ there is a category $\mathbb{A} \times \mathbb{B}$, called their **product**, together with functors $\mathbb{A} \times \mathbb{B} \xrightarrow{p} \mathbb{A}$, $\mathbb{A} \times \mathbb{B} \xrightarrow{q} \mathbb{B}$, called projections, such that

$$\forall \mathbb{C}, \forall f, \forall g\ \mathbb{C} \xrightarrow{f} \mathbb{A} \wedge \mathbb{C} \xrightarrow{g} \mathbb{B} \Rightarrow \exists! h[hp = f \wedge hq = g].$$

$$\begin{tikzcd} & \mathbb{C} \arrow[dl, "f"'] \arrow[d, dashed, "h"] \arrow[dr, "g"] & \\ \mathbb{A} & \mathbb{A}\times\mathbb{B} \arrow[l, "p"] \arrow[r, "q"'] & \mathbb{B} \end{tikzcd}$$

Taking $\mathbb{C} = \mathbf{2}$, it is clear that the maps in $\mathbb{A} \times \mathbb{B}$ are in one-to-one correspondence with ordered pairs $\langle x, y\rangle$ where $x \in \mathbb{A}$ and $y \in \mathbb{B}$. (A similar statement holds for objects, as is seen by taking $\mathbb{C} = \mathbf{1}$.) In fact, for any $\mathbb{C}, f, g, h$, as in the diagram, we write $h = \langle f, g\rangle$. The **sum** $\mathbb{A} + \mathbb{B}$ and the associated injections are described by the dual (coproduct) diagram. Every member $\mathbf{2} \xrightarrow{u} \mathbb{A}+\mathbb{B}$ factors through exactly one of the injections $\mathbb{A} \longrightarrow \mathbb{A}+\mathbb{B}$, $\mathbb{B} \longrightarrow \mathbb{A}+\mathbb{B}$; however, this does not follow from the definition above. Both products and sums are unique up to unique isomorphisms which commute with the 'structural maps' (projections and injections, respectively). A similar statement holds for equalizers, coequalizers, and functor categories, which we now define.

Given any categories $\mathbb{A}$, $\mathbb{B}$ and any functors f, g such that $\mathbb{A} \overset{f}{\underset{g}{\rightrightarrows}} \mathbb{B}$ (i.e. $\mathbb{A} \xrightarrow{f} \mathbb{B} \wedge \mathbb{A} \xrightarrow{g} \mathbb{B}$), there is a category $\mathbb{K}$ and a functor $\mathbb{K} \xrightarrow{k} \mathbb{A}$ such that

$$kf = kg \wedge \forall \mathbb{C}\ \forall h\ [\mathbb{C} \xrightarrow{h} \mathbb{A} \wedge hf = hg \Rightarrow \exists! j\ [\mathbb{C} \xrightarrow{j} \mathbb{K} \wedge h = jk]].$$

$$\begin{tikzcd} & \mathbb{C} \arrow[dl, dotted, "j"'] \arrow[d, "h"] & \\ \mathbb{K} \arrow[r, "k"'] & \mathbb{A} \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & \mathbb{B} \end{tikzcd}$$

$\mathbb{K}$ is called the **equalizer** of f, g, but since this is really an abuse of language, we usually write $k = fEg$. Taking $\mathbb{C} = \mathbf{2}$, it is clear that the members of $\mathbb{K}$ are in one-to-one correspondence with those members of $\mathbb{A}$ at which f, g are equal.

Coequalizers are defined dually, and we write $k^* = fE^*g$, where $\mathbb{B} \xrightarrow{k*} \mathbb{K}^*$ is the structural map of the coequalizer $\mathbb{K}^*$ of f, g. The objects in $\mathbb{K}^*$ are equivalence classes of objects of $\mathbb{B}$, B being equivalent to B' if there exists $a \in \mathbb{A}$ such that $af = B \wedge ag = B'$. The maps in $\mathbb{K}^*$ are equivalence classes of admissible finite nonempty strings of maps in $\mathbb{B}$; here a string $\langle u_0, \ldots, u_{n-1}\rangle$ is admissible iff for every $i < n-1$, the codomain of u_i is *equivalent* to the domain of u_{i+1}, in the above sense; two strings are equivalent if their being so follows by composition (i.e. concatenation), reflexivity, symmetry, and transitivity from the following two types of relations: two strings $\langle u_0\rangle$ and $\langle u_0'\rangle$ of length one are equivalent if $\exists x \in \mathbb{A}\ [xf = u_0 \wedge xg = u_0']$; a string $\langle u_0, u_1\rangle$ of length two is equivalent to a string $\langle v\rangle$ of length one if $u_0u_1 = v$ in $\mathbb{B}$. Thus, for example, the coequalizer of $\mathbf{1} \overset{0}{\underset{1}{\rightrightarrows}} \mathbf{2}$ is the additive monoid $\mathbb{N}$ of non-negative integers. (A **monoid** is a category $\mathbb{A}$

such that $\exists!\ \mathbb{1} \longrightarrow \mathbb{A}$.) This example shows that, in contrast to the situation for algebraic categories (see Chapter III), coequalizers in the category of categories need not be onto, although they are of course 'epimorphisms', i.e. maps which satisfy the left cancellation law.

Before defining exponential (functor) categories, we point out that the operation of forming the product is functorial. That is, given any functors $\mathbb{A} \xrightarrow{f} \mathbb{A}'$, $\mathbb{B} \xrightarrow{g} \mathbb{B}'$, there is a unique functor $\mathbb{A}\times\mathbb{B} \longrightarrow \mathbb{A}'\times\mathbb{B}'$ which commutes with the projections; we denote it by $f\times g$. If $\mathbb{A}' \xrightarrow{f_1} \mathbb{A}''$, $\mathbb{B}' \xrightarrow{g_1} \mathbb{B}''$ are further functors, then $(f\times g)(f_1\times g_1) = ff_1\times gg_1$ by uniqueness. Analogous propositions hold for sums.

Now given any two categories $\mathbb{A}$, $\mathbb{B}$, there exists a category $\mathbb{B}^{\mathbb{A}}$ called the **exponential** or **functor** category of $\mathbb{A}$, $\mathbb{B}$, together with a functor $\mathbb{A}\times\mathbb{B}^{\mathbb{A}} \xrightarrow{e} \mathbb{B}$ called the evaluation functor, such that

$$\forall\mathbb{C}\ \forall f\ [\mathbb{A}\times\mathbb{C} \xrightarrow{f} \mathbb{B} \Rightarrow \exists!g\ [\mathbb{C} \xrightarrow{g} \mathbb{B}^{\mathbb{A}} \wedge f = (\mathbb{A}\times g)e]].$$

We sometimes write $g = \{f\}$.

$$\begin{tikzcd}
\mathbb{A}\times\mathbb{C} \arrow[d, shift right, "\mathbb{A}"'] \arrow[d, shift left, dashed, "g"] \arrow[dr, "f"] & \\
\mathbb{A}\times\mathbb{B}^{\mathbb{A}} \arrow[r, "e"'] & \mathbb{B}
\end{tikzcd}$$

Taking $\mathbb{C} = \mathbb{1}$, and noting that $\mathbb{A}\times\mathbb{1} \cong \mathbb{A}$ (isomorphic) for all $\mathbb{A}$, it follows that the objects in the functor category $\mathbb{B}^{\mathbb{A}}$ are in one-to-one correspondence with the functors $\mathbb{A} \longrightarrow \mathbb{B}$. Since $\mathbb{A}\times\mathbb{2} \cong \mathbb{2}\times\mathbb{A}$, it follows also that the maps in $\mathbb{B}^{\mathbb{A}}$ are in one-to-one correspondence with the functors $\mathbb{A} \longrightarrow \mathbb{B}^{\mathbb{2}}$ with domain $\mathbb{A}$ and codomain $\mathbb{B}^{\mathbb{2}}$. The latter is not to be confused with $\mathbb{B}\times\mathbb{B}$, which is isomorphic to $\mathbb{B}^{|\mathbb{2}|}$ (defined below). In fact the members of $\mathbb{B}^{\mathbb{2}}$ are in one-to-one correspondence with functors $\mathbb{2}\times\mathbb{2} \longrightarrow \mathbb{B}$, which in turn can be identified with **commutative squares** in $\mathbb{B}$, for $\mathbb{2}\times\mathbb{2}$ is a single commutative square

$$\begin{tikzcd}
00\ \bullet \arrow[r] \arrow[d] & \bullet\ 01 \arrow[d] \\
10\ \bullet \arrow[r] & \bullet\ 11
\end{tikzcd}$$

as can be proved using methods described below. The maps in a functor category are usually called **natural transformations**. Exponentiation is functorial in the sense that given

$$\mathbb{A}' \xrightarrow{f} \mathbb{A}\ ,\ \mathbb{B} \xrightarrow{g} \mathbb{B}',$$

there is a unique functor $\mathbb{B}^{\mathbb{A}} \xrightarrow{g^f} \mathbb{B}'^{\mathbb{A}'}$ such that the diagram

$$\begin{tikzcd}
\mathbb{A}' \times \mathbb{B}^{\mathbb{A}} \arrow[rr, "f \times \mathbb{B}^{\mathbb{A}}"] \arrow[d, shift right, "\mathbb{A}'"'] \arrow[d, shift left, dashed, "g^f"] & & \mathbb{A} \times \mathbb{B}^{\mathbb{A}} \arrow[d, "e"] \\
& & \mathbb{B} \arrow[d, "g"] \\
\mathbb{A}' \times \mathbb{B}'^{\mathbb{A}'} \arrow[rr, "e'"'] & & \mathbb{B}'
\end{tikzcd}$$

is commutative. If $\mathbb{A}'' \xrightarrow{f_1} \mathbb{A}'$, $\mathbb{B}' \xrightarrow{g_1} \mathbb{B}''$ are further functors, then by uniqueness

$$(gg_1)^{ff_1} = g^f g_1^{f_1}.$$

Godement's 'cinq règles de calcul functoriel' [Godement, 1958] now follow immediately. Products are associative; in particular for any three categories $\mathbb{A}$, $\mathbb{B}$, $\mathbb{C}$, there is a unique isomorphism

$$\mathbb{A} \times (\mathbb{B}^{\mathbb{A}} \times \mathbb{C}^{\mathbb{B}}) \xrightarrow{\alpha} (\mathbb{A} \times \mathbb{B}^{\mathbb{A}}) \times \mathbb{C}^{\mathbb{B}}$$

which commutes with the projections. Using this we can construct a unique 'composition' functor γ such that

$$\begin{tikzcd}
\mathbb{A} \times (\mathbb{B}^{\mathbb{A}} \times \mathbb{C}^{\mathbb{B}}) \arrow[rr, "\alpha"] \arrow[d, shift right, "\mathbb{A}"'] \arrow[d, shift left, dashed, "\gamma"] & & (\mathbb{A} \times \mathbb{B}^{\mathbb{A}}) \times \mathbb{C}^{\mathbb{B}} \arrow[d, shift right, "e"'] \arrow[d, shift left, "\mathbb{C}^{\mathbb{B}}"] \\
& & \mathbb{B} \times \mathbb{C}^{\mathbb{B}} \arrow[d, "\overline{e}"] \\
\mathbb{A} \times \mathbb{C}^{\mathbb{A}} \arrow[rr, "\overline{\overline{e}}"'] & & \mathbb{C}
\end{tikzcd}$$

One has isomorphisms

$$\begin{aligned}
\mathbb{B}^{(\mathbb{A}+\mathbb{A}')} &\cong \mathbb{B}^{\mathbb{A}} \times \mathbb{B}^{\mathbb{A}'} \\
(\mathbb{B} \times \mathbb{B}')^{\mathbb{A}} &\cong \mathbb{B}^{\mathbb{A}} \times \mathbb{B}'^{\mathbb{A}} \\
\mathbb{B}^{\mathbb{A}\times\mathbb{A}'} &\cong (\mathbb{B}^{\mathbb{A}})^{\mathbb{A}'} \\
\mathbb{B}^{\mathbf{1}} &\cong \mathbb{B} \\
\mathbb{B}^{\mathbf{0}} &\cong \mathbf{1}.
\end{aligned}$$

Furthermore, the existence of exponentiation implies that

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

this being a special case of a general theorem concerning adjoint functors Theorem 2.3. The two maps $\mathbf{1} \overset{0}{\underset{1}{\rightrightarrows}} \mathbf{2}$ induce the domain and codomain functors $\mathbb{B}^{\mathbf{2}} \xrightarrow{D_0} \mathbb{B}$ and $\mathbb{B}^{\mathbf{2}} \xrightarrow{D_1} \mathbb{B}$ for each category $\mathbb{B}$.

We now define a **set** to be any category $\mathbb{A}$ such that the unique map $\mathbf{2} \longrightarrow \mathbf{1}$ induces an isomorphism

$$\mathbb{A} \cong \mathbb{A}^{\mathbf{1}} \cong \mathbb{A}^{\mathbf{2}}.$$

That is, a set is a category in which every map is an object (identity). Note that the word ‘set’ here carries no ‘size’ connotation; we could equally well use the word ‘class’. Two further operations on categories can now be defined. For any category $\mathbb{A}$, there exists a *set* $|\mathbb{A}|_0$ and a functor $|\mathbb{A}|_0 \xrightarrow{i} \mathbb{A}$, such that

$$\forall \mathbb{S}\ \forall f\ [\mathbb{S} \cong \mathbb{S}^{\mathbf{2}} \wedge \mathbb{S} \xrightarrow{f} \mathbb{A} \Rightarrow \exists! h\ [hi = f]].$$

$$\begin{tikzcd}
\mathbb{S} \arrow[d, dashed, "h"'] \arrow[dr, "f"] & \\
|\mathbb{A}|_0 \arrow[r, "i"'] & \mathbb{A}
\end{tikzcd}$$

We call $|\mathbb{A}|_0$ the **set of objects** of $\mathbb{A}$. Every functor $\mathbb{A} \xrightarrow{f} \mathbb{A}'$ induces a unique functor $|\mathbb{A}|_0 \xrightarrow{|f|_0} |\mathbb{A}|_0$ commuting with i, i', and $\mathbb{A}' \xrightarrow{f_1} \mathbb{A}''$ implies $|ff_1|_0 = |f|_0|f_1|_0$.

The members of $|\mathbb{A}|_0$ (i.e functors $\mathbf{2} \longrightarrow |\mathbb{A}|_0$) are in one-to-one correspondence with functors $\mathbf{1} \longrightarrow \mathbb{A}$, and we sometimes, by abuse of notation, use $A \in |\mathbb{A}|_0$ to mean that $\mathbf{1} \xrightarrow{A} \mathbb{A}$. Note that $|\mathbb{A}^{\mathbf{2}}|_0$ is the **set of maps** of $\mathbb{A}$.

Dually, we define $|\mathbb{A}|_1$, the **set of components** of $\mathbb{A}$ to be a *set* which has a map $\mathbb{A} \longrightarrow |\mathbb{A}|_1$ universal with respect to functors from $\mathbb{A}$ to sets. The members of $|\mathbb{A}|_1$ may be regarded as equivalence classes of objects of $\mathbb{A}$, two objects A, A' being equivalent if there exists a finite sequence of objects and maps

$$A \longrightarrow A_1 \longleftarrow A_2 \longrightarrow A_3 \longleftarrow \ \ldots \ \longrightarrow A'$$

in $\mathbb{A}$. Clearly this operation is also functorial. The composite map

$$|\mathbb{A}|_0 \longrightarrow \mathbb{A} \longrightarrow |\mathbb{A}|_1$$

is always onto. It is an isomorphism iff $\mathbb{A}$ is a sum (possibly infinite) of monoids. $\mathbb{A}$ is **connected** if $|\mathbb{A}|_1 \cong \mathbf{1}$.

The existence of the two operations $|\ |_0$ and $|\ |_1$ implies that products, sums, equalizers, and coequalizers of sets are again sets (again appealing to Theorem 2.3).

We often abbreviate $|\ |_0$ to $|\ |$ and denote by $|\mathbb{A}|$ the set of *objects* of $\mathbb{A}$.

Given two functors $\mathbb{A} \xrightarrow{f} \mathbb{B}$, $\mathbb{A}' \xrightarrow{f'} \mathbb{B}$ with common codomain, we define their **meet** $\mathbb{M}$ to be the equalizer of the pair $\mathbb{A} \times \mathbb{A}' \rightrightarrows \mathbb{B}$ consisting of pf and $p'f'$ (p, p' being the projections from the product). The meet satisfies the following:

$$g'f' = gf \wedge \forall \mathbb{C}\ \forall h\ \forall h'\ [\mathbb{C} \xrightarrow{h} \mathbb{A} \wedge \mathbb{C} \xrightarrow{h'} \mathbb{A}' \wedge h'f' = hf \Rightarrow \\ \exists! k[kg = h\ \wedge kg' = h']].$$

$$\begin{tikzcd}
\mathbb{C} \arrow[dr, dotted, "k"] \arrow[drr, "h'"] \arrow[ddr, "h"'] & & \\
& \mathbb{M} \arrow[r, "g'"] \arrow[d, "g"] & \mathbb{A}' \arrow[d, "f'"] \\
& \mathbb{A} \arrow[r, "f"'] & \mathbb{B}
\end{tikzcd}$$

For example, if f' is a monomorphism (i.e. cancellable on the right) then so is $\mathbb{M} \xrightarrow{g} \mathbb{A}$, which may be called the inverse image of f' under f. If f is also a monomorphism then so is $\mathbb{M} \longrightarrow \mathbb{B}$, which may be called the intersection of f, f'.

Comeets are defined dually. For example, the category 3 may be defined as a comeet as follows:

$$\begin{tikzcd}
1 \arrow[r, "0"] \arrow[d, "1"'] & 2 \arrow[d, "12"] \\
2 \arrow[r, "01"'] & 3
\end{tikzcd}$$

Although it does not follow from this definition above, it is a fact about the category of categories that there is exactly one non-constant map $2 \longrightarrow 3$ in addition to the two displayed above; we denote this third map by 02. Intuitively the category 3 may be pictured thus:

$$\begin{tikzcd} & 1 \arrow[dr, "12"] & \\ 0 \arrow[ur, "01"] \arrow[rr, "02"'] & & 2 \end{tikzcd}$$

Like $\mathbf{2}$, $\mathbf{3}$ is an ordinal; in particular it is a **preorder**, i.e. a category $\mathbb{A}$ such that any pair of maps $\mathbf{1} \underset{B}{\overset{A}{\rightrightarrows}} \mathbb{A}$ has *at most one* extension to a map u such that

$$\mathbf{2} \xrightarrow{u} \mathbb{A} \wedge A = 0u \wedge B = 1u.$$

$$\begin{tikzcd} \mathbf{1} \arrow[r, "0"] \arrow[dr, "A"'] & \mathbf{2} \arrow[d, "u"] \\ & \mathbb{A} \end{tikzcd} \qquad \begin{tikzcd} \mathbf{1} \arrow[r, "1"] \arrow[dr, "B"'] & \mathbf{2} \arrow[d, "u"] \\ & \mathbb{A} \end{tikzcd}$$

We have already pointed out that members of any category $\mathbb{A}$ 'are' simply functors $\mathbf{2} \longrightarrow \mathbb{A}$. Among such members, we now show how to define the basic predicates domain, codomain, and composition entirely in terms of functors in the category of categories. Suppose x, y, u, a, b are all functors $\mathbf{2} \longrightarrow \mathbb{A}$. Then

a is the $\mathbb{A}$-**domain** *of x* iff

$$\begin{tikzcd} \mathbf{2} \arrow[r, "a"] \arrow[d] & \mathbb{A} \\ \mathbf{1} \arrow[r, "0"'] & \mathbf{2} \arrow[u, "x"'] \end{tikzcd}$$

is commutative.

b is the $\mathbb{A}$-**codomain** *of x* iff

$$\begin{tikzcd}
\mathbf{2} \arrow[r, "b"] \arrow[d] & \mathbb{A} \\
\mathbf{1} \arrow[r, "1"'] & \mathbf{2} \arrow[u, "x"']
\end{tikzcd}$$

is commutative.

and u *is a* $\mathbb{A}$-**composition** *of* $\langle x, y\rangle$ iff

$$\begin{tikzcd}
 & \mathbf{2} \arrow[dr, "12"] \arrow[drr, "y"] & \mathbf{2} \arrow[d, "02"] \arrow[dr, "u"] & \\
\mathbf{1} \arrow[ur, "0"] \arrow[dr, "1"'] & & 3 \arrow[r, dashed] & \mathbb{A} \\
 & \mathbf{2} \arrow[ur, "01"] \arrow[urr, "x"'] & &
\end{tikzcd}$$

is commutative.

By relativising quantifiers to functors $\mathbf{2} \longrightarrow \mathbb{A}$, one extends the above scheme so that to every formula φ of the usual first-order theory of categories, there is another formula $\varphi_{\mathbb{A}}$ with one more free variable $\mathbb{A}$, which states in the language of the category of categories that φ holds in $\mathbb{A}$. In particular, one can discuss $a \xrightarrow{x} b$, commutative triangles and squares, etc. *in* $\mathbb{A}$. We have the obvious

Metatheorem. *If φ is a sentence provable in the usual first-order theory of categories, then the sentence $\forall\mathbb{A}[\varphi_{\mathbb{A}}]$ is provable in the first-order theory of the category of categories.*

By use of the technique just outlined one can, e.g., deduce the precise nature of domain, codomain, and composition in a functor category or product category.

We say that two functors $\mathbb{A} \underset{g}{\overset{f}{\rightrightarrows}} \mathbb{B}$ are **equivalent** iff the two objects $\mathbf{1} \xrightarrow{\{f\}} \mathbb{B}^{\mathbb{A}}$, $\mathbf{1} \xrightarrow{\{g\}} \mathbb{B}^{\mathbb{A}}$, which correspond to them in the functor category are isomorphic in $\mathbb{B}^{\mathbb{A}}$. Categories $\mathbb{A}$, $\mathbb{B}$ are **equivalent** if there exist functors $\mathbb{A} \underset{g}{\overset{f}{\rightleftarrows}} \mathbb{B}$ such that $\{fg\}$ is isomorphic in the category $\mathbb{A}^{\mathbb{A}}$ to the object $\{\mathbb{A}\}$ and $\{gf\}$ is isomorphic in $\mathbb{B}^{\mathbb{B}}$ to the object $\{\mathbb{B}\}$ ($\{\mathbb{A}\}$ and $\{\mathbb{B}\}$ being the objects in the functor categories corresponding to the respective identity functors). In the latter case, f and g are called **equivalences**.

There are still some special objects in the category of categories which we will need to describe. Chief among these are $\mathcal{C}_1$, the **category of small categories** and $\mathcal{C}_2$, the **category of large categories**. In terms of the set theory ZF_3, a model for $\mathcal{C}_i$ may be obtained by considering the category of all categories and functors which are of rank less than θ_i, the i-th inaccessible ordinal ($i < 3$).

In terms of the theory of the category of categories, $\mathcal{C}_1$ may be described as an object such that:

(1) $\mathcal{C}_1$ has all properties we have thus far attributed to the category of categories. (Here we use the $\varphi_{\mathbb{A}}$ technique.) This includes the existence of 0, 1, 2, E, E^*, $|\ |_0$, $|\ |_1$, and exponentiation.

(2) $\mathcal{C}_1$ is closed under products and sums over all index sets $\mathbb{S}$ which are equal in size to some object in $\mathcal{C}_1$. (The precise meaning of infinite products and sums will be explained in Section 2.) A way of expressing the equipotency condition will be explained below.

(3) For any category $\mathbb{C}$ having the properties (1) and (2), there is a functor $\mathcal{C}_1 \longrightarrow \mathbb{C}$ which preserves 0, 1, 2, E, E^*, Π, Σ, $|\ |_0$, $|\ |_1$, and exponentiation, and which is unique up to equivalence.

One can then describe $\mathcal{C}_2$ by the same three properties, except that (1) now includes the existence of an object in $\mathcal{C}_2$ having all the properties in $\mathcal{C}_2$ which $\mathcal{C}_1$ has in the master category (i.e. in the theory of the category of categories).

Proposition. *$\mathcal{C}_1$ and $\mathcal{C}_2$ are unique up to equivalence and there is a canonical $\mathcal{C}_1 \longrightarrow \mathcal{C}_2$.*

The category $\mathcal{C}_0$ of finite categories is somewhat more difficult to describe in the first-order theory of the category of categories, because, as already pointed out, it is not closed under E^*.

The existence of the category $\mathcal{S}_0$ of finite sets, the category $\mathcal{S}_1$ of small sets, and the category $\mathcal{S}_2$ of large sets now follows, as does the existence of categories $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$ of monoids. Note that these are all *objects* in our master category. There are also objects $\mathbf{1} \xrightarrow{\{\mathcal{S}_1\}} \mathcal{C}_2$, $\mathbf{1} \xrightarrow{\{\mathcal{M}_1\}} \mathcal{C}_2$ in $\mathcal{C}_2$ which correspond to $\mathcal{S}_1$, $\mathcal{M}_1$. Of course there are many categories which appear as objects in our master category but which are larger than any object in our category of 'large' categories $\mathcal{C}_2$, for example $\mathcal{C}_2$ itself, the Boolean algebra $\mathbf{2}^{|\mathcal{C}_2|}$, etc. Whenever we refer to 'the' category of sets, monoids, etc., we ordinarily mean $\mathcal{S}_1$,$\mathcal{M}_1$, in general the full category of *small* objects of the stated sort, choosing any (not larger than) large version.

There is also a category ω which is an ordinal number and which is such that $|\omega| = |\mathbb{N}^2|$. Of course ω, $\mathbb{N}$, $\mathcal{S}_0$ are all quite different as categories, but we can choose a version of $\mathcal{S}_0$ such that

$$|\mathcal{S}_0| = |\omega| = |\mathbb{N}^2| = N.$$

(Note that $|\ |$ is not preserved under equivalence.) The common*(set)* value N has the property that there exists an object $\mathbf{1} \xrightarrow{0} N$ and a functor $N \xrightarrow{s} N$ such that for every category $\mathbb{C}$, for every object $\mathbf{1} \xrightarrow{C} \mathbb{C}$, and every functor $\mathbb{C} \xrightarrow{t} \mathbb{C}$ there exists a unique functor $N \xrightarrow{f} \mathbb{C}$ such that

$$\begin{tikzcd}
\mathbf{1} \arrow[r, "0"] \arrow[d, "\cong"'] & N \arrow[r, "s"] \arrow[d, dashed, "f"] & N \arrow[d, dashed, "f"] \\
\mathbf{1} \arrow[r, "C"'] & \mathbb{C} \arrow[r, "t"'] & \mathbb{C}
\end{tikzcd}$$

is commutative. This 'Peano postulate' characterizes the triple $\langle N, 0, s\rangle$ up to a unique isomorphism which preserves 0, s. The elementary properties of recursion follow easily.

Finally we mention two very important operations in the category of categories whose existence can be derived from what we have said. One is dualization, which assigns to each category $\mathbb{A}$ the category $\mathbb{A}^*$ obtained by interchanging domains and codomains and reversing the order of composition. Each functor $\mathbb{A} \xrightarrow{f} \mathbb{B}$ induces $\mathbb{A}^* \xrightarrow{f^*} \mathbb{B}^*$, and $(fg)^* = f^*g^*$ for $\mathbb{B} \xrightarrow{g} \mathbb{C}$. Also $A^{**} \cong \mathbb{A}$. This shows that we must take the two maps $\mathbf{1} \overset{0}{\underset{1}{\rightrightarrows}} \mathbf{2}$ as primitives in formalizing the first-order theory of the category of categories, because dualization is an automorphism which preserves the categorical structure but interchanges 0 and 1, whereas we need to distinguish between these in a canonical fashion in order to define $\varphi_{\mathbb{A}}$, $\mathcal{C}_1$, etc.

The other operation is one which we denote by (,). It is defined for any pair of functors $\mathbb{A} \xrightarrow{f} \mathbb{B}$, $\mathbb{A}' \xrightarrow{f'} \mathbb{B}$ with a common codomain, and is determined up to unique isomorphism by the requirement that all three squares below be *meet* diagrams:

$$\begin{tikzcd}
& & (f,f') \arrow[dl] \arrow[dr] & & \\
& (f,\mathbb{B}) \arrow[dl] \arrow[dr] & & (\mathbb{B},f') \arrow[dl] \arrow[dr] & \\
\mathbb{A} \arrow[dr, "f"'] & & \mathbb{B}^{\mathbf{2}} \arrow[dl, "D_0"] \arrow[dr, "D_1"'] & & \mathbb{A}' \arrow[dl, "f'"] \\
& \mathbb{B} & & \mathbb{B} &
\end{tikzcd}$$

It follows that the maps in the category (f, f') are in one-to-one correspondence with quadruples $\langle u_0, x, x', u_1 \rangle$ where $u_0 \in \mathbb{B}$, $u_1 \in \mathbb{B}$, $x \in \mathbb{A}$, $x' \in \mathbb{A}'$ and such that

$$\begin{tikzcd} \bullet \arrow[r, "xf"] \arrow[d, "u_0"'] & \bullet \arrow[d, "u_1"] \\ \bullet \arrow[r, "x'f'"'] & \bullet \end{tikzcd}$$

is a commutative square in $\mathbb{B}$. The domain of the above map is $\langle u_0, xD_0, x'D_0, u_0 \rangle$ and the codomain is $\langle u_1, xD_1, x'D_1, u_1 \rangle$, and the composition of $\langle u_0, x, x', u_1 \rangle$, $\langle u_1, y, y', u_2 \rangle$ is $\langle u_0, xy, x'y', u_2 \rangle$. In particular, the objects of (f, f') are in one-to-one correspondence with triples $\langle a, u, a' \rangle$ where $a \in |\mathbb{A}|$, $a' \in |\mathbb{A}'|$, $u \in \mathbb{B}$, and $uD_0 = af$, $uD_1 = a'f'$.

For example, if $\mathbb{A} = \mathbb{1}$, then f is an object in $\mathbb{B}$, say $f = B$, the objects of (B, f') are in one-to-one correspondence with pairs $\langle u, A' \rangle$ where $B \xrightarrow{u} A'f'$, and the maps in (B, f') are commutative triangles in $\mathbb{B}$ of the form:

$$\begin{tikzcd} & A'_0 f' \arrow[dd, "x'f'"] \\ B \arrow[ur, "u_0"] \arrow[dr, "u_1"'] & \\ & A'_1 f' \end{tikzcd} \qquad x' \in \mathbb{A}'$$

In particular, $\mathbb{A}'$ can be a subcategory of $\mathbb{B}$.

A case of the (,) construction which will play an important role in Chapter III is $(\mathcal{C}_2, \{\mathcal{S}_1\})$. The objects of this category are functors $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ whose codomain is the category of small sets and functions and whose domain is arbitrary large (smaller than θ_2) category $\mathcal{X}$. Maps in $(\mathcal{C}_2, \{\mathcal{S}_1\})$ are commutative triangles

$$\begin{tikzcd} \mathcal{X} \arrow[rr, "T"] \arrow[dr, "U"'] & & \mathcal{X}' \arrow[dl, "U'"] \\ & \mathcal{S}_1 & \end{tikzcd}$$

If $\mathbb{A} = \mathbb{A}' = \mathtt{1}$, then f, f' are both objects in $\mathbb{B}$, and (f, f') always reduces to a *set*, the set of maps in $\mathbb{B}$ with domain f and codomain f'. The condition referred to earlier in the discussion of $\mathcal{C}_i$ can now be formalized. According to condition (1) in the description of $\mathcal{C}_i$, there is an object $\mathtt{1}_i : \mathtt{1} \longrightarrow \mathcal{C}_i$ having the properties in $\mathcal{C}_i$ that $\mathtt{1}$ has in the master category.

Definition. *A set* $\mathbb{S}$ *is* **equipollent with a set of** $\mathcal{C}_i$ *iff there is an object* $\mathtt{1} \xrightarrow{S} \mathcal{C}_i$ *in* $\mathcal{C}_i$ *such that* $(\mathtt{1}_i, S) \cong \mathbb{S}$.

As the final topic of this section, we discuss full, faithful, and dense functors. Let $\mathbb{A} \xrightarrow{f} \mathbb{B}$ be a functor. Note that for every pair of objects $\mathtt{1} \overset{a}{\underset{a'}{\rightrightarrows}} \mathbb{A}$ in $\mathbb{A}$, there is an induced map

$$(a, a') \longrightarrow (af, a'f').$$

Definition. f *is* **full** *iff the above induced map is an epimorphism of sets for every pair* $\langle a, a' \rangle$ *of objects.* f *is* **faithful** *iff the induced map is a monomorphism of sets for every pair of objects.* f *is* **dense** *iff for every object* $b \in |\mathbb{B}|$ *there is an object* $a \in |\mathbb{A}|$ *such that* $af \cong b$ *in* $\mathbb{B}$*. For example, the inclusions* $\mathcal{C}_1 \longrightarrow \mathcal{C}_2$, $\mathcal{M}_i \longrightarrow \mathcal{C}_i$, *and* $\mathcal{S}_i \longrightarrow \mathcal{C}_i$ *are full and faithful, but not dense.*

A proof of the following proposition will be found, for example, in Freyd's dissertation [Freyd, 1960].

Proposition. *A functor* $\mathbb{A} \xrightarrow{f} \mathbb{B}$ *is an equivalence iff it is full, faithful, and dense.*

2. Adjoint functors

As pointed out in Section 1, for any two objects $\mathtt{1} \overset{A}{\underset{A'}{\rightrightarrows}} \mathbb{A}$ in a category $\mathbb{A}$, the category (A, A') is a set; however it need not be a small set (or even a 'large' set in our sense) so that in general $(\ ,\ )$ does not define a functor $\mathbb{A}^* \times \mathbb{A} \longrightarrow \mathcal{S}_1$ (although the latter is of course true for many categories of interest). This fact prevents us from giving the definition of adjointness in a functor category $\mathcal{S}_1^{\mathbb{B}^* \times \mathbb{A}}$ as done by [Kan, 1958]. However we are able to give a definition free of this difficulty by making use of the broader domain of our $(\ ,\ )$ operation.

Note that if $\mathbb{A} \xrightarrow{f} \mathbb{B}$, $\mathbb{B} \xrightarrow{g} \mathbb{A}$ are any functors, then there is a functor

$$(\mathbb{B}, f) \xrightarrow{\overline{f}} \mathbb{B} \times \mathbb{A}$$

defined by the outer functors $(\mathbb{B}, f) \longrightarrow \mathbb{B}^{\mathbf{2}} \xrightarrow{D_0} \mathbb{B}$ and $(\mathbb{B}, f) \longrightarrow \mathbb{A}$ in the diagram

$$\begin{tikzcd}
 & & (\mathbb{B}, f) \arrow[dl] \arrow[dr] & \\
 & \mathbb{B}^{\mathbf{2}} \arrow[dl, "D_0"] \arrow[dr, "D_1"'] & & \mathbb{A} \arrow[dl, "f"] \\
\mathbb{B} & & \mathbb{B} &
\end{tikzcd}$$

Similarly, there is a functor $(g, \mathbb{A}) \xrightarrow{\overline{g}} \mathbb{B} \times \mathbb{A}$ induced by the outer functors $(g, \mathbb{A}) \longrightarrow \mathbb{B}$ and $(g, \mathbb{A}) \longrightarrow \mathbb{A}^{\mathbf{2}} \xrightarrow{D_1} \mathbb{A}$ in the diagram

$$\begin{tikzcd}
 & (g, \mathbb{A}) \arrow[dl] \arrow[dr] & & \\
\mathbb{B} \arrow[dr, "g"'] & & \mathbb{A}^{\mathbf{2}} \arrow[dl, "D_0"] \arrow[dr, "D_1"'] & \\
 & \mathbb{A} & & \mathbb{A}
\end{tikzcd}$$

Note that if $\langle x, b, a, x' \rangle$ is a typical map in $(\mathbb{B}, f)$ (i.e. $bx' = x(af)$) then $\langle x, b, a, x' \rangle \overline{f} = \langle b, a \rangle$, and analogously for $\overline{g}$.

Definition. *If* $\mathbb{A} \xrightarrow{f} \mathbb{B}$ *and* $\mathbb{B} \xrightarrow{g} \mathbb{A}$, *then we say* g **is adjoint to** f *(and* f *is* **co-adjoint** *to* g*) iff there exists an* isomorphism h *rendering the triangle of functors*

$$\begin{tikzcd}
(\mathbb{B}, f) \arrow[rr, dashed, "h"] \arrow[dr, "\overline{f}"'] & & (g, \mathbb{A}) \arrow[dl, "\overline{g}"] \\
 & \mathbb{B} \times \mathbb{A} &
\end{tikzcd}$$

commutative, where $\overline{f}$, $\overline{g}$ *are the functors defined above.*

Theorem 1. *For each functor* $\mathbb{A} \xrightarrow{f} \mathbb{B}$*, there exists a functor* $\mathbb{B} \xrightarrow{g} \mathbb{A}$ *such that* g *is adjoint to* f *iff for every object* $B \in |\mathbb{B}|$ *there exists an object* $A \in |\mathbb{A}|$ *and a map* $B \xrightarrow{\varphi} Af$ *in* $\mathbb{B}$ *such that for every object* $A' \in |\mathbb{A}|$ *and every map* $B \xrightarrow{x} A'f$ *in* $\mathbb{B}$*, there exists a unique map* $A \xrightarrow{y} A'$ in $\mathbb{A}$ *such that* $x = \varphi(yf)$ *in* $\mathbb{B}$*:*

$$\begin{tikzcd} B \arrow[r, "\varphi"] \arrow[dr, "x"'] & Af \arrow[d, dashed, "yf"] \arrow[dl, phantom, "\mathbb{B}"] \\ & A'f \end{tikzcd}$$

Proof. If g is adjoint to f, let B be an object in $\mathbb{B}$. Regarding the identity map $(B)g \longrightarrow Bg$ as an object in $(g, \mathbb{A})$ and applying the functor h^{-1} we obtain an object $B \xrightarrow{\varphi} Bgf$ in $(\mathbb{B}, f)$ such that $(\varphi)h = Bg$, since $\overline{f} = h\overline{g}$. If $B \xrightarrow{x} A'f$ is any object in $(\mathbb{B}, f)$ of the displayed form, then $(x)h$ is an object $Bg \longrightarrow A$. We wish to show that $A = Bg$, φ satisfy the condition in the statement of the theorem. For this we show that $y = (x)h$ satisfies the above commutative triangle and is uniquely determined by that condition. Consider the objects φ, $x \in |(\mathbb{B}, f)|$ and Bg, $y \in |(g, \mathbb{A})|$ in the diagrams below. We have

$$(\varphi)h = Bg$$

$$(x)h = y$$

$$\begin{tikzcd} B \arrow[r, dashed, "u"] \arrow[d, "\varphi"'] \arrow[dr, phantom, "\mathbb{B}"] & B \arrow[d, "x"] & & Bg \arrow[r, "Bg"] \arrow[d] \arrow[dr, phantom, "\mathbb{A}"] & Bg \arrow[d, "y"] \\ Bgf \arrow[r, dashed, "v"'] & A'f & {} \arrow[r, phantom, "\xrightarrow{h}\ Bg"] & Bg \arrow[r, "y"'] & A' \end{tikzcd}$$

Obviously $\langle Bg, y\rangle$ defines a map $Bg \longrightarrow y$ in $(g, \mathbb{A})$, i.e. the right hand square above is commutative. Let $\langle u, v\rangle$ be h^{-1} of $\langle Bg, y\rangle$. Then the left hand square is commutative (i.e. defines a map $\varphi \longrightarrow x$ in $(\mathbb{B}, f)$) and, again because of the assumed commutativity property of h, we have $u = B$ (identity map) and $v = yf$. Hence $\varphi(yf) = x$ in $\mathbb{B}$. To show uniqueness, suppose $Bg \xrightarrow{y'} A'$ is such that $\varphi(y'f) = x$. Then taking $u = B$, $v = y'f$, one gets a map $\varphi \longrightarrow x$ in $(\mathbb{B}, f)$ and applying h to it one gets $(Bg)y = (Bg)y'$, where as before $y = (x)h$ (again using the fact that h commutes with the two functors $\overline{f}$, $\overline{g}$ to $\mathbb{B} \times \mathbb{A}$); but Bg is an identity map, hence $y = y'$.

Conversely, if the condition of the theorem holds, then we *choose* for each $B \in |\mathbb{B}|$ a pair $\langle A, \varphi\rangle$ satisfying the condition. Then setting $Bg = A$, a functor $|\mathbb{B}| \longrightarrow |\mathbb{A}|$ is defined for objects, which by the condition has a well-defined extension to a functor

$\mathbb{B} \xrightarrow{g} \mathbb{A}$ defined for all maps in $\mathbb{B}$. We then define a functor $(\mathbb{B}, f) \xrightarrow{h} (g, \mathbb{A})$ as follows. Given any map $x \longrightarrow x'$ in $(\mathbb{B}, f)$ defined, say, by $\langle b, a\rangle$ as below,

$$\begin{tikzcd}
B \arrow[rrr, "b"] \arrow[dr, "\varphi"] \arrow[dd, "x"'] & & & B' \arrow[dl, "\varphi'"'] \arrow[dd, "x'"] \\
& Bgf \arrow[r, dashed, "bgf"] \arrow[dl, dotted, "yf"] & B'gf \arrow[dr, dotted, "y'f"'] & \\
Af \arrow[rrr, "af"'] & & & A'f
\end{tikzcd}$$

the resulting square

$$\begin{tikzcd}
Bg \arrow[r, "bg"] \arrow[d, "y"'] & B'g \arrow[d, "y'"] \\
A \arrow[r, "a"'] & A'
\end{tikzcd}$$

is a map in $(g, \mathbb{A})$, by uniqueness. Define $\langle x, b, a, x'\rangle h = \langle y, b, a, y'\rangle$. Then h is clearly a functor $(\mathbb{B}, f) \xrightarrow{h} (g, \mathbb{A})$ such that $\overline{f} = h\overline{g}$. It is also clear that h is one-to-one and onto, hence an isomorphism. ∎

The above theorem shows that our notion of adjointness coincides with that of [Kan, 1958].

Corollary. *For any functor f, there is up to equivalence at most one g such that g is adjoint to f. If g is adjoint to f and g' is adjoint to f' where*

$$\mathbb{A} \xrightarrow{f} \mathbb{B} \xrightarrow{f'} \mathbb{C}$$

then $g'g$ is adjoint to ff'. Further if g is adjoint to f and t is co-adjoint to g, then t is equivalent to f.

The above theorem and corollary have obvious dualizations for co-adjoints.

For any categories $\mathbb{A}$, $\mathbb{D}$ the unique functor $\mathbb{D} \longrightarrow \mathbb{1}$ induces a functor $\mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}}$.

Definition. *$\mathbb{A}$ is said to have* **inverse limits** *over $\mathbb{D}$ if the functor $\mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}}$ has a* co-adjoint. *Dually $\mathbb{A}$ has* **direct limits** *over $\mathbb{D}$ if $\mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}}$ has an* adjoint. *We denote these functors co-adjoint and adjoint to $\mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}}$ by $\lim\limits_{\leftarrow \mathbb{D}}$ and $\lim\limits_{\rightarrow \mathbb{D}}$, respectively, when they exist, or by $\lim\limits_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}$, $\lim\limits_{\rightarrow \mathbb{D}}{}^{\mathbb{A}}$ if there is any danger of confusion. (We will violate our customary convention for the order of composition when evaluating limit functors.)*

If $\mathbb{D} \xrightarrow{f} \mathbb{A}$, then we also sometimes write

$$\overleftarrow{f} = \lim_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}(\{f\})$$

$$\overrightarrow{f} = \lim_{\rightarrow \mathbb{D}}{}^{\mathbb{A}}(\{f\})$$

if these exist, where $\{f\}$ is the object in $\mathbb{A}^{\mathbb{D}}$ corresponding to f. Note that the latter notation is unambiguous since f determines its domain $\mathbb{D}$ and codomain $\mathbb{A}$. (It is 'ambiguous' in the sense that the limit functors are defined only up to a unique equivalence of functors.) $\overleftarrow{f}$ and $\overrightarrow{f}$ are *objects* in $\mathbb{A}$ if they exist.

In particular, if $\mathbb{D}$ is a *set* we write

$$\prod_{\mathbb{D}} = \lim_{\leftarrow \mathbb{D}}$$

$$\mathop{\star}_{\mathbb{D}} = \lim_{\rightarrow \mathbb{D}}$$

in any $\mathbb{A}$ for which the latter exist, and we call these operations *product* and *coproduct*, respectively. In the categories $\mathcal{S}_i$, $\mathcal{C}_i$, or in $\mathcal{A}_i$ (the categories of finite, small, and large abelian groups), where the practice is customary, we replace $\star$ by $\sum$, and in the category $\mathcal{R}_c$ of (small) commutative rings with unit, we replace $\star$ by $\otimes$. In particular if $\mathbb{D} = |\mathbf{2}|$ and $\mathbb{D} \xrightarrow{A} \mathbb{A}$, we write

$$A_0 \times A_1 = \lim_{\leftarrow \mathbb{D}}(A)$$

$$A_0 \star A_1 = \lim_{\rightarrow \mathbb{D}}(A)$$

when these exist.

Also, when $\mathbb{D}$ is a set, say $S = |\mathbb{D}| \cong \mathbb{D}$, and whenever $\mathbf{1} \xrightarrow{A} \mathbb{A}$, we write

$$A^S = \lim_{\leftarrow \mathbb{D}}(A\mathbb{A}^{\mathbb{D} \to \mathbf{1}})$$

$$S \cdot A = \lim_{\rightarrow \mathbb{D}}(A\mathbb{A}^{\mathbb{D} \to \mathbf{1}})$$

for the 'S-fold product and S-fold coproduct of A with itself', respectively, when these exist. That is, A^S is the composite

$$\mathbf{1} \xrightarrow{A} \mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}} \xrightarrow{\prod_{\mathbb{D}}} \mathbb{A}$$

and $S \cdot A$ is the composite

$$\mathbf{1} \xrightarrow{A} \mathbb{A} \longrightarrow \mathbb{A}^{\mathbb{D}} \xrightarrow{\overset{\star}{\mathbb{D}}} \mathbb{A}$$

where $S = |\mathbb{D}| \cong \mathbb{D}$. Note that in the categories $\mathcal{C}_i$ (and $\mathcal{S}_i$) the notation A^S agrees essentially with the exponential notation already adopted, and that similarly $S \cdot A = S' \times A$ in $\mathcal{C}_i$, where S' in an object in $\mathcal{C}_i$, which is a *'set' in* $\mathcal{C}_i$, such that $(\mathbf{1}_i, S') \cong S \cong \mathbb{D}$, when such exist (i.e. when S is equipollent to a set of $\mathcal{C}_i$.)

The category $\mathbb{E}$ is defined by the requirement that

$$\begin{tikzcd}
|\mathbf{2}| \arrow[r, "i"] \arrow[d, "i"'] & \mathbf{2} \arrow[d] \\
\mathbf{2} \arrow[r] & \mathbb{E}
\end{tikzcd}$$

be a comeet diagram. $\mathbb{E}$ may be pictured thus:

$$0\bullet \rightrightarrows \bullet 1$$

The functors $\mathbb{E} \xrightarrow{a} \mathbb{A}$ are in one-to-one correspondence with pairs $\langle a', a'' \rangle$ of maps in $\mathbb{A}$ such that $a'D_0 = a''D_0 \wedge a'D_1 = a''D_1$ in $\mathbb{A}$. If $\underset{\leftarrow \mathbb{E}}{\lim}^{\mathbb{A}}$ and $\underset{\rightarrow \mathbb{E}}{\lim}^{\mathbb{A}}$ exist, then we denote by

$$\underset{\leftarrow \mathbb{E}}{\lim}(a) \xrightarrow{a'Ea''} A$$

and

$$A' \xrightarrow[\underset{\rightarrow \mathbb{E}}{\lim}]{a'E^*a''} (a)$$

the canonical maps associated with these limits (analogous to φ in Theorem 2.1), where $A = a'D_0 = a''D_0$ and $A'' = a'D_1 = a''D_1$ in $\mathbb{A}$. We call these limits the *equalizer* and *coequalizer*, respectively, of a in $\mathbb{A}$. Note that (by Theorem 2.1), we have

$$(a'Ea'')a' = (a'Ea'')a''$$

$$a'(a'E^*a'') = a''(a'E^*a'')$$

in $\mathbb{A}$.

If $\underset{\leftarrow \mathbf{o}}{\lim}^{\mathbb{A}}$ and $\underset{\rightarrow \mathbf{o}}{\lim}^{\mathbb{A}}$ exist, then, since $\mathbb{A}^{\mathbf{o}} \cong \mathbf{1}$, these functors may be regarded as *objects* in $\mathbb{A}$, which we denote by $1_{\mathbb{A}}$ and $0_{\mathbb{A}}$, respectively; these are unique up to unique isomorphism in $\mathbb{A}$. By Theorem 2.1, $1_{\mathbb{A}}$ is characterized by the property that for every $A \in |\mathbb{A}|$ there is a unique $A \longrightarrow 1_{\mathbb{A}}$ in $\mathbb{A}$; dually for $0_{\mathbb{A}}$.

Definition. *$\mathbb{A}$ is said to* **have finite limits** *iff* $\lim_{\leftarrow\mathbb{D}}{}^{\mathbb{A}}$ *and* $\lim_{\rightarrow\mathbb{D}}{}^{\mathbb{A}}$ *exist for every finite category $\mathbb{D}$ (i.e. for every $\mathbb{D}$ such that $|\mathbb{D}^2|$ is equipollent to a set of $\mathcal{C}_0$ (or $\mathcal{S}_0$)). $\mathbb{A}$ is said to be* **left complete** *iff* $\lim_{\leftarrow\mathbb{D}}{}^{\mathbb{A}}$ *exists for every* small *category $\mathbb{D}$ (i.e. for every $\mathbb{D}$ such that $|\mathbb{D}^2|$ is equipollent to a set of $\mathcal{C}_1$ (or $\mathcal{S}_1$)). Dually, $\mathbb{A}$ is* **right complete** *if it has small* direct *limits. $\mathbb{A}$ is* **complete** *iff it is left complete and right complete.*

The following fact was pointed out to the author by Peter Freyd.

Theorem 2. *A category $\mathbb{A}$ is left complete iff $\mathbb{A}$ has equalizers and arbitrary small products.*

Proof. Let $\mathbb{D}$ be any small category and $\mathbb{D} \xrightarrow{f} \mathbb{A}$ any functor. We construct $\overleftarrow{f}$ as follows. Let $|\mathbb{D}| \xrightarrow{\bar{f}} \mathbb{A}$ be the functor determined by the diagram

$$\begin{tikzcd} |\mathbb{D}| \arrow[r] \arrow[d, "|f|"'] \arrow[dr, dotted, "\bar{f}"] & \mathbb{D} \arrow[d, "f"] \\ |\mathbb{A}| \arrow[r] & \mathbb{A} \end{tikzcd}$$

and consider also the functor $|\mathbb{D}^2| \xrightarrow{|D_1|} |\mathbb{D}|$ where D_1 is the codomain functor of $\mathbb{D}$. Then, since $|\mathbb{D}|$, $|\mathbb{D}^2|$ are both small sets, the existence of the two objects in the diagram below is assured.

$$\begin{tikzcd} \prod\limits_{|\mathbb{D}|}\bar{f} \arrow[r, shift left, "a'"] \arrow[r, shift right, "a''"'] & \prod\limits_{|\mathbb{D}^2|}|D_1|\bar{f} \end{tikzcd}$$

We must now define the maps a', a'' in $\mathbb{A}$. For each $\mathbb{D} \in |\mathbb{D}|$, let $\prod\limits_{|\mathbb{D}|}\bar{f} \xrightarrow{p_D} Df$ denote the canonical projection. (The family of p_D's determines the map of Theorem 2.1 in this case.) For each $x \in |\mathbb{D}^2|$ consider

$$\prod_{|\mathbb{D}|}\bar{f} \xrightarrow{p_{(xD_1)}} (xD_1)f.$$

This family of maps determines a unique map $\prod\limits_{|\mathbb{D}|}\bar{f} \xrightarrow{a''} \prod\limits_{|\mathbb{D}^2|}|D_1|\bar{f}$ such that for every $x \in |\mathbb{D}^2|$, $a''q_x = p_{(xD_1)}$ where $\prod\limits_{|\mathbb{D}^2|}|D_1|\bar{f} \xrightarrow{q_x} (xD_1)f$ is the canonical projection of the second product. Now consider also, for $x \in |\mathbb{D}^2|$, the composite map

$$\prod_{|\mathbb{D}|}\bar{f} \xrightarrow{p_{(xD_0)}} (xD_0)f \xrightarrow{xf} (xD_1)f.$$

This family determines a unique $\prod_{|\mathbb{D}|}\overline{f} \xrightarrow{a'} \prod_{|\mathbb{D}^2|}|D_1|\overline{f}$ such that $a'q_x = p_{(xD_0)}(xf)$ for every $x \in |\mathbb{D}^2|$. We now let A denote the equalizer of the functor $\mathbb{E} \xrightarrow{a} \mathbb{A}$ determined by $\langle a', a''\rangle$:

$$A \xrightarrow{a'Ea''} \prod_{|\mathbb{D}|}\overline{f} \overset{a'}{\underset{a''}{\rightrightarrows}} \prod_{|\mathbb{D}^2|}|D_1|\overline{f}$$

We wish to show $A \cong \overleftarrow{f}$. For each $D \in |\mathbb{D}|$, let $A \xrightarrow{\varphi_D} Df$ be the composite $\varphi_D = (a'Ea'')p_D$. We need to show that the family φ_D for $D \in |\mathbb{D}|$ defines a map $(A)\mathbb{A}^{(\mathbb{D}\to 1)} \xrightarrow{\varphi} \{f\}$ in $\mathbb{A}^{\mathbb{D}}$, and that $\langle A, \varphi\rangle$ is universal with respect to that property. Because the maps in $\mathbb{A}^{\mathbb{D}}$ are natural transformations, the first is true since for every $D \xrightarrow{x} D'$ in $\mathbb{D}$, we have

$$\begin{aligned} \varphi_D(xf) &= (a'Ea'')p_D xf &&= (a'Ea'')a'q_x \\ &= (a'Ea'')a''q_x &&= (a'Ea'')p_{D'} \\ &= \varphi_{D'}. \end{aligned}$$

To show the universality we consider any other family ψ_D, $D \in |\mathbb{D}|$ such that for every $D \xrightarrow{x} D'$ in $\mathbb{D}$, $\psi_D(xf) = \psi_{D'}$:

$$\begin{tikzcd} & Df \arrow[dd, "xf"] \\ X \arrow[ur, "\psi_D"] \arrow[dr, "\psi_{D'}"'] & \\ & D'f \end{tikzcd}$$

By the universal property of products (i.e. by Theorem 2.1 applied to $\prod_{|\mathbb{D}|}$) the family ψ determines a unique map $X \xrightarrow{b} \prod_{|\mathbb{D}|}\overline{f}$ such that $bp_D = \psi_D$ for all $D \in |\mathbb{D}|$. Then for every $x \in \mathbb{D}$, we have (in $\mathbb{A}$)

$$\begin{aligned} ba''q_x &= bp_{(xD_1)} &&= \psi_{(xD_1)} = \psi_{(xD_0)}(xf) \\ &= bp_{(xD_0)}(x\overline{f}) &&= ba'q_x. \end{aligned}$$

By uniqueness, $ba'' = ba'$; i.e. b 'equalizes' $\langle a', a''\rangle$. Therefore $\exists! X \xrightarrow{y} A$ such that $y(a'Ea'') = b$. But by construction y is also the unique map satisfying $\psi_D = y\varphi_D$ for all $D \in |\mathbb{D}|$. This proves that $A \cong \overleftarrow{f}$. ■

Because Theorem 2.2 obviously has a dual, we have the

Corollary. *A category $\mathbb{A}$ is complete iff $\mathbb{A}$ has equalizers, coequalizers, small products, and small coproducts.*

Also we can establish a second

Corollary. *A category has finite limits iff $\mathbb{A}$ has inverse and direct limits over the three categories $\mathbf{0}$, $|\mathbf{2}|$, $\mathbb{E}$.*

Proof. By the proof of Theorem 2.2, we need only show that $\mathbb{A}$ has products and coproducts over finite sets. We have assumed that $\lim\limits_{\leftarrow \mathbf{0}}{}^{\mathbb{A}}$ and $\lim\limits_{\rightarrow \mathbf{0}}{}^{\mathbb{A}}$ exist, and for any $\mathbb{A}$ we have that

$$\lim_{\rightarrow \mathbf{1}}{}^{\mathbb{A}} \cong \lim_{\leftarrow \mathbf{1}}{}^{\mathbb{A}} \cong \{\mathbb{A}\} \text{ in } \mathbb{A}^{\mathbb{A}^{\mathbf{1}}}.$$

We have also assumed that limits exist over $|\mathbf{2}| \cong \mathbf{1} + \mathbf{1}$. Then if n is any finite set for which we know $\prod\limits_n$ exists, and if $n + \mathbf{1} \xrightarrow{A} \mathbb{A}$ is any functor, consider the injections $n \xrightarrow{j} n + \mathbf{1}$, $\mathbf{1} \xrightarrow{(n)} n + \mathbf{1}$ and the binary product $(\prod\limits_n jA) \times A_n$. Defining projections $\overline{p}_0, \dots, \overline{p}_{n-1}, \overline{p}_n$ by the compositions

$$\begin{tikzcd}
 & & & (\prod\limits_n jA) \times A_n \arrow[dl, "q"'] \arrow[ddr, "q' = \overline{p}_n"] & \\
 & & \prod\limits_n jA \arrow[dll, "p_0"'] \arrow[dl] \arrow[dr, "p_{n-1}"] & & \\
A_0 & A_1 & \cdots & A_{n-1} & A_n
\end{tikzcd}$$

it is clear that the correct universal mapping property holds, so that $(\prod\limits_n jA) \times A_n \cong \prod\limits_{n+1} A$, completing the proof by induction. ■

In particular, one can define meets and comeets in $\mathbb{A}$, for any category $\mathbb{A}$ satisfying the condition of the corollary, as inverse and direct limits over the categories defined respectively by the comeet diagrams

$$\begin{tikzcd}
\mathbf{1} \arrow[r, "1"] \arrow[d, "1"'] & \mathbf{2} \arrow[d, "0'1"] \\
\mathbf{2} \arrow[r, "01"'] & \boxed{\begin{matrix} & 0' \\ & \downarrow \\ 0 \to & 1 \end{matrix}}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathbf{1} \arrow[r, "0"] \arrow[d, "0"'] & \mathbf{2} \arrow[d, "01'"] \\
\mathbf{2} \arrow[r, "01"'] & \boxed{\begin{matrix} 0 \to & 1' \\ \downarrow & \\ 1 & \end{matrix}}
\end{tikzcd}$$

Note that $\lim\limits_{\leftarrow \mathbf{2}}{}^{\mathbb{A}} \cong D_0$ and $\lim\limits_{\rightarrow \mathbf{2}}{}^{\mathbb{A}} \cong D_1$ exist for any category.

Definition. *A functor* $\mathbb{A} \xrightarrow{f} \mathbb{B}$ *is said to* **commute with inverse limits over** $\mathbb{D}$ *iff* $\lim\limits_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}$ *and* $\lim\limits_{\leftarrow \mathbb{D}}{}^{\mathbb{B}}$ *exist and the diagram*

$$\begin{tikzcd}
\mathbb{A}^{\mathbb{D}} \arrow[r, "f^{\mathbb{D}}"] \arrow[d, "\lim\limits_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}"'] & \mathbb{B}^{\mathbb{D}} \arrow[d, "\lim\limits_{\leftarrow \mathbb{D}}{}^{\mathbb{B}}"] \\
\mathbb{A} \arrow[r, "f"'] & \mathbb{B}
\end{tikzcd}$$

is commutative up to an equivalence in $\mathbb{B}^{(\mathbb{A}^{\mathbb{D}})}$. *Similarly for direct limits.* f *is said to be* **left exact** *iff* f *commutes with inverse limits over every* finite $\mathbb{D}$, *and* f *is* **left continuous** *iff it commutes with inverse limits over every* small $\mathbb{D}$. *Similarly,* **right exact**, **right continuous**, **exact**, **continuous** *are defined.*

Remark. For additive functors between abelian categories, our notions of exactness are equivalent to the customary ones, as was shown by [Freyd, 1960].

Definition. *A functor* $\mathbb{C} \xrightarrow{u} \mathbb{D}$ *will be called* **left pacing** *iff* $\mathbb{C}$ *is small and for every* $\mathbb{D} \xrightarrow{t} \mathbb{A}$ *with* $\mathbb{A}$ *left complete,*

$$\lim_{\leftarrow \mathbb{C}}{}^{\mathbb{A}}(ut) \cong \lim_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}(t) \text{ in } \mathbb{A}$$

(and in particular, the latter exists).

The following two theorems are also due in essence to [Freyd, 1960].

Theorem 3. *Let* $\mathbb{A} \xrightarrow{f} \mathbb{B}$ *be a functor with* $\mathbb{A}$, $\mathbb{B}$ *left complete. Then there exists* g *adjoint to* f *iff* f *is left continuous and for every* $B \in |\mathbb{B}|$*, there exists a small category* $\mathbb{C}_B$ *and a left pacing functor* $\mathbb{C}_B \xrightarrow{u} (B, f)$*.*

PROOF. Suppose f has an adjoint g, and suppose $\mathbb{D}$ is any small category and $\mathbb{D} \xrightarrow{t} \mathbb{A}$ is any functor. Let $L = \lim_{\leftarrow \mathbb{D}}{}^{\mathbb{A}}(t)$ and let λ_D, for $D \in |\mathbb{D}|$, denote the associated canonical maps, i.e. for every $D \xrightarrow{x} D'$ in $\mathbb{D}$, $\lambda_D(xt) = \lambda_{D'}$, where $L \xrightarrow{\lambda_D} Dt$. We wish to show that Lf, together with $\lambda_D f$ for $D \in |\mathbb{D}|$, satisfies the universal property (Theorem 2.1) which characterizes $\lim_{\leftarrow \mathbb{D}}{}^{\mathbb{B}}(tf)$. Since $(\lambda_D f)(xtf) = \lambda_{D'} f$ for every $D \xrightarrow{x} D'$ in $\mathbb{D}$, we need only show that for any family ψ_D, having the property that $\psi_D(xtf) = \psi_{D'}$, for all $D \xrightarrow{x} D'$ in $\mathbb{D}$, there is a unique map $X \xrightarrow{y} Lf$ such that for all $D \in |\mathbb{D}|$, $\psi_D = y(\lambda_D f)$, where X is the common domain of the ψ_D:

$$\begin{tikzcd}
 & & Dtf \arrow[dd, "xtf"] \\
X \arrow[urr, "\psi_D"] \arrow[r, dashed, "y"] \arrow[drr, "\psi_{D'}"'] & Lf \arrow[ur, "\lambda_D f"] \arrow[dr, "\lambda_{D'} f"'] & \\
 & & D'tf
\end{tikzcd} \quad \text{in } \mathbb{B}.$$

To establish this, note that by the definition of adjointness there is an isomorphism $(\mathbb{B}, f) \xrightarrow{h} (g, \mathbb{A})$ such that $h\overline{g} = \overline{f}$. In particular, we have

$$\begin{tikzcd} & Dt \arrow[dd, "xt"] \\ Xg \arrow[ur, "(\psi_D)h"] \arrow[dr, "(\psi_{D'})h"'] & \\ & D't \end{tikzcd} \qquad \text{in } \mathbb{A}.$$

Thus there is a unique map $Xg \xrightarrow{\overline{y}} L$ in $\mathbb{A}$ such that $(\psi_D)h = \overline{y}\lambda_D$ for all $D \in |\mathbb{D}|$. Then applying h^{-1} we get a unique $y \in \mathbb{B}$ such that $\psi_D = y(\lambda_D f)$ for all $D \in |\mathbb{D}|$, as required. Also, if f has an adjoint g, the object $B \xrightarrow{\varphi} Bgf$ is isomorphic to $1_{(B,f)}$, hence the functor $\mathbf{1} \longrightarrow (B, f)$ determined by the object φ is left pacing.

Conversely, suppose that conditions of the theorem are satisfied. The canonical $(B, f) \xrightarrow{t} \mathbb{A}$ has an inverse limit A in $\mathbb{A}$, which we will show satisfies the condition of Theorem 2.1. First we must define a map $B \xrightarrow{\varphi} Af$. For this, let λ_x denote the canonical map $A \longrightarrow xt$, for each object $B \xrightarrow{x} (xt)f$ in (B, f). For every map $x \xrightarrow{a} x'$ in (B, f)

$$\begin{tikzcd} & B \arrow[dl, "x"'] \arrow[dr, "x'"] & \\ (xt)f \arrow[rr, "af"'] & & (x't)f \end{tikzcd}$$

we have $\lambda_x a = \lambda_{x'}$. Now since A is also an inverse limit over the small category $\mathbb{C}_B$ and since f is left continuous, Af is the inverse limit of the functor tf, with associated maps $Af \xrightarrow{\lambda_x f} xtf$. Since for every map $x \xrightarrow{a} x'$ in (B, f), $x(af) = x'$ and $a = \langle x, a, x' \rangle t$, there is a unique map $B \xrightarrow{\varphi} Af$ such that $\varphi(\lambda_x f) = x$ in $\mathbb{B}$ for every $x \in |(B, f)|$. This defines φ and also shows, for each x, the existence of y satisfying the condition of Theorem 2.1; we need now only show uniqueness, i.e. if $\varphi(yf) = x$, then $y = \lambda_x$. For this consider the equalizer K of the functor $\mathbb{E} \longrightarrow \mathbb{A}$ defined by $\langle y, \lambda_x \rangle$:

$$K \xrightarrow{yE\lambda_x} A \underset{\lambda_x}{\overset{y}{\rightrightarrows}} xt.$$

Because f is left continuous, Kf is also the equalizer of $\langle yf, \lambda_x f\rangle$, and since φ 'equalizes' the latter, there is a unique z such that the left hand triangle below is commutative in $\mathbb{B}$:

$$
\begin{tikzcd}
 & & B \arrow[dll, dotted, "z"'] \arrow[d, "\varphi"'] \arrow[dr, "x"] & \\
Kf \arrow[rr, "(yE\lambda_x)f = (yf)E(\lambda_x f)"'] & & Af \arrow[r, shift left, "y"] \arrow[r, shift right, "\lambda_x"'] & xtf
\end{tikzcd}
$$

As $B \xrightarrow{z} Kf$ is an object in (B, f), there is a map $A \xrightarrow{\lambda_z} K = zt$. For any $x' \in |(B, f)|$ we have:

$$
\begin{tikzcd}
 & Kf \arrow[d, "(yE\lambda_x)f"] \\
B \arrow[ur, "z"] \arrow[r, "\varphi"] \arrow[dr, "x'"'] & Af \arrow[d, "\lambda_{x'}f"] \\
 & x'tf
\end{tikzcd}
$$

i.e. $\langle z, (yE\lambda_x)\lambda_{x'}, x'\rangle$ is a map in (B, f) at which t takes the value $(yE\lambda_x)\lambda_{x'}$. Hence

$$
\begin{tikzcd}
 & K = zt \arrow[dd, "(yE\lambda_x)\lambda_{x'}"] \\
A \arrow[ur, "\lambda_z"] \arrow[dr, "\lambda_{x'}"'] & \\
 & x't
\end{tikzcd}
$$

i.e. $(\lambda_z(yE\lambda_x))\lambda_{x'} = (A)\lambda_{x'}$ for all $x' \in |(B, f)|$, (A being an identity map). Hence, by the uniqueness stipulation of Theorem 2.1 applied to the case of $\varprojlim(t)$ we have $\lambda_z(yE\lambda_x) = A$. Then we have immediately $y = Ay = \lambda_z(yE\lambda_x)y = \lambda_z(yE\lambda_x)\lambda_x = A\lambda_x = \lambda_x$, proving that $y = \lambda_x$ is the unique $A \longrightarrow xt$ such that $\varphi(yf) = x$ in $\mathbb{B}$. Since this is true for all $B \xrightarrow{x} xt$ in $\mathbb{B}$, the condition of Theorem 2.1 is true, so that there exists g adjoint to f. ∎

The dual of Theorem 2.3 implies in particular that a functor which has a co-adjoint must be right continuous (if its domain and codomain categories are complete).

Theorem 4. *A functor $\mathbb{A} \xrightarrow{f} \mathbb{B}$, with $\mathbb{A}$, $\mathbb{B}$ left complete has an adjoint if and only if f commutes with equalizers and all small products and for every $B \in |\mathbb{B}|$, there exists a small set $\mathbb{S}_B$ of objects in $\mathbb{A}$ and maps $B \xrightarrow{v_A} Af$, $A \in \mathbb{S}_B$, such that for every $A' \in |\mathbb{A}|$ and for every $B \xrightarrow{x} A'f$ in $\mathbb{B}$ there is some $A \in \mathbb{S}_B$ and a map $A \xrightarrow{y} A'$ in $\mathbb{A}$ such that $x = v_A(yf)$.*

Proof. The necessity of conditions is clear by Theorem 2.3. The first condition is clearly equivalent to left continuity, by Theorem 2.2. Thus by Theorem 2.3 we need only construct a small left pacing $\mathbb{C}_B \xrightarrow{v} (B, f)$ to complete the proof of Theorem 2.4. Now the second condition of the theorem may be phrased thus: there is a small set $\mathbb{S}_B$ and a functor $\mathbb{S}_B \xrightarrow{v} (B, f)$ such that the property **(P)** holds for $u = v$ (by taking $\overline{x} = A$, $\overline{y} = yf$).

(P) For every object x in the codomain of u, there is an object $\overline{x}$ in the domain of u and a map $\overline{x}u \xrightarrow{\overline{y}} x$ in the codomain of u.

Since it is clear that property **(P)** also holds for u in the diagram below, where $\mathbb{C}_B$ is the full subcategory of (B, f) determined by the image of v,

$$\begin{tikzcd}
\mathbb{S}_B \arrow[rr, "v"] \arrow[dr, dotted] & & (B, f) \\
& \mathbb{C}_B \arrow[ur, "u"'] &
\end{tikzcd}$$

the following lemma proves Theorem 2.4; it is also clear that since $\mathbb{A}$ is complete, $\mathbb{D} = (B, f)$ **has pseudomeets** in the sense that for every pair of maps $D' \xrightarrow{d'} D$, $D'' \xrightarrow{d''} D$ in $\mathbb{D}$ with common codomain, there exists a commutative diagram:

$$\begin{tikzcd}
\overline{D} \arrow[r, dashed] \arrow[d, dashed] & D'' \arrow[d, "d''"] \\
D' \arrow[r, "d'"'] & D
\end{tikzcd}$$

Lemma. *Let $\mathbb{C} \xrightarrow{u} \mathbb{D}$ be a full functor with property* **(P)** *where $\mathbb{C}$ is small and $\mathbb{D}$ has pseudomeets. Then u is left pacing.*

Proof of Lemma. Consider $\mathbb{D} \xrightarrow{t} \mathbb{A}$, $\mathbb{A}$ left complete; we must show $\overleftarrow{ut} \cong \overleftarrow{t}$ (so that in particular $\overleftarrow{t}$ exists). By **(P)** we can, for each $D \in |\mathbb{D}|$, choose a definite $C_D u \xrightarrow{\overline{y}_D} D$ ($C_D \in |\mathbb{C}|$) and then define $\overline{\lambda}_D$ to be the composition

$$\overleftarrow{ut} \xrightarrow{\lambda_{C_D}} C_D ut \xrightarrow{\overline{y}_D t} Dt$$

where λ_C are the canonical maps associated with $\overleftarrow{ut}$. We wish to show that for every $D \xrightarrow{d} D'$ in $\mathbb{D}$, $\overline{\lambda}_D(dt) = \overline{\lambda}_{D'}$. Since $\mathbb{D}$ has pseudomeets we can find $\overline{D}$ and maps such that

$$\begin{tikzcd}
 & & C_D u \arrow[r, "\overline{y}_D"] & D \arrow[dd, "d"] \\
C_{\overline{D}}u \arrow[r, "\overline{y}_{\overline{D}}"] & \overline{D} \arrow[ur] \arrow[dr] & & \\
 & & C_{D'}u \arrow[r, "\overline{y}_{D'}"'] & D'
\end{tikzcd}$$

is commutative. Since u is full, the composite maps $Cu \longrightarrow C_D u$ and $Cu \longrightarrow C_{D'}u$ 'come from' $\mathbb{C}$. Thus on applying t we find that the parts, and hence the whole of the diagram below are commutative:

$$\begin{tikzcd}
 & & C_D ut \arrow[r, "\overline{y}_D t"] & Dt \arrow[dd, "dt"] \\
\overleftarrow{ut} \arrow[urr, "\lambda_{C_D}"] \arrow[r, "\lambda_{C_{\overline{D}}}"] \arrow[drr, "\lambda_{C_{D'}}"'] & C_{\overline{D}}ut \arrow[ur] \arrow[dr] & & \\
 & & C_{D'}ut \arrow[r, "\overline{y}_{D'}t"'] & D't
\end{tikzcd}$$

Thus $\overline{\lambda}_D(dt) = \overline{\lambda}_{D'}$. We now show the universality of $\overline{\lambda}$. Suppose that for each $D \in |\mathbb{D}|$, $X \xrightarrow{\psi_D} Dt$, and for every $D \xrightarrow{d} D'$, $\psi_D(dt) = \psi_{D'}$ in $\mathbb{A}$. Then in particular $\psi_{Cu}(ct) = \psi_{C'u}$ for every $C \xrightarrow{c} C'$ in $\mathbb{C}$. Therefore there is a unique z such that

$X \xrightarrow{z} \overleftarrow{u}t$ and for all $C \in |\mathbb{C}|$, $\psi_{Cu} = z\lambda_C$ in $\mathbb{A}$. But then we have (in $\mathbb{A}$)

$$z\overline{\lambda}_D = z\lambda_{C_D}(\overline{y}_d t) = \psi_{C_D u}(\overline{y}_D t) = \psi_D.$$

Since z is also the unique map such that the latter relation holds, the proof of the lemma, and hence of Theorem 2.4, is complete. ■

Remark. In view of the above two theorems, it is a reasonable conjecture that any given left continuous functor has an adjoint. However, as shown by [Gaifman, 1961], the inclusion of complete small Boolean algebras into all small Boolean algebras does not have an adjoint; hence this family of conjectures cannot be made into a general theorem which omits 'smallness' hypotheses like those of Theorems 2.3 and 2.4. The above theorems also suggest that inverse limits are in a sense the 'canonical' means for constructing adjoints, whereas the usual constructions of particular adjoint in algebra often look like direct limits (e.g. the tensor product is a quotient of a sum). Some light is shed on this 'mystery' by Theorem 2.5 below, together with the observation (substantiated by Chapters III and IV of this paper) that the common functors in algebra are usually closely associated with induced functors between functor categories.

We mention some propositions concerning adjoints of such functors before stating and proving our theorem.

Proposition 1. *Let $\mathbb{A}$ be any small category, $\mathcal{X} \xrightarrow{T} \mathcal{Y}$ any functor with an adjoint $\hat{T}$. Then the induced functor*

$$\mathcal{X}^{\mathbb{A}} \xrightarrow{T^{\mathbb{A}}} \mathcal{Y}^{\mathbb{A}}$$

has the adjoint $\hat{T}^{\mathbb{A}}$.

Proposition 2. *If $\mathcal{X}$ is left complete and $\mathbb{A}$ small, then $\mathcal{X}^{\mathbb{A}}$ is left complete. If $\mathbb{1} \xrightarrow{A} \mathbb{A}$ is any object in $\mathbb{A}$, then the 'evaluation' functor*

$$\mathcal{X}^{\mathbb{A}} \xrightarrow{\mathcal{X}^{A}} \mathcal{X}$$

is left continuous (i.e limits in $\mathcal{X}^{\mathbb{A}}$ are computed 'pointwise').

Proposition 3. *If $\mathcal{X}$ is left complete and $\mathbb{C}$, $\mathbb{D}$ are any small categories, then*

$$\underset{\leftarrow\mathbb{C}}{\lim}^{\mathcal{X}^{\mathbb{D}}} \underset{\leftarrow\mathbb{D}}{\lim}^{\mathcal{X}} \cong \underset{\leftarrow\mathbb{D}}{\lim}^{\mathcal{X}^{\mathbb{C}}} \underset{\leftarrow\mathbb{C}}{\lim}^{\mathcal{X}}$$

is a natural equivalence of functors $\mathcal{X}^{\mathbb{C}\times\mathbb{D}} \longrightarrow \mathcal{X}$.

The above propositions have obvious dualizations. Proofs will be found, e.g. in [Gray, 1962]

Theorem 5. *Let $\mathcal{X}$ be complete, $\mathbb{A}$, $\mathbb{B}$ small, and let $\mathbb{B} \xrightarrow{f} \mathbb{A}$ be* any *functor. Then the induced functor*

$$\mathcal{X}^{\mathbb{A}} \xrightarrow{\mathcal{X}^f} \mathcal{X}^{\mathbb{B}}$$

has an adjoint. More explicitly, if $\mathbb{B} \xrightarrow{U} \mathcal{X}$ is any functor, then the value $\overline{U}$ of the adjoint at U is given by the formula

$$A\overline{U} = \varinjlim_{(f,A)}{}^{\mathcal{X}}(d_0^A U)$$

where $\mathbf{1} \xrightarrow{A} \mathbb{A}$ is any object in $\mathbb{A}$ and where d_0^A is the canonical functor in the meet diagram:

$$\begin{tikzcd}
(f,A) \arrow[r] \arrow[d, "d_0^A"'] & \mathbb{A}^{\mathbf{2}} \arrow[d, "D_0"] \\
\mathbb{B} \arrow[r, "f"'] & \mathbb{A}
\end{tikzcd}$$

Proof. We use again the characterization of Theorem 2.1. Let $\overline{U}$ be defined by the above formula. Since $\overline{U}\mathcal{X}^f = f\overline{U}$, our first task is to construct a map

$$U \xrightarrow{\varphi} f\overline{U}$$

in $\mathcal{X}^{\mathbb{B}}$. Since for $\mathbf{1} \xrightarrow{B} \mathbb{B}$,

$$Bf\overline{U} = \varinjlim[(f, Bf) \xrightarrow{d_0} \mathbb{B} \xrightarrow{U} \mathcal{X}],$$

we have for each map

$$\begin{tikzcd}
B'f \arrow[rr, "bf"] \arrow[dr, "x'"'] & \mathbb{A} & B''f \arrow[dl, "x''"] \\
& Bf &
\end{tikzcd}$$

in (f, Bf), maps $\lambda_{x'}^{Bf}$, $\lambda_{x''}^{Bf}$ in $\mathcal{X}$ such that:

$$\begin{tikzcd}
B'U \arrow[dd, "bU"'] \arrow[dr, "\lambda_{x'}^{Bf}"] & \\
& (Bf)\overline{U} \\
B''U \arrow[ur, "\lambda_{x''}^{Bf}"'] &
\end{tikzcd}$$

satisfying the universal properties of direct limits. In particular, taking $B = B'$ and $x' = Bf = B'f = x''$, we get a map

$$BU \xrightarrow{\varphi_B} Bf\overline{U}$$

by defining $\varphi_B = \lambda_{Bf}^{Bf}$.

We need to show that whenever $B \xrightarrow{b} B'$ in $\mathbb{B}$,

$$\begin{tikzcd}
BU \arrow[r, "\varphi_B"] \arrow[d, "bU"'] \arrow[dr, phantom, "\mathcal{X}"] & Bf\overline{U} \arrow[d, "(bf)\overline{U}"] \\
B'U \arrow[r, "\varphi_{B'}"'] & B'f\overline{U}
\end{tikzcd}$$

Now $\overline{U}$ is defined for maps $A \xrightarrow{a} A'$ in $\mathbb{A}$ as follows. Since there is an induced functor $(f, A) \longrightarrow (f, A')$, there is a map linking the direct limits over these two categories, defined uniquely by the universal property of λ^A in the following typical diagram

$$\begin{tikzcd}
BU = xd_0U \arrow[dd, "bU"'] \arrow[dr, "\lambda_x^A"] \arrow[drr, "\lambda_{xa}^{A'}"] & & \\
& A\overline{U} \arrow[r, dashed, "a\overline{U}"] & A'\overline{U} \\
B'U = x'd_0U \arrow[ur, "\lambda_{x'}^A"'] \arrow[urr, "\lambda_{x'a}^{A'}"'] & &
\end{tikzcd}$$

where

$$\begin{tikzcd} Bf \arrow[rr, "bf"] \arrow[dr, "x"'] & \mathbb{A} & B'f \arrow[dl, "x'"] \\ & A & \end{tikzcd} \xrightarrow{(f,a)} \begin{tikzcd} Bf \arrow[rr, "bf"] \arrow[dr, "xa"'] & \mathbb{A} & B'f \arrow[dl, "x'a"] \\ & A' & \end{tikzcd}$$

for a typical map in (f, A).

In particular, for $A = Bf$, $a = bf$, we have $\lambda_x^{Bf}(bf) = \lambda_{x(bf)}^{Bf}$. Taking $x = bf$, this gives

$$\lambda_{Bf}^{Bf}(bf)\overline{U} = \lambda_{bf}^{B'f}.$$

On the other hand, since

$$\begin{tikzcd} Bf \arrow[rr, "bf"] \arrow[dr, "bf"'] & & B'f \arrow[dl, "B'f"] \\ & B'f & \end{tikzcd}$$

is a map in $(f, B'f)$, we have

$$(bU)\lambda_{B'f}^{B'f} = \lambda_{bf}^{B'f}.$$

Recalling the definition of φ, this shows

$$\varphi_B(bfU) = (bU)\varphi_{B'}$$

for every $B \xrightarrow{b} B'$ in $\mathbb{B}$. Hence φ is natural, i.e. $U \xrightarrow{\varphi} f\overline{U}$ is a map in $\mathcal{X}^{\mathbb{B}}$.

Now suppose T is any object in $\mathcal{X}^{\mathbb{A}}$ and

$$U \xrightarrow{\xi} fT$$

any map in $\mathcal{X}^{\mathbb{B}}$. We wish to show that there is a unique $\overline{U} \xrightarrow{\eta} T$ in $\mathcal{X}^{\mathbb{A}}$ that $\varphi(\eta\mathcal{X}^f) = \xi$. Now by assumption

$$\begin{tikzcd} Bf \arrow[rr, "bf"] \arrow[dr, "x"'] & & B'f \arrow[dl, "x'"] \\ & A & \end{tikzcd}$$

implies

$$\begin{tikzcd}
BU \arrow[rr, "\xi_B"] \arrow[dd, "bu"'] & & BfT \arrow[dd, "bfT"'] \arrow[dr, "xT"] & \\
& & & AT \\
B'U \arrow[rr, "\xi_{B'}"'] & & B'fT \arrow[ur, "x'T"'] &
\end{tikzcd}$$

i.e. for every $x \in |(f, A)|$ we have a map $xd_0U = BU \xrightarrow{\xi_B(xT)} AT$ which commutes with every $x \xrightarrow{b} x'$. Since $A\overline{U} = \varinjlim d_0^A U$, we have a unique $A\overline{U} \xrightarrow{\eta_A} AT$ such that $\lambda_x^A \eta_A = \xi_B(xT)$ for all objects $Bf \xrightarrow{x} A$ in (f, A). For every $A \xrightarrow{a} A'$ in $\mathbb{A}$, $\lambda_x^A \eta_A(aT) = \xi_B(xT)(aT) = \xi_B(xa)T = \lambda_{xa}^{A'} \eta_{A'} = \lambda_x^A (a\overline{U})\eta_{A'}$. That is $\eta_A(aT) = (a\overline{U})\eta_{A'}$ at each λ_x^A, hence by uniqueness η is natural, i.e. a map

$$\overline{U} \xrightarrow{\eta} T$$

in $\mathcal{X}^{\mathbb{A}}$. A particular case of the foregoing calculation is that associated with the object $Bf \xrightarrow{Bf} Bf$ in (f, Bf). We have

$$\xi_B = \xi_B(Bf)T = \lambda_{Bf}^{Bf} \eta_{Bf} = \varphi_B \eta_{Bf} = \varphi_B(\eta \mathcal{X}^f)_B$$

for every $B \in |\mathbb{B}|$, i.e. $\varphi(\eta \mathcal{X}^f) = \xi$ as required. However, since the latter is only a special case of the condition which originally defined η, its uniqueness may be in doubt; but this follows from $\varphi(\eta \mathcal{X}^f) = \xi$, together with the required fact that η is natural. That is, from

$$\begin{aligned} \xi_B &= \lambda_{Bf}^{Bf} \eta_{Bf} \\ \eta_A(aT) &= (a\overline{U})\eta_{A'}\ , \text{ for all } A \xrightarrow{a} A' \end{aligned}$$

it *follows* that

$$\xi_B(xT) = \lambda_{Bf}^{Bf} \eta_{Bf} xT = \lambda_{Bf}^{Bf}(x\overline{U}\eta_A) = \lambda_{(Bf)x}^{A} \eta_A = \lambda_x^A \eta_A$$

for any $Bf \xrightarrow{x} A$, and the latter condition *does* determine η. ■

In particular, if $\mathbb{B} = \mathbf{1}$, $\mathbf{1} \xrightarrow{f=A_0} \mathbb{A}$, then $(f, A) = (A_0, A)$ is a set, $d_0 U$ is constant, so $\varinjlim$ is the (A_0, A)-fold coproduct. Thus we have the

Corollary. *If $\mathcal{X}$ is complete, $\mathbb{A}$ small, $A_0 \in |\mathbb{A}|$, then the evaluation at A_0*

$$\mathcal{X}^{\mathbb{A}} \longrightarrow \mathcal{X}$$

has an adjoint. For each $X \in \mathcal{X}$, the value of the adjoint at X is the functor whose value at $A \in |\mathbb{A}|$ is $(A_0, A) \cdot X$.

Thus for every $\mathbb{A} \xrightarrow{T} \mathcal{X}$ we have

$$(H^{A_0} \cdot X, T) \cong (X, A_0 T)$$

where $\mathbb{A} \xrightarrow{H^{A_0}} \mathcal{S}_1$ is the functor whose value at A is (A_0, A) and where $H^{A_0} \cdot X$ is the functor $\mathbb{A} \longrightarrow \mathcal{X}$, its value at A is the (A, A_0)-fold coproduct of X with itself.

In particular, taking $\mathcal{X} = \mathcal{S}_1$, $X = 1_{\mathcal{S}_1}$, we have another

Corollary. *For every functor $\mathbb{A} \xrightarrow{T} \mathcal{S}_1$ where $\mathbb{A}$ is small, and for every $A_0 \in |\mathbb{A}|$,*

$$(H^{A_0}, T) = A_0 T.$$

Corollary. *For any small $\mathbb{A}$, the functor*

$$\mathbb{A}^* \longrightarrow \mathcal{S}_1^{\mathbb{A}}$$

which takes A_0 to H^{A_0} is full.

3. Regular epimorphisms and monomorphisms

In this section we work *in* an arbitrary but fixed category, which we will presently require to have finite limits.

Definition. *A map $K \xrightarrow{k} A$ is said to be a* **regular monomap** *iff there exist $A \overset{f}{\underset{g}{\rightrightarrows}} B$ such that $k = fEg$. Dually, a regular epimap is any map having the properties of a co-equalizer.*

Remark. The notions of regular monomaps and epimaps seem better suited (say in the category of topological spaces) for discussing subobjects and quotient objects than do the more inclusive notions of monomorphisms and epimorphisms. Clearly in any category every retract is a regular monomap and every regular monomap is a monomorphism, (and dually); all these notions, however, can be different. We require in this paper only two or three propositions from the theory of regular epimaps and monomaps.

Proposition 1. *If k is an epimorphism and also a regular monomap, then k is an isomorphism.*

Proof. If $k = fEg$, then since k is an epimorphism $f = g$. Hence $K \underset{k}{\cong} A$. ∎

For the next two propositions assume that our category has finite limits.

Proposition 2. *A map k is a regular monomap iff $k = (j_1 q)E(j_2 q)$ where $q = (kj_1)E^*(kj_2)$.*

$$\begin{tikzcd} K \arrow[r, "k"] & A \arrow[r, shift left, "j_1"] \arrow[r, shift right, "j_2"'] & A \star A \arrow[r, "q"] & Q \end{tikzcd}$$

Proof. Suppose $k = fEg$. Define t by

$$\begin{tikzcd} A \arrow[r, "j_1"] \arrow[dr, "f"'] & A \star A \arrow[d, dashed, "t"'] & A \arrow[l, "j_2"'] \arrow[dl, "g"] \\ & B & \end{tikzcd}$$

and let $h = (j_1 q)E(j_2 q)$. Then obviously $k \leq h$. To show $h \leq k$, note that $kj_1 t = kj_2 t$ since $k = (j_1 t)E(j_2 t)$. That is, t 'coequalizes' kj_1, kj_2. Hence $\exists! Q \xrightarrow{u} B$ such that $t = qu$. Then

$$hf = hj_1 t = hj_1 qu = hj_2 qu = hj_2 t = hg$$

where the third equation follows from $hj_1 q = hj_2 q$. Therefore $\exists! M \xrightarrow{z} K$ such that $zk = h$, i.e $h \leq k$.

$$\begin{tikzcd} K \arrow[r, "k"] \arrow[d, dashed, shift right] & A \arrow[r, shift left, "j_1"] \arrow[r, shift right, "j_2"'] \arrow[d, shift right, "f"'] \arrow[d, shift left, "g"] & A \star A \arrow[r, "q"] \arrow[dl, "t"] & Q \arrow[dll, dotted, "u"] \\ H \arrow[u, dashed, shift right, "z"'] \arrow[ur, "h"] & B & & \end{tikzcd}$$

∎

Dually an epimap is regular iff it is the coequalizer of the relation it induces on its domain.

Proposition 3. *If k_1 is a monomorphism and $k_2 k_1$ is a regular monomap, then k_2 is a regular monomap, and dually.*

PROOF. Consider

$$\begin{aligned} q_3 &= (k_2k_1j_1)E^*(k_2k_1j_2) \\ k_2k_1 &= (j_1q_3)E(j_2q_3) \\ q_2 &= (k_2i_1)E^*(k_2i_2) \end{aligned}$$

$$\begin{tikzcd}
 & K_2 \arrow[dd, "k_2"] & \\
X \arrow[ur, dotted, "z"] \arrow[dr, "x"'] & & \\
 & K_1 \arrow[r, "k_1"] \arrow[d, shift right, "i_1"'] \arrow[d, shift left, "i_2"] & A \arrow[d, shift right, "j_1"'] \arrow[d, shift left, "j_2"] \\
 & K_1 \star K_1 \arrow[r, "k_1 \star k_1"] \arrow[d, "q_2"'] & A \star A \arrow[d, "q_3"] \\
 & Q_2 \arrow[r, dashed, "v"'] & Q_3
\end{tikzcd}$$

We must show $k_2 = (i_1q_2)E(i_2q_2)$. Now k_2 'equalizes' i_1q_2, i_2q_2 by definition of q_2. We get v since $(k_1 \star k_1)q_3$ coequalizes k_2i_1, k_2i_2, i.e. is a 'candidate' for q_2. Thus if $xi_1q_2 = xi_2q_2$, it follows that xk_1 equalizes j_1q_3, j_2q_3, i.e. is a candidate for k_2k_1. Therefore $\exists! z[xk_1 = zk_2k_1]$. But k_1 is a monomorphism, so $x = zk_2$. Finally, z satisfying the last equation is unique, since k_2 is a monomorphism. ■

Chapter II

Algebraic theories

1. The category of algebraic theories

Before discussing the category of algebraic theories, we briefly consider the category $(\{\mathcal{S}_0\}, \mathcal{C}_1)$, of which the category of algebraic theories will be a subcategory. Recall that the maps in the category $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ may be identified as commutative triangles

$$\begin{tikzcd} & \mathcal{S}_0 \arrow[dl, "A"'] \arrow[dr, "B"] & \\ \mathbb{A} \arrow[rr, "f"'] & & \mathbb{B} \end{tikzcd}$$

of functors where $\mathcal{S}_0$ is (any fixed *small* version of) the category of finite sets and $\mathbb{A}$, $\mathbb{B}$ are any small categories. For definiteness suppose $|\mathcal{S}_0| \cong N$.

Definition. *Let $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ be an object in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$. If n is any object in $\mathcal{S}_0$, we will denote by ${}_nA$ the value at n of A; thus ${}_nA$ is an object in $\mathbb{A}$. If σ is any map in $\mathcal{S}_0$, we will sometimes simply write σ for the value of σ at A. However, in the special case of maps $1 \xrightarrow{i} n$ in $\mathcal{S}_0$, we will usually write π_i^n for the value at i of A. For any objects n is $\mathcal{S}_0$, an* n**-ary operation of** $\mathbb{A}$ *means any map ${}_1A \xrightarrow{\theta} {}_nA$ in $\mathbb{A}$. (Note that the notion of n-ary operation really depends on A, not just on $\mathbb{A}$; however, in what follows we will be justified in the abuse of notation which confuses A with $\mathbb{A}$.) In particular, each π_i^n, where $i \in n$ in $\mathcal{S}_0$, is an n-ary operation of every $\mathbb{A}$.*

Proposition 1. *The category $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ has products and coproducts. In fact, the codomain functor $(\{\mathcal{S}_0\}, \mathcal{C}_1) \to \mathcal{C}_1$ is left continuous and (binary) coproducts in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ are defined* *by comeet diagrams of the form*

$$\begin{tikzcd}
\mathcal{S}_0 \arrow[r, "B"] \arrow[d, "A"'] \arrow[dr, phantom, "\mathcal{C}_1"] & \mathbb{B} \arrow[d] \\
\mathbb{A} \arrow[r] & \mathbb{A} \underset{\mathcal{S}_0}{\star} \mathbb{B}
\end{tikzcd}$$

PROOF. These assertions in fact remain valid if $\mathcal{C}_1$ is replaced by any complete category and $\{\mathcal{S}_0\}$ by any object in it. The left continuity statement is obvious. We explicitly verify the 'coproduct = comeet in $\mathcal{C}_1$' assertion. Let $A \xrightarrow{f} C$, $B \xrightarrow{g} C$ be any maps in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$. Since this implies that $Af = C = Bg$ in $\mathcal{C}_1$, and since $\mathbb{A} \underset{\mathcal{S}_0}{\star} \mathbb{B}$ is a comeet in $\mathcal{C}_1$, there exists a unique h in $\mathcal{C}_1$ such that

$$\begin{tikzcd}
\mathcal{S}_0 \arrow[r, "B"] \arrow[d, "A"'] \arrow[dr, "A \underset{\mathcal{S}_0}{\star} B"] & \mathbb{B} \arrow[d, "k"] \arrow[ddr, "g"] & \\
\mathbb{A} \arrow[r, "j"] \arrow[drr, "f"'] & \mathbb{A} \underset{\mathcal{S}_0}{\star} \mathbb{B} \arrow[dr, dotted, "h"] & \\
 & & \mathbb{C}
\end{tikzcd}$$

is commutative in $\mathcal{C}_1$. But then h defines the unique map $\mathbb{A} \underset{\mathcal{S}_0}{\star} \mathbb{B} \xrightarrow{h} \mathbb{C}$ in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ such that $f = jh$ and $g = kh$, i.e. $\mathcal{S}_0 \xrightarrow{A \underset{\mathcal{S}_0}{\star} B} \mathbb{A} \underset{\mathcal{S}_0}{\star} \mathbb{B}$ is the coproduct in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ of A, B. (The same proof clearly works for infinite coproducts.) In view of the nature of comeets in $\mathcal{C}_1$, this means that in the $(\{\mathcal{S}_0\}, \mathcal{C}_1)$-coproduct any map $n \xrightarrow{x} m$ is represented by a string

$$n \xrightarrow{x_0} n_0 \xrightarrow{x_1} n_1 \to \cdots \to n_{\ell-2} \xrightarrow{x_{\ell-1}} m$$

where each x_i is a map $n_{i-1} \to n_i$ in either $\mathbb{A}$ or $\mathbb{B}$ (n_{-1} being n and $n_{\ell-1}$ being m); the only relations imposed on strings are that $\langle \sigma A \rangle = \langle \sigma B \rangle$ for all $\sigma \in \mathcal{S}_0$ and that $\langle x_0 x_1 \rangle = \langle y_0 \rangle$ if $x_0 x_1 = y_0$ in $\mathbb{A}$ or in $\mathbb{B}$ (and consequences of these relations). ∎

Definition. The category $\mathcal{T}$ of algebraic theories *is the full subcategory of* $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ *determined by those objects* $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ *such that* A *commutes with finite coproducts and such that* $|A|$ *is an isomorphism.*

Thus for an algebraic theory $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ we have in essence that every object in $\mathbb{A}$ is of the form ${}_nA$ and that furthermore ${}_{(n+m)}A = {}_nA \star {}_mA$ in $\mathbb{A}$ for all $n, m \in |\mathcal{S}_0|$. In view of the structure of $\mathcal{S}_0$, the latter condition is equivalent to

$$ {}_nA = n \cdot {}_1A $$

for every $n \in |\mathcal{S}_0|$, where the right-hand side is the n-fold coproduct of ${}_1A$ with itself in $\mathbb{A}$. It follows immediately that ${}_1A$ is a *generator* for $\mathbb{A}$, and that ${}_0A = \varinjlim_{\mathbf{0}}^{\mathbb{A}}$ (i.e. $\forall n\ \exists!\ [{}_0A \longrightarrow {}_nA]$). The equation ${}_nA = n \cdot {}_1A$ implies that the maps $n \longrightarrow m$ in $\mathbb{A}$ are in one-to-one correspondence with n-tuples of m-ary operations of $\mathbb{A}$.

Proposition 2. *Let $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ be an algebraic theory and consider the map ${}_0A \longrightarrow {}_1A$ in $\mathbb{A}$. This map is always a monomorphism, and if there exists x such that ${}_1A \xrightarrow{x} {}_0A$ in $\mathbb{A}$ then ${}_0A \longrightarrow {}_1A$ is a retract.*

Proof. If there is no such x, then there is no map from any ${}_nA$ to ${}_0A$ and so ${}_0A \longrightarrow {}_1A$ is vacuously a monomorphism. If there is such an x, then ${}_0A \longrightarrow {}_1A \xrightarrow{x} {}_0A$ must be the identity since ${}_0A = \varinjlim_{\mathbf{0}}^{\mathbb{A}}$. ■

A particular example of an algebraic theory is the identity functor $\mathcal{S}_0$, which we will sometimes call 'the theory of equality'. It is clear that $\mathcal{S}_0 \cong \varinjlim_{\mathbf{0}}^{\mathcal{T}}$. Any theory $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ such that $\mathbb{A}$ is equivalent to either $\mathbf{2}$ or $\mathbf{1}$ will be called **inconsistent**.

Proposition 3. *Let $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$ be an algebraic theory. If there are $m \underset{\tau}{\overset{\sigma}{\rightrightarrows}} n'$ in $\mathcal{S}_0$, $\sigma \neq \tau$, such that $\sigma A = \tau A$, then the theory is inconsistent.*

Proof. For any $n \geq 2$, there is $n' \xrightarrow{\tau'} n$ such that $\sigma\tau' \neq \tau\tau'$ and $1 \xrightarrow{j} m$ such that $i_0 = j\sigma\tau' \neq j\tau\tau' = i_0'$, but $\pi^n_{i_0} = \pi^n_{i_0'}$. There are also $n \xrightarrow{\overline{\sigma}} n-1$ and $n-1 \xrightarrow{\overline{\tau}} n$ such that $i_0\overline{\sigma} = i_0'\overline{\sigma}$ and $\overline{\tau}\overline{\sigma} = n-1$. Then for all $1 \xrightarrow{i} n$, $\pi^n_{i\overline{\sigma}\overline{\tau}} = \pi^n_i$. Since ${}_{n-1}A$ and ${}_nA$ are coproducts, there are unique ${}_nA \xrightarrow{f} {}_{n-1}A$ and ${}_{n-1}A \xrightarrow{g} {}_nA$ such that

$$\begin{array}{rcll} \pi^{n-1}_{i\overline{\sigma}} & = & \pi^n_i f & \text{for all } i \in n \\ \pi^n_{j\overline{\tau}} & = & \pi^{n-1}_j g & \text{for all } j \in n-1. \end{array}$$

Now we have $gf = {}_{n-1}A$ in any case, and due to our hypotheses we have $\pi^n_i fg = \pi^{n-1}_{i\overline{\sigma}} = \pi^n_{i\overline{\sigma}\overline{\tau}} = \pi^n_i$ for all $1 \xrightarrow{i} n$ and thus

$$ {}_{n-1}A \cong {}_nA \text{ for all } n \geq 2. $$

It follows that

$$ {}_nA \cong {}_1A \text{ for all } n \geq 1, $$

where the isomorphisms are all induced by maps coming from $\mathcal{S}_0$. From this it follows that $\pi^2_0 = \pi^2_1$, since we always have $\pi^2_0\sigma = {}_1A = \pi^2_0\sigma$ for $2 \xrightarrow{\sigma} 1$, and in our case σA has an inverse. From this it is immediate that

$$({}_nA, {}_mA) = 1$$

for any n and for $m \neq 0$. But this implies that $\mathbb{A}$ is equivalent to either $\mathtt{1}$ or $\mathtt{2}$. ∎

Now, since the compositions $0 \longrightarrow 1 \rightrightarrows 2$ are equal in $\mathcal{S}_0$, for any algebraic theory $\mathbb{A}$ we have equal compositions

$${}_0A \longrightarrow {}_1A \rightrightarrows {}_2A$$

in $\mathbb{A}$ and hence equal compositions of maps between sets

$$({}_1A, {}_0A) \longrightarrow ({}_1A, {}_1A) \rightrightarrows ({}_1A, {}_2A).$$

However, $({}_1A, {}_0A)$ need not to be the equalizer of the other two maps. If we denote this equalizer by K, then members of K are called **definable constants** of the theory. Clearly every expressible constant determines a definable one via $0 \longrightarrow 1$. Explicitly, a definable constant is a unary operation θ such that $\theta\pi^2_0 = \theta\pi^2_1$; an expressible constant is any zero-ary operation.

Proposition 4. *If θ is a definable constant of an algebraic theory $\mathbb{A}$ and if φ is any unary operation, then $\varphi\theta$ is a definable constant and $\theta\varphi = \theta$.*

Proof. The first assertion is obvious and the second follows from the diagram

$$\begin{tikzcd}
& {}_1A \arrow[r, "\varphi"] \arrow[d, "\pi^2_0"'] & {}_1A \arrow[d, "\pi^2_0"'] \arrow[dr, "{}_1A"] & \\
{}_1A \arrow[ur, "\theta"] \arrow[dr, "\theta"'] & {}_2A \arrow[r, dashed] & {}_2A \arrow[r, dashed] & {}_1A \\
& {}_1A \arrow[u, "\pi^2_1"] \arrow[r, "{}_1A"'] & {}_1A \arrow[u, "\pi^2_1"] \arrow[ur, "{}_1A"'] &
\end{tikzcd}$$

∎

Proposition 5. *For any algebraic theory* $\mathcal{S}_0 \xrightarrow{A} \mathbb{A}$*, either every definable constant is expressible or none are. That is, either* $({}_1A, {}_0A) \cong K$ *or* $({}_1A, {}_0A) = 0$.

Proof. Suppose there exists ${}_1A \xrightarrow{x} {}_0A$ and let θ be any definable constant. Then by the second assertion of Proposition 1.4, the composite

$$ {}_1A \xrightarrow{\theta} {}_1A \xrightarrow{x} {}_0A \longrightarrow {}_1A $$

equals θ, i.e. θx expresses θ. ∎

Note that the 'expression' is faithful by Proposition 1.2.

Remark. Evidently one could 'complete' (or deplete) algebraic theories with regard to expressibility of constants; however, there seems to be no need to do so. As an example, the algebraic theory of groups can be 'presented' (see Section 2) in two ways, one involving a single binary operation $x, y \longrightarrow x \cdot y^{-1}$ as generator, and the second involving a 0-ary generator e, a unary generator $x \longrightarrow x^{-1}$, and a binary generator $x, y \longrightarrow x \cdot y$. This actually gives two theories, for in the first case no constant is expressible and in the second case the (only) constant e is expressible. These two theories also give rise to different algebraic categories (see Chapter III), for according to the first theory the empty set is a group, whereas to the second it is not.

Theorem 1. *The category* $\mathcal{T}$ *of algebraic theories is complete. Neither the first functor nor the composite in the diagram below*

$$ \mathcal{T} \longrightarrow (\{\mathcal{S}_0\}, \mathcal{C}_1) \longrightarrow \mathcal{C}_1 $$

is left continuous; coproducts and coequalizers are as described in Lemma 1.1 and Lemma 1.2 below.

Proof. The completeness of the middle category follows from two facts

(1) coproducts are computed coordinate-wise in a product of categories

(2) if $k = fEg$ in $\mathcal{C}_1$

$$\begin{tikzcd}
 & \mathcal{S}_0 \arrow[dl, dotted, "K"'] \arrow[d, "A"] \arrow[dr, "B"] & \\
\mathbb{K} \arrow[r, "k"'] & \mathbb{A} \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & \mathbb{B}
\end{tikzcd}$$

where A and B are algebraic theories, then K commutes with finite coproducts. To see this consider ${}_1K \xrightarrow{\theta_i} {}_mK$, $\theta_i \in \mathbb{K}$, $i \in n$. There is a unique $y \in \mathbb{A}$ such that

$$\begin{tikzcd}
{}_1Kk \arrow[d, "\pi_i^n"'] \arrow[dr, "\theta_i k"] & \\
{}_nKk \arrow[r, dashed, "y"'] & {}_mKk
\end{tikzcd}$$

We need only show $\exists x \in \mathbb{K}\ [y = xk]$, which will follow from the fact that $yf = yg$. But this follows at once from the fact that $\theta_i kf = \theta_i kg$, $i \in n$, and A, B, f, g commute with finite coproducts. ∎

Direct limits in $\mathcal{T}$ are described by the following lemmas.

Lemma 1. *Let Λ be any small set and $\Lambda \xrightarrow{\mathbb{A}} \mathcal{T}$ any functor. Then the coproduct $\overline{\mathbb{A}} = \underset{\lambda\in\Lambda}{\star} \mathbb{A}_\lambda$ in $\mathcal{T}$ may be constructed as follows.*

0. If $1 \xrightarrow{\theta} n$ is a map in some $\mathbb{A}_\lambda$, then $1 \xrightarrow{\theta} n$ is a map in $\overline{\mathbb{A}}$.

1. If $1 \xrightarrow{\phi_i} m$, $i \in n$ are any maps in $\overline{\mathbb{A}}$, then $\{\phi_0 \ldots \phi_{n-1}\}$ is a map $n \longrightarrow m$ in $\overline{\mathbb{A}}$.

2. If $n \xrightarrow{\phi} n' \xrightarrow{\psi} n''$ are any maps in $\overline{\mathbb{A}}$, then $n \xrightarrow{\phi\psi} n''$ is a map in $\overline{\mathbb{A}}$.

3. All maps in $\overline{\mathbb{A}}$ are represented by expressions obtained by some finite number of applications of 0,1,2. However, the following relations *are imposed on these expressions, as are all relations that follow by reflexivity, symmetry, or transitivity from (a) through (i):*

(a) *If $\phi \equiv \phi'$ and $\psi \equiv \psi'$ then $\phi\psi \equiv \phi'\psi'$,*

(b) *In some $\mathbb{A}_\lambda$,*

$$\begin{tikzcd}
1 \arrow[dd, "\pi_i^n"'] \arrow[dr, "\theta_i"] \arrow[ddrr, "\psi_i"] & 1 \arrow[d, "\pi_j^{n'}"'] \arrow[ddr, "\varphi_j"] & \\
 & n' \arrow[dr, "\varphi"] & \\
n \arrow[ur, "\theta"] \arrow[rr, "\psi"'] & & n''
\end{tikzcd}$$

$$\begin{array}{rcll} \pi_i^n\theta & = & \theta_i & , \quad i \in n \\ \pi_j^{n'}\varphi & = & \varphi_j & , \quad j \in n' \\ \pi_i^n\psi & = & \psi_i & , \quad i \in n \\ \theta\varphi & = & \psi & \end{array}$$

the θ_i being n'-ary operations, the φ_j being n''-ary operations and the ψ_i also being n''-ary operations, all of $\mathbb{A}_\lambda$, then

$$\{\theta_0 \dots \theta_{n-1}\}\{\varphi_0 \dots \varphi_{n'-1}\} \equiv \{\psi_0 \dots \psi_{n-1}\}.$$

(c) *For any $\sigma \in \mathcal{S}_0$ and any $\lambda, \lambda' \in \Lambda$, $\sigma A_\lambda \equiv \sigma A_{\lambda'}$.*

(d) $\pi_i^n\{\phi_0 \dots \phi_{n-1}\} \equiv \phi_i$, *$i \in n$, where $1 \xrightarrow{\phi_i} m$ for all $i \in n$.*

(e) *If $\theta_i \equiv \theta_i'$, $i \in n$, where θ_i, θ_i' are m-ary, then* $\{\theta_0 \dots \theta_{n-1}\} \equiv \{\theta_0' \dots \theta_{n-1}'\}$.

(f) *For any $n \xrightarrow{\phi} m$, $\phi = \{\pi_0^n\phi \dots \pi_{n-1}^n\phi\}$.*

(g) $\{\theta_0 \dots \theta_{n-1}\}\{\pi_o^m \dots \pi_{m-1}^m\} \equiv \{\theta_0 \dots \theta_{n-1}\}$
$\{\pi_0^m \dots \pi_{m-1}^m\}\{\phi_0 \dots \phi_{m-1}\} \equiv \{\phi_0 \dots \phi_{m-1}\}$
where the θ_i are all m-ary and ϕ_j are all k-ary operations of $\overline{\mathbb{A}}$.

(h) $\theta(\phi\psi) \equiv (\theta\phi)\psi$.

(i) *For each n, $\{\pi_0^n \dots \pi_{n-1}^n\} \equiv n$.*

4. Domain and codomain in $\overline{\mathbb{A}}$ were specified in the construction, and composition in $\overline{\mathbb{A}}$ is by concatenation of representative expressions. Inclusion functors $\mathbb{A}_\lambda \longrightarrow \overline{\mathbb{A}}$ are defined in the obvious fashion, and, in view of 3.(c), there is a unique $\mathcal{S}_0 \xrightarrow{\overline{A}} \overline{\mathbb{A}}$ which, for every λ, is equal to A_λ composed with the λ-th inclusion.

Proof. $\overline{\mathbb{A}}$ is clearly a small category and the inclusions $\mathbb{A}_\lambda \longrightarrow \mathbb{A}$ are clearly functors, as is $\overline{A}$. We first show that $\overline{\mathbb{A}}$ is a theory, i.e. that $\overline{\mathbb{A}}$ commutes with finite coproducts. It suffices to show

$$_n\overline{A} \cong n \cdot {_1\overline{A}}$$

the right-hand side being an n-fold coproduct in $\overline{\mathbb{A}}$. So let ϕ_i

$$\begin{tikzcd} {_1\overline{A}} \arrow[d, "\pi_i^n"'] \arrow[dr, "\phi_i"] & \\ {_n\overline{A}} \arrow[r, dashed, "\phi"'] & {_mA} \end{tikzcd}$$

be any maps in $\overline{\mathbb{A}}$ and define $\phi = \{\phi_0 \dots \phi_{n-1}\}$. By 3.(d), $\pi^n_i \phi = \phi_i$. If $\pi^n_i \phi' = \phi_i$, then $\phi' = \{\pi^n_o \phi' \dots \pi^n_{n-1}\phi'\} = \{\phi_0 \dots \phi_{n-1}\} = \phi$. Thus maps $n \xrightarrow{\phi} m$ in $\overline{\mathbb{A}}$ correspond exactly to n-tuples of m-ary operations of $\overline{\mathbb{A}}$, i.e. $\overline{A}$ commutes with finite coproducts.

We must also show that $\overline{\mathbb{A}} = \underset{\lambda\in\Lambda}{\star} \mathbb{A}_\lambda$ in $\mathcal{T}$. So let $\mathcal{S}_0 \xrightarrow{C} \mathbb{C}$ be any object in $(\{\mathcal{S}_0\}, \mathcal{C}_1)$ such that $|C|$ is an isomorphism and C commutes with finite coproducts, and let

$$\begin{tikzcd}
\mathcal{S}_0 \arrow[d, "A_\lambda"'] \arrow[dr, "C"] & \\
\mathbb{A}_\lambda \arrow[r, "f_\lambda"'] \arrow[ur, phantom, "\lambda \in \Lambda"] & \mathbb{C}
\end{tikzcd}$$

be commutative diagrams. Define $\overline{\mathbb{A}} \xrightarrow{f} \mathbb{C}$ as follows:

0. If $1 \xrightarrow{\theta} n$ in $\mathbb{A}_\lambda$, $(\theta)f = (\theta)f_\lambda$.
1. If $n \xrightarrow{\phi} n' \xrightarrow{\psi} n''$ then $(\phi\psi)f = (\phi)f(\psi)f$.
2. If $1 \xrightarrow{\phi_i} m$, $i \in n$, then $\{\phi_0 \dots \phi_{n-1}\}f = y$ where y is the unique map $n \xrightarrow{y} m$ in $\mathbb{C}$ such that $\pi^n_i y = (\phi_i)f$.

This defines f on the expressions, and by 3.(a) and 3.(e), f remains well defined on $\overline{\mathbb{A}}$. By definition f is a functor, and λ-th inclusion composed with f is f_λ. If f' has the latter two properties, then f' satisfies the conditions 0. and 1. in the definition of f. Since $\overline{A}$ and C commute with coproducts, so must f', i.e. f' satisfies the condition 2. in the definition of f. Thus $f' = f$, so that f is unique. ■

In an arbitrary algebraic theory we will sometimes use the notation introduced in Lemma 1.1, namely if $\langle\theta_0 \dots \theta_{n-1}\rangle$ is an n-tuple of m-ary operations, then $\{\theta_0 \dots \theta_{n-1}\}$ is the unique $n \longrightarrow m$ such that

$$\pi^n_i\{\theta_0 \dots \theta_{n-1}\} = \theta_i \text{ for } i \in n.$$

Definition. *By a* **congruence relation** *R in an algebraic theory $\mathbb{B}$ is meant the following.*

(0) For each n, $m \in |\mathcal{S}_0|$, $R_{n,m} \subseteq ({}_nB, {}_mB) \times ({}_nB, {}_mB)$.

(1) If $n \overset{\theta}{\underset{\theta'}{\rightrightarrows}} n' \overset{\varphi}{\underset{\varphi'}{\rightrightarrows}} n''$ in $\mathbb{B}$ with $\langle\theta, \theta'\rangle \in R_{n,n'}$ and $\langle\varphi, \varphi'\rangle \in R_{n',n''}$, then $\langle\theta\varphi, \theta'\varphi'\rangle \in R_{n,n''}$.

(2) $1 \overset{\theta_i}{\underset{\theta'_i}{\rightrightarrows}} m$ *in* $\mathbb{B}$ *and* $\langle \theta_i, \theta'_i \rangle \in R_{1,m}$ *for* $i \in n$ *implies*

$$\langle \{\theta_0 \dots \theta_{n-1}\}, \{\theta'_0 \dots \theta'_{n-1}\} \rangle \in R_{n,m}.$$

(3) *Each* $R_{n,m}$ *is reflexive, symmetric, and transitive.*

It is clear that for any algebraic theory $\mathbb{B}$ and any congruence relation R in $\mathbb{B}$, we have $\mathbb{B} \xrightarrow{\eta} \mathbb{B}/R$ in $\mathcal{T}$ such that $\langle \theta, \theta' \rangle \in R_{n,m} \Rightarrow \theta\eta = \theta'\eta$ and such that given any $\mathbb{B} \xrightarrow{f} \mathbb{C}$ in $\mathcal{T}$ with the same property, there is a unique completion of the commutative triangle

$$\begin{tikzcd} & \mathbb{B} \arrow[dl, "\eta"'] \arrow[dr, "f"] & \\ \mathbb{B}/R \arrow[rr, dashed] & & \mathbb{C} \end{tikzcd}$$

and that, furthermore, η is full (as a map of categories).

Lemma 2. *Let* $\mathbb{A} \overset{f}{\underset{g}{\rightrightarrows}} \mathbb{B}$ *in* $\mathcal{T}$. *Define* $R_{n,m}$ *to be the set of all pairs* $n \overset{\theta}{\underset{\theta'}{\rightrightarrows}} m$ *in* $\mathbb{B}$ *such that* $\varphi f = \theta$ *and* $\varphi g = \theta'$ *for some* φ *in* $\mathbb{A}$, *together with all pairs obtained from these by repeated applications of reflexivity, symmetry, transitivity, composition, and* $\{\ \}$. *Then* R *is a congruence relation in* $\mathbb{B}$ *and the natural* $\mathbb{B} \xrightarrow{\eta} \mathbb{B}/R$ *is the coequalizer of* f, g; *i.e.* $\eta = fE^*g$. *Furthermore,* $R_{n,m}$ *is also the set of all* $\langle \theta, \theta' \rangle$ *such that* $\theta\eta = \theta'\eta$

Proof. It is obvious from the definition that R is the smallest congruence relation containing the set of pairs $\langle \theta, \theta' \rangle$ for which $\varphi f = \theta$ and $\varphi g = \theta'$ for some $\varphi \in \mathbb{A}$, and the other assertions follow readily from this fact. ∎

2. Presentations of algebraic theories

We define a functor

$$\mathcal{T} \xrightarrow{T} \mathcal{S}_1^N$$

as follows. For each algebraic theory $\mathbb{A} \in |\mathcal{T}|$, $\mathbb{A}T$ is the sequence of sets whose n-th term is the set $({}_1A, {}_nA)$. That is, $(\mathbb{A}T)_n$ is the set of n-ary operations of $\mathbb{A}$. For each $\mathbb{A} \xrightarrow{f} \mathbb{B}$ in $\mathcal{T}$, fT is the sequence of maps in $\mathcal{S}_1$ such that $(fT)_n = ({}_1A, f_n) : ({}_1A, {}_nA) \longrightarrow ({}_1B, {}_nB)$. Clearly T is a functor.

Theorem 1. *There is a free algebraic theory over each sequence of small sets, i.e.* T *has an adjoint* F. *Further, given any algebraic theory* $\mathbb{A}$, *there exists a free algebraic theory* $\mathbb{F}$ *and regular epimap* $\mathbb{F} \longrightarrow \mathbb{A}$ *in* $\mathcal{T}$.

PROOF. We first consider a sequence of special cases. Let $n \in N = |\mathcal{S}_0|$. Define $\mathbb{I}_n$ to be the algebraic theory constructed as follows: adjoin a single n-ary operation to $\mathcal{S}_0$, consider all expressions formed from $\mathcal{S}_0$ and this n-ary operation by means of composition and $\{\ \}$, and impose on these expressions relations analogous to those under 3 in Lemma 1.1. Then for any algebraic theory $\mathbb{A}$,

$$(\mathbb{I}_n, \mathbb{A}) \cong ({}_1A, {}_nA) \cong (\delta_n, \mathbb{A}T)$$

where δ_n is the sequence of sets whose k-th entry is 1 if $k = n$, otherwise 0. Hence $\mathbb{I}_n$ is the free algebraic theory over δ_n.

Now for any sequence $N \xrightarrow{S} \mathcal{S}_1$ of small sets

$$S = \sum_{n \in N} S_n \cdot \delta_n.$$

Hence, since adjoint functors are right continuous,

$$SF = \mathop{\star}_{n \in N} S_n \cdot \mathbb{I}_n$$

is the free algebraic theory over S. ∎

The second assertion of Theorem 2.1 follows from the more refined statements of Lemmas 2.1 and 2.2.

Lemma 1. *For each algebraic theory $\mathbb{A}$, any $n \in N$, and any $\mathbb{I}_n \longrightarrow \mathbb{A}$ in $\mathcal{T}$, there exists a lifting*

$$\begin{tikzcd} & \mathbb{I}_n \arrow[dl, dotted] \arrow[d] \\ \mathbb{A}TF \arrow[r] & \mathbb{A} \end{tikzcd}$$

PROOF. Since $\mathbb{I}_n = \delta_n F$, the proof is immediate from the characterization of coadjoints (i.e. the dual of Theorem I.2.1) ∎

Lemma 2. *If $\mathbb{A} \xrightarrow{f} \mathbb{B}$ in $\mathcal{T}$, then f is a regular epimap iff for every $n \in N$, every $\mathbb{I}_n \longrightarrow \mathbb{B}$ admits a lifting*

$$\begin{tikzcd} & \mathbb{I}_n \arrow[dl, dotted] \arrow[d] \\ \mathbb{A} \arrow[r, "f"'] & \mathbb{B} \end{tikzcd}$$

Proof. The lifting condition is equivalent to f being onto, and it is clear from our discussion of direct limits in $\mathcal{T}$ that a map of theories is onto iff it is the coequalizer of the congruence it induces, which by Proposition I.3.2 is equivalent to being a regular epimap. ∎

Definition. *By a* **presentation** *of an algebraic theory is meant a triple $\langle S, E, f\rangle$ where $S, E \in |\mathcal{S}_1^N|$ and where $f : E \longrightarrow SFT \times SFT$ in $\mathcal{S}_1^N$. The* **theory presented by** *$\langle S, E, f\rangle$ is the coequalizer in $\mathcal{T}$ of the maps*

$$EF \rightrightarrows SF$$

corresponding to fp, fp' under the natural isomorphism $(EF, SF) \cong (E, SFT)$. Members of S_n are called **basic polynomials in n variables** *of the presentation (or, by abuse of language, of the theory presented) and members of E_n are called* **basic identities (or axioms) in n variables** *of the presentation. Members of $(SFT)_n$ are called* **polynomials in n variables** *of the presentation, and members of EFT are called* **identities** *or* **theorems** *of the presentation. Clearly every axiom determines a theorem by means of the canonical $E \longrightarrow EFT$. The particular polynomial π_i^n in n variables will be called the* **i-th n-ary variable**. *Note that we avoid the usual practice of lumping together the n-ary variables for various n. The n-ary variables and the m-ary variables are related only by specified 'substitutions' σ.*

For any polynomial θ in n variables, we have that

$$\theta \equiv \theta\{\pi_0^n \dots \pi_{n-1}^n\}$$

is a theorem. More generally if φ_i is a polynomial in m variables for each $i \in n$, then the polynomial

$$\theta\{\varphi_0 \dots \varphi_{n-1}\}$$

in m variables is called **the result of substituting φ_i for the i-th variable in θ**. (This order of writing for polynomial composition is the one consistent with writing xf for the value at x of a homomorphism f (see Chapter V).) Note that for a given presentations $\langle S, E, f\rangle$, the polynomials in n variables of the presentation are mapped canonically onto the set of n-ary operations of the presented algebraic theory $\mathbb{A}$ by $(\eta T)_n$ where $EF \rightrightarrows SF \xrightarrow{\eta} \mathbb{A}$ is the coequalizer diagram defining $\mathbb{A}$.

Example. *The theory of associative rings with unity* is presented as follows (writing $+$ and $\cdot$ between the arguments, as usual, rather than writing the operation symbol in front or back).

n	S	E
0	$\odot$, 1	empty
1	–	$\pi_0^1 + (-\pi_0^1) \equiv (\odot)(0 \longrightarrow 1)$ $(\odot)(0 \longrightarrow 1) + \pi_0^1 \equiv \pi_0^1$, $\pi_0^1 + (\odot)(0 \longrightarrow 1) \equiv \pi_0^1$ $\pi_0^1 \cdot (1)(0 \longrightarrow 1) \equiv \pi_0^1$, $(1)(0 \longrightarrow 1) \cdot \pi_0^1 \equiv \pi_0^1$
2	$+, \cdot$	empty
3	empty	$\pi_0^3 \cdot (\pi_1^3 \cdot \pi_2^3) \equiv (\pi_0^3 \cdot \pi_1^3) \cdot \pi_2^3$
4	empty	$(\pi_0^4 + \pi_1^4) + (\pi_2^4 + \pi_3^4) \equiv (\pi_0^4 + \pi_2^4) + (\pi_1^4 + \pi_3^4)$ $(\pi_0^4 + \pi_1^4) \cdot (\pi_2^4 + \pi_3^4) \equiv ((\pi_0^4 \cdot \pi_2^4) + (\pi_1^4 \cdot \pi_2^4)) + ((\pi_0^4 \cdot \pi_3^4) + (\pi_1^4 \cdot \pi_3^4))$

$S_n = E_n = 0$ for $n \geq 5$.

Proposition. *Let $\langle S, E, f\rangle, \langle S', E', f'\rangle$ be presentations of algebraic theories. Let $S \xrightarrow{h} S'FT$ be a map in $\mathcal{S}_1^N$ (i.e a sequence of functions) such that there exists a map $E \xrightarrow{g} E'FT$ for which*

$$\begin{tikzcd}
E \arrow[r, dashed, "g"] \arrow[d, "f"'] & E'FT \arrow[d] & E' \arrow[l] \arrow[dl, "f'"] \\
(SF \times SF)T \arrow[r, "(\overline{h} \times \overline{h})T"'] & (S'F \times S'F)T &
\end{tikzcd}$$

where $\overline{h}$ corresponds to h via $(S, S'FT) = (SF, S'F)$; then there is a unique map $\overline{\overline{h}}$ of the theory $\mathbb{A}$ presented by $\langle S, E, f\rangle$ into the theory $\mathbb{A}'$ presented by $\langle S', E', f'\rangle$ for which

$$\begin{tikzcd}
SF \arrow[r, "\overline{h}"] \arrow[d] & S'F \arrow[d] \\
\mathbb{A} \arrow[r, "\overline{\overline{h}}"'] & \mathbb{A}'
\end{tikzcd}$$

PROOF. From the assumption one gets in $\mathcal{T}$ by adjointness:

$$\begin{tikzcd}
EF \arrow[rr, dashed, "\overline{g}"] \arrow[d, shift left] \arrow[d, shift right] & & E'F \arrow[d, shift left] \arrow[d, shift right] \\
SF \arrow[rr, "\overline{h}"] \arrow[d] & & S'F \arrow[d] \\
\mathbb{A} \arrow[rr, dashed, "\overline{\overline{h}}"'] & & \mathbb{A}'
\end{tikzcd}$$

Because of the existence of $\overline{g}$, $SF \longrightarrow S'F \longrightarrow \mathbb{A}'$ coequalizes $EF \rightrightarrows SF$. Hence the unique $\overline{\overline{h}}$ as required exists. ■

In the terminology we have introduced, the proposition states that a theory map $\mathbb{A} \longrightarrow \mathbb{A}'$ may be defined by assigning, for each n, a polynomial in n variables of $\langle S', E', f' \rangle$ to every basic polynomial in n variables of $\langle S, E, f \rangle$, in such a way that axioms of $\langle S, E, f \rangle$ are mapped into theorems of $\langle S', E', f' \rangle$.

Example. Let $\mathbb{A}'$ be the theory of associative rings with unity and let $\mathbb{A}$ be the theory of Lie rings. 'The' presentation of the latter differs from that of $\mathbb{A}'$ in that there is no $1 \in S_0$, in that $[,]$ replaces $\cdot$ in S_2, and in that antisymmetry replaces identity in E_1, while the Jacobi identity replaces associativity in E_3. Then there is a unique map $\mathbb{A} \longrightarrow \mathbb{A}'$ of the theory of Lie rings into the theory of associative rings with unity defined by interpreting $0, -, +$ as themselves and by interpreting

$$[\pi_0^2, \pi_1^2] \longrightarrow (\pi_0^2 \cdot \pi_1^2) + (-(\pi_1^2 \cdot \pi_0^2)).$$

That the Jacobi identity (and also antisymmetry) are mapped into theorems is shown by the usual calculations.

Chapter III

Algebraic categories

1. Semantics as a coadjoint functor

Definition. *Let $\mathbb{A}$ be an algebraic theory. We say that X* **is a pre-algebra of type** $\mathbb{A}$ *iff X is a functor $\mathbb{A}^* \longrightarrow \mathcal{S}_1$, $\mathbb{A}^*$ being the dual of the small category $\mathbb{A}$. X is called an* **algebra of type** $\mathbb{A}$ *iff X is a pre-algebra of type $\mathbb{A}$ and X commutes with finite products. Denote by $\mathcal{S}_1^{(\mathbb{A}^*)}$ the full subcategory of $\mathcal{S}_1^{\mathbb{A}^*}$ determined by the objects which are algebras. Say that $\mathcal{X}$ is an* **algebraic category** *iff for some algebraic theory $\mathbb{A}$, $\mathcal{X}$ is equivalent to $\mathcal{S}_1^{(\mathbb{A}^*)}$.*

Thus a pre-algebra X of type $\mathbb{A}$ is a sequence of sets together with, for every ${}_nA \xrightarrow{\theta} {}_mA$ in $\mathbb{A}$, a map $X_m \xrightarrow{\theta_X} X_n$ in $\mathcal{S}_1$, such that $(\theta\varphi)_X = \varphi_X\theta_X$. X is an algebra iff $X_n \cong X_1^n$ for all finite sets n. Thus in an algebra X of type $\mathbb{A}$, every n-ary operation θ of $\mathbb{A}$ determines a map $X_1^n \xrightarrow{\theta_X} X_1$ in $\mathcal{S}_1$.

The maps in the category of pre-algebras are natural transformations of functors; if f is such a map, then in particular $\pi_i^n f_1 = f_n\pi_i^n$ for $i \in n$. In an *algebra* the n-ary variables π_i^n are mapped into projections, so that $f_n = f_1^n$. Therefore if X, Y are algebras of type $\mathbb{A}$, then every map $X \xrightarrow{f} Y$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$ is determined by a single map $X_1 \xrightarrow{f_1} Y_1$ in $\mathcal{S}_1$ such that for every n and every n-ary operation θ of $\mathbb{A}$, the diagram

$$\begin{tikzcd} X_1^n \arrow[r, "f_1^n"] \arrow[d, "\theta_X"'] & Y_1^n \arrow[d, "\theta_Y"] \\ X_1 \arrow[r, "f_1"'] & Y_1 \end{tikzcd}$$

is commutative in $\mathcal{S}_1$. Thus it is clear that if we are given a presentation of an algebraic theory $\mathbb{A}$, then $\mathcal{S}_1^{(\mathbb{A}^*)}$ is precisely the usual category of all algebras of the type associated with the presentation ('type' here being used to include specification of identities as well as operations).

We will regard certain facts about algebraic categories as well known, in particular the elementary properties of limits, free algebras, and congruence relations. These needed facts, as well as certain additional statements which are obvious, are recorded in the following three propositions.

Proposition 1. *If $\mathbb{A}$ is any algebraic theory, then the composite functor $U_{\mathbb{A}}$ given by*

$$\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*} \longrightarrow \mathcal{S}_1$$

where the second is evaluation at 1, *is faithful, and has an adjoint, the value of this adjoint at a small set S being the free $\mathbb{A}$-algebra over S. In particular, the above composite is left continuous.*

The faithfulness is clear from the preceding remarks. Actually, we have shown in a corollary to Theorem I.2.5 that the functor $\mathcal{S}_1^{\mathbb{A}^*} \longrightarrow \mathcal{S}_1$ has an adjoint, and we will show in Chapter IV that the inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ has an adjoint, from which it will also follow that $\mathcal{S}_1^{(\mathbb{A}^*)}$ is right complete. (It is obvious from the definition and Proposition I.2.3 that $\mathcal{S}_1^{(\mathbb{A}^*)}$ is left complete.) Explicitly, the left continuity assertion of Proposition 1.1 means that the underlying set of a product of $\mathbb{A}$-algebras or of an equalizer of $\mathbb{A}$-algebra maps is the product or equalizer, respectively, of the underlying sets or $\mathcal{S}_1$-maps.

Proposition 2. *For any algebraic theory $\mathbb{A}$, the functor $\mathbb{A} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ defined by assigning to ${}_nA$ the pre-algebra X such that $X_k = ({}_kA, {}_nA)$, has algebras as values, is full and faithful, and commutes with finite coproducts. Identifying the X just defined with ${}_nA$, we thus have $\mathbb{A} \subset \mathcal{S}_1^{(\mathbb{A}^*)}$, a full subcategory. For any $\mathbb{A}$-algebra Y, $({}_1A, Y) = Y_1$. It follows that ${}_1A$ is a generator for $\mathcal{S}_1^{(\mathbb{A}^*)}$, and that the free algebra over a small set S is the S-fold coproduct $S \cdot {}_1A$. Thus $\mathbb{A}$ is precisely the full category of finitely generated free $\mathbb{A}$-algebras.*

$${}_0A = \lim_{\to \circ} {}^{\mathcal{S}_1^{(\mathbb{A}^*)}}.$$

Proposition 3. *A map of $\mathbb{A}$-algebras is onto iff it is the coequalizer of the congruence relation which it induces on its domain. In other words, $X \xrightarrow{f} Y$ is a regular epimap in $\mathcal{S}_1^{(\mathbb{A}^*)}$ iff for every ${}_1A \xrightarrow{y} Y$ there is ${}_1A \xrightarrow{x} X$ such that $xf = y$.*

Now if $\mathbb{A} \xrightarrow{f} \mathbb{B}$ is a map of algebraic theories, then viewed as a functor $\mathbb{A}^* \xrightarrow{f^*} \mathbb{B}^*$ commutes with finite products. Thus we have a factorization

$$\begin{tikzcd}
\mathcal{S}_1^{(\mathbb{B}^*)} \arrow[r, "\mathcal{S}_1^{(f^*)}"] \arrow[d] & \mathcal{S}_1^{(\mathbb{A}^*)} \arrow[d] \\
\mathcal{S}_1^{\mathbb{B}^*} \arrow[r, "\mathcal{S}_1^{f^*}"'] & \mathcal{S}_1^{\mathbb{A}^*}
\end{tikzcd}$$

That is, $\mathcal{S}_1^{f^*}$ takes algebras into algebras. $\mathcal{S}_1^{(f^*)}$ defined by the above diagram will be called an algebraic functor (of degree 1; algebraic functors of higher degree are defined and discussed in Chapter IV).

We now point out two facts about algebraic categories and algebraic functors.

Proposition 4. *Every algebraic functor of degree* 1 *commutes with the underlying set functors. That is, if* $\mathbb{A} \xrightarrow{f} \mathbb{B}$ *in* $\mathcal{T}$, *then*

$$\begin{tikzcd} \mathcal{S}_1^{(\mathbb{B}^*)} \arrow[rr, "\mathcal{S}_1^{(f^*)}"] \arrow[dr, "U_{\mathbb{B}}"'] & & \mathcal{S}_1^{(\mathbb{A}^*)} \arrow[dl, "U_{\mathbb{A}}"] \\ & \mathcal{S}_1 & \end{tikzcd}$$

is commutative.

Proof. For $X \in \mathcal{S}_1^{(\mathbb{B}^*)}$, $X\mathcal{S}_1^{(f^*)}U_{\mathbb{A}} = fXU_{\mathbb{A}} = (fX)_1 = X_{(1)f} = X_1 = XU_{\mathbb{B}}$. ■

In particular, each $\mathcal{S}_1^{(f^*)}$ is faithful and left continuous.

Theorem 1. *For any algebraic theory* $\mathbb{A}$, *the underlying set functor* $U_{\mathbb{A}}$ *has the property that* $(\{U_{\mathbb{A}}\}^n, \{U_{\mathbb{A}}\}^m)$, *the indicated set of natural transformations in* $\mathcal{S}_1^{\mathcal{S}_1^{(\mathbb{A}^*)}}$, *is small for any finite sets* n, m. *In fact*

$$(\{U_{\mathbb{A}}\}^n, \{U_{\mathbb{A}}\}^m) \cong ({}_mA, {}_nA)$$

where ${}_mA$, ${}_nA$ *are viewed as objects in* $\mathbb{A}$ *(or in* $\mathcal{S}_1^{(\mathbb{A}^*)}$*).*

Proof. The proof of Theorem I.2.5 clearly works equally well if we consider 'large' rather then small categories and properties throughout; in particular the category must have a large coproducts. Thus by a corollary to that theorem we can state the following (which one could actually prove directly without appeal to limits):

Let $\mathcal{B}$ be any large category and let $\mathcal{B} \xrightarrow{T} \mathcal{S}_2$ be any functor. If for any $B \in |\mathcal{B}|$, H^B denotes the functor $\mathcal{B} \longrightarrow \mathcal{S}_2$ whose value at X is (B, X) then

$$(H^B, T) \cong BT.$$

We prove Theorem 1.1 by applying this statement with $\mathcal{B} = \mathcal{S}_1^{\mathcal{S}_1^{(\mathbb{A}^*)}}$. Since $\{U_{\mathbb{A}}\} = H^{{}_1A}$, $(H^B)^n = H^{n \cdot B}$, and $n \cdot {}_1A = {}_nA$, we have for any $\mathcal{S}_1^{(\mathbb{A}^*)} \xrightarrow{T} \mathcal{S}_1$ that $(\{U_{\mathbb{A}}\}^n, T) \cong (H^{{}_nA}, T) \cong {}_nAT$. In the case $T = \{U_{\mathbb{A}}\}^m = H^{{}_mA}$, we therefore have $(\{U_{\mathbb{A}}\}^n, \{U_{\mathbb{A}}\}^m) \cong ({}_mA, {}_nA)$. ■

Definition. *Let $\mathcal{K}$ be the full subcategory of $(\mathcal{C}_2, \{\mathcal{S}_1\})$ determined by those objects $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ for which $(\{U\}^n, U)$ is small for every finite set n. The functor $\mathcal{T}^* \xrightarrow{\mathfrak{S}} \mathcal{K}$ defined by*

$$\begin{aligned} (\mathbb{A})\mathfrak{S} &= U_{\mathbb{A}} \\ (f)\mathfrak{S} &= \mathcal{S}_1^{(f^*)} \end{aligned}$$

will be called (algebraic) **semantics.** *Here, for $\mathbb{A} \xrightarrow{f} \mathbb{B}$ in $\mathcal{T}$, $(f)\mathfrak{S} = \mathcal{S}_1^{(f^*)}$ is regarded as a map $U_{\mathbb{B}} \longrightarrow U_{\mathbb{A}}$ in $\mathcal{K}$, where of course $U_{\mathbb{B}}$ and $U_{\mathbb{A}}$ are the underlying set functors, $\mathcal{S}_1^{(\mathbb{B}^*)} \longrightarrow \mathcal{S}_1$ and $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1$, respectively.*

Thus intuitively the semantics functor assigns to each algebraic theory the category of algebras of which it is a theory. However, we find it necessary to include the underlying set functor $U_{\mathbb{A}}$ as part of the value at $\mathbb{A}$ of $\mathfrak{S}$, because the left continuity of semantics, which follows from Theorem 1.2 below, would be destroyed if we defined $\mathbb{A}\mathfrak{S}$ to be $\mathcal{S}_1^{(\mathbb{A}^*)}$ in $\mathcal{C}_2$ rather than $U_{\mathbb{A}}$ in $\mathcal{K}$.

Note that in $(\mathcal{C}_2, \{\mathcal{S}_1\})$, and hence in $\mathcal{K}$, direct limits agree with those in $\mathcal{C}_2$, but binary products, for example, are defined by meet diagrams of the form

$$\begin{tikzcd}
\mathcal{X} \underset{\mathcal{S}_1}{\times} \mathcal{X} \arrow[r] \arrow[d] \arrow[dr, phantom, "\mathcal{C}_2"] & \mathcal{X}' \arrow[d, "U'"] \\
\mathcal{X} \arrow[r, "U"'] & \mathcal{S}_1
\end{tikzcd}$$

Definition. *We define a functor $\mathcal{K} \xrightarrow{\hat{\mathfrak{S}}} \mathcal{T}^*$, as follows. Given an object $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ in $\mathcal{K}$, we consider the object $\{U\} \in |\mathcal{S}_1^{\mathcal{X}}|$. There is a unique functor $\mathcal{S}_0^* \longrightarrow \mathcal{S}_1^{\mathcal{X}}$ which commutes with finite products and which maps 1 into $\{U\}$. Let $U\hat{\mathfrak{S}}$ be the algebraic theory $\mathcal{S}_0 \longrightarrow \mathbb{A}$, where $\mathbb{A}$ is the dual of the* full *image of $\mathcal{S}_0^* \longrightarrow \mathcal{S}_1^{\mathcal{X}}$. We call $U\hat{\mathfrak{S}}$ the* **algebraic structure** *of U. If $U \xrightarrow{T} U'$ in $\mathcal{K}$, then $U'\hat{\mathfrak{S}} \xrightarrow{T\hat{\mathfrak{S}}} U\hat{\mathfrak{S}}$ in $\mathcal{T}$ is defined by dualizing the small triangle in*

$$\begin{tikzcd}
 & & \mathcal{S}_0^* \arrow[dl] \arrow[dr] & & \\
 & A'^* \arrow[rr, "(T\hat{\mathfrak{S}})^*"] \arrow[dl] & & \mathbb{A}^* \arrow[dr] & \\
\mathcal{S}_1^{\mathcal{X}'} \arrow[rrrr, "\mathcal{S}_1^T"'] & & & & \mathcal{S}_1^{\mathcal{X}}
\end{tikzcd}$$

the big triangle being commutative since $T(U'^n) = (TU')^n = U^n$ *for* $n \in \mathcal{S}_0$.

Explicitly, the n-ary operations of the algebraic structure of $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ are in one-to-one correspondence with the natural transformations $\{U\}^n \longrightarrow \{U\}$ in $\mathcal{S}_1^{\mathcal{X}}$. Theorem 1.1 states that $\mathfrak{S}\hat{\mathfrak{S}} \cong \{\mathcal{T}^*\}$, i.e. $\mathcal{T}^*$ is a retract of $\mathcal{K}$ (up to equivalence).

Theorem 2. *Algebraic structure is adjoint to (algebraic) semantics.*

Proof. We use again Theorem I.2.1, i.e. given $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ in $\mathcal{K}$, we define a 'universal' Φ such that

$$\begin{tikzcd}
\mathcal{X} \arrow[rr, "\Phi"] \arrow[dr, "U"'] & & \mathcal{S}_1^{(U\hat{\mathfrak{S}}*)} \arrow[dl, "U\hat{\mathfrak{S}}\mathfrak{S}"] \\
& \mathcal{S}_1 &
\end{tikzcd}$$

Since in the construction of $\mathbb{A} = U\hat{\mathfrak{S}}$, $\mathbb{A}^*$ is the full image of the functor $\mathcal{S}_0^* \longrightarrow \mathcal{S}_1^{\mathcal{X}}$ defined by taking n-fold products of U with itself for each $n \in \mathcal{S}_0$, there is a functor

$$\mathbb{A}^* \longrightarrow \mathcal{S}_1^{\mathcal{X}}$$

depending on U such that the corresponding

$$\mathcal{X} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$$

has algebras as values. Define Φ as the resulting

$$\mathcal{X} \xrightarrow{\Phi} \mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}.$$

Thus for each $X \in |\mathcal{X}|$, $(X\Phi)_n = (XU)^n$, and for each natural transformation

$$\{U\}^n \xrightarrow{\theta} \{U\}$$

i.e. for each n-ary operation of $\mathbb{A} = U\hat{\mathfrak{S}}$, the corresponding $(X\Phi)_1^n \longrightarrow (X\Phi)_1$ is just θ_X. Thus each map $X \xrightarrow{x} X'$ in $\mathcal{X}$ determines a map $X\Phi \xrightarrow{x\Phi} X'\Phi$ of the corresponding $U\hat{\mathfrak{S}}$-algebras. By construction $\Phi(U\hat{\mathfrak{S}}\mathfrak{S}) = U$. We need to show that Φ has the universal mapping property of Theorem I.2.1.

Consider any algebraic theory $\mathbb{A}'$ and any functor $\mathcal{X} \xrightarrow{T} \mathcal{S}_1^{(\mathbb{A}'^*)}$ for which $TU_{\mathbb{A}'} = U$, i.e. $U \xrightarrow{T} \mathbb{A}'\mathfrak{S}$ in $\mathcal{K}$. We must show that there is a unique $\mathbb{A}' \xrightarrow{f} \mathbb{A} = U\hat{\mathfrak{S}}$ in $\mathcal{T}$ such that $\Phi\mathcal{S}_1^{(f^*)} = T$, i.e. such that $\Phi(f\mathfrak{S}') = T$. Now by Theorem 1.1, $U\hat{\mathfrak{S}}\mathfrak{S}\hat{\mathfrak{S}} \xrightarrow{\Phi\hat{\mathfrak{S}}} U\hat{\mathfrak{S}}$ is an isomorphism in $\mathcal{T}$. Thus if $T = \Phi(f\mathfrak{S})$, then $T\hat{\mathfrak{S}} = (\Phi(f\mathfrak{S}))\hat{\mathfrak{S}} = (f\mathfrak{S}\hat{\mathfrak{S}})\Phi\hat{\mathfrak{S}}$, so that

$$f\mathfrak{S}\hat{\mathfrak{S}} = (T\hat{\mathfrak{S}})(\Phi\hat{\mathfrak{S}})^{-1}.$$

But by Theorem 1.1, the functor $\mathcal{T}^* \xrightarrow{\mathfrak{S}\hat{\mathfrak{S}}} \mathcal{T}^*$ is equivalent to the identity, in particular full and faithful, so that the equation above, and hence the equation $T = \Phi(f\mathfrak{S})$, has exactly one solution f. ∎

According to Theorems 1.1 and 1.2 , given any large category $\mathcal{X}$ and any 'underlying set' functor $\mathcal{X} \xrightarrow{U} \mathcal{S}_1$ such that (U^n, U) is small for all n, there is a well-defined 'algebraic closure' of $\langle \mathcal{X}, U \rangle$, i.e. an algebraic category $\mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$ and functor $\mathcal{X} \xrightarrow{\Phi} \mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$, preserving underlying sets, which is universal with respect to underlying set-preserving functors of $\mathcal{X}$ into any algebraic category. Further, this universality is accomplished by *algebraic* functors in the factorization, and an algebraic category is algebraically closed. In fact, $\mathcal{X}$ is algebraic iff: $\Phi : \mathcal{X} \cong \mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$. We will presently discuss some ways of constructing reasonable underlying set functors U given only the category $\mathcal{X}$. Note however that Theorems 1.1 and 1.2 place no conditions of faithfulness, left continuity, etc. on the functor U considered; the only condition is that (U^n, U) be small.

Example 1. *Let $\mathcal{X}$ be any large category such that $1 \overset{X}{\underset{Y}{\rightrightarrows}} \mathcal{X} \Rightarrow (X, Y)$ small, and such that $\mathcal{X}^* \times \mathcal{X} \xrightarrow{(,)} \mathcal{S}_1$ defines an object in $\mathcal{K}$. Consider the algebraic structure of $(,)$ and let $\overline{\mathcal{X}}$ be the resulting algebraic category. There results*

$$\begin{tikzcd} \mathcal{X}^* \times \mathcal{X} \arrow[rr, "\mathrm{Hom}_{\mathcal{X}}"] \arrow[dr, "{(,)}"'] & & \overline{\mathcal{X}} \arrow[dl] \\ & \mathcal{S}_1 & \end{tikzcd}$$

We call $\mathrm{Hom}_{\mathcal{X}}$ *the* algebraic Hom-functor *of the category $\mathcal{X}$. If, e.g., $\mathcal{X}$ is the category of modules over a ring, then $\overline{\mathcal{X}}$ is the category of modules over the center of the ring.*

Example 2. *Let $\mathcal{D}$ be the category of division rings. $\mathcal{D}$ is not algebraic (e.g. it does not have products); however there is an inclusion $\mathcal{D} \longrightarrow \mathcal{R}$ where $\mathcal{R}$ is the category of associative rings with unity and an obvious underlying set functor U. We thus have*

$$\begin{tikzcd} & \mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)} \arrow[d] \\ \mathcal{D} \arrow[ur] \arrow[r] \arrow[dr, "U"'] & \mathcal{R} \arrow[d] \\ & \mathcal{S}_1 \end{tikzcd}$$

and a map $\mathbb{A} \longrightarrow U\hat{\mathfrak{S}}$ *in* $\mathcal{T}$, *where* $\mathcal{R} = \mathbb{A}\mathfrak{S}$. *However, this map is not an isomorphism; in fact there is an additional unary operation* θ *in* $U\hat{\mathfrak{S}}$, *which satisfies the identities*

$$\begin{aligned} \pi_0^1 \cdot \theta(\pi_0^1) \cdot \pi_0^1 &\equiv \pi_0^1 \\ \theta(\pi_0^2 \cdot \pi_1^2) &\equiv \theta(\pi_1^2) \cdot \theta(\pi_0^2) \\ \theta(1) &\equiv 1. \end{aligned}$$

The maps in the associated category are ring homomorphisms which commute with θ, *and the category does of course have products.*

Example 3. *Let* $\mathcal{G}$ *denote the category of groups,* $\mathcal{R}$, $\mathbb{A}$ *as above. Define* U' *as* $\mathcal{R}_c \times \mathcal{G} \longrightarrow \mathcal{R} \longrightarrow \mathcal{S}_1$, *where the first assigns to* $\langle R, G\rangle$ *the group algebra of* G *with coefficients in the commutative ring* R. *The codomain of* $\mathbb{A} \longrightarrow U'\hat{\mathfrak{S}}$ *has two additional unary operations,* φ, τ, *which in this case satisfy the identities:*

$$\begin{aligned} \varphi\varphi &\equiv \pi_0^1 \\ \varphi(\pi_0^2 + \pi_1^2) &\equiv \varphi(\pi_0^2) + \varphi(\pi_1^2) \\ \varphi(\pi_0^2 \cdot \pi_1^2) &\equiv \varphi(\pi_1^2) \cdot \varphi(\pi_0^2) \\ \varphi(1) &\equiv 1 \\ \tau(1) &\equiv 1 \\ \tau(\pi_0^2 + \pi_1^2) &\equiv \tau(\pi_0^2) + \tau(\pi_1^2) \\ \tau(\pi_0^2 \cdot \pi_1^2) &\equiv \tau(\pi_1^2 \cdot \pi_0^2) \\ \varphi\tau &\equiv \tau \\ \tau\varphi &\equiv \tau \\ \tau\tau &\equiv \tau \end{aligned}$$

If, in this example, we replace $\mathcal{G}$ *by the category of group* mono*morphisms, then evaluation at the neutral element is also part of the resulting algebraic structure.*

Example 4. *Let* $\mathcal{S}_1^*$ *be the dual of the category of small sets and define* $\mathcal{S}_1^* \xrightarrow{P} \mathcal{S}_1$ *by* $SP = (S, 2)$ *for* $S \in \mathcal{S}_1$ $(2 = |\mathfrak{2}|)$. *Then* $P\hat{\mathfrak{S}}$ *is the theory of Boolean algebras, for by the corollary to Theorem I.2.5, we have* $(P^n, P) = ((H^2)^n, H^2) = (H^{2^n}, H^2) = (2^n, 2)$, *(note the two dualizations) and known facts about truth tables complete the proof. The resulting* $\mathcal{S}_1^* \longrightarrow \mathcal{S}_1^{(P\hat{\mathfrak{S}}^*)}$ *takes each small set into its Boolean algebra of subsets.*

Example 5. *Let* $\mathcal{X}$ *be the category of* compact *topological spaces and* I *the unit interval. Consider* $\mathcal{X}^* \xrightarrow{(\cdot, I)} \mathcal{S}_1$. *There results an embedding of* $\mathcal{X}^*$ *in an algebraic category in which the* n*-ary operations are arbitrary continuous* $I^n \xrightarrow{\theta} I$.

2. Characterization of algebraic categories

Definition. *Let $\mathcal{X}$ be any large category having finite limits and let $G \in |\mathcal{X}|$ be such that for any small set S, $S \cdot G$, the S-fold coproduct of G with itself, exists in $\mathcal{X}$. Then G is a* **generator** *for $\mathcal{X}$ iff*

$$\forall X\ \forall Y\ \forall f\ \forall g\ [X \underset{g}{\overset{f}{\rightrightarrows}} Y \Rightarrow [f = g \Leftrightarrow \forall x\ [G \xrightarrow{x} X \Rightarrow xf = xg]]].$$

G is **projective** *iff*

$$\begin{gathered}\forall X\ \forall Y\ \forall f\ [X \xrightarrow{f} Y \Rightarrow [\exists h\ \exists g\ [f = hE^* g] \\ \Leftrightarrow \forall y\ [G \xrightarrow{y} Y \Rightarrow \exists x\ [G \xrightarrow{x} X \wedge xf = y]]]].\end{gathered}$$

G is **abstractly finite** *iff*

$$\forall S\ \forall f\ [S \textit{ is small set } \wedge G \xrightarrow{f} S \cdot G \Rightarrow \exists F\ \exists g\ \exists h\ [F \textit{ is a finite set } \wedge F \xrightarrow{g} S \wedge$$

$$\begin{tikzcd} G \arrow[rr, "f"] \arrow[dr, "h"'] & & S \cdot G \\ & F \cdot G \arrow[ur, "g \cdot G"'] & \end{tikzcd} \quad]].$$

The term 'abstractly finite' is due to [Freyd, 1960]. Note that our concept of 'projective' coincides with the usual one for abelian categories, but differs for some other categories. The definition of G being projective may be rephrased: 'For every f, f is 'onto' (with respect to G) *iff* f is a *regular* epimap.' If $\mathbb{A}$ is any algebraic theory, then ${}_1A$ is an abstractly finite projective generator for $\mathcal{S}_1^{(\mathbb{A}^*)}$. The abstract finiteness is due to the nature of coproducts in an algebraic category: the value at π_0^1 of a map ${}_1A \longrightarrow S \cdot {}_1A$ is some n-ary operation θ applied to the various copies of π_0^1; but this, and hence the whole image of the map, involves at most n of the S-copies of ${}_1A$.

Definition. *Let $\mathcal{X}$ be a category and $G \in |\mathcal{X}|$. Consider a pair*

$$X \underset{f_1}{\overset{f_0}{\rightrightarrows}} Y$$

of maps in $\mathcal{X}$. We say that $\langle f_0, f_1 \rangle$ is **RST with respect to** *G iff the following three conditions hold (in $\mathcal{X}$):*

R: $\forall y\ [G \xrightarrow{y} Y \Rightarrow \exists x\ [G \xrightarrow{x} X \wedge xf_0 = y \wedge xf_1 = y]]$

S: $\forall x\ [G \xrightarrow{x} X \Rightarrow \exists x'\ [G \xrightarrow{x'} X\ x'f_0 = xf_1 \wedge x'f_1 = xf_0]]$

T: $\forall u\, \forall v\ [G \xrightarrow{u} X \wedge G \xrightarrow{v} X \wedge uf_1 = vf_0 \Rightarrow \exists w\ [G \xrightarrow{w} X \wedge wf_0 = uf_0 \wedge wf_1 = vf_1]]$.

We say that $\langle f_0, f_1\rangle$ *is* **monomorphic with respect to** G *iff*

$$\forall x\, \forall x'\ [G \xrightarrow{x} X \wedge G \xrightarrow{x'} X \wedge xf_0 = x'f_0 \wedge xf_1 = x'f_1 \Rightarrow x = x'].$$

We say that $\langle f_0, f_1\rangle$ *is a* **congruence relation with respect to** G *iff*

$$\forall y\, \forall y'\ [G \xrightarrow{y} Y \wedge G \xrightarrow{y'} Y \wedge [\forall Z \forall z[Y \xrightarrow{z} Z \wedge f_0 z = f_1 z \Rightarrow yz = y'z]]$$
$$\Rightarrow \exists! x\ [G \xrightarrow{x} X \wedge xf_0 = y \wedge xf_1 = y']].$$

It is well known that (in our language) if $\mathbb{A}$ is an algebraic theory, $\mathcal{X} = \mathcal{S}_1^{(\mathbb{A}^*)}$, $G = {}_1A$, then a pair $\langle f_0, f_1\rangle$ in $\mathcal{X}$ is a congruence relation with respect to ${}_1A$ iff it is monomorphic and RST with respect to ${}_1A$. In an algebraic category, a pair $\langle f_0, f_1\rangle$ is a congruence relation with respect to ${}_1A$ iff the map $X \xrightarrow{\langle f_0, f_1\rangle} Y \times Y$ is the equalizer of $Y \times Y \underset{p_2}{\overset{p_1}{\rightrightarrows}} Y \xrightarrow{q} Q$ where q is the coequalizer of the pair $\langle f_0, f_1\rangle$.

Theorem 1. *Let $\mathcal{X}$ be any large category with finite limits. Then $\mathcal{X}$ is algebraic iff there exists $G \in |\mathcal{X}|$ such that*

(1) arbitrary small coproducts of G with itself exist in $\mathcal{X}$;

(2) for every $X \in |\mathcal{X}|$, (G, X) is small;

(3) G is an abstractly finite projective generator for $\mathcal{X}$;

(4) for any small set I and any object X, any pair $\langle f_0, f_1\rangle$ of maps $X \underset{f_1}{\overset{f_0}{\rightrightarrows}} I \cdot G$ which is monomorphic and RST with respect to G is also a congruence relation with respect to G.

Proof. Necessity is clear. Suppose there is an object G satisfying the four conditions, let $U = (G, \cdot)$ and consider the functor $\mathcal{X} \xrightarrow{\Phi} \mathcal{S}_1^{(U\hat{\mathfrak{G}}^*)}$. We must show that Φ is full, dense, and faithful. Since G is a generator, faithfulness is clear.

We now show that Φ is full. Note that for any $X \in |\mathcal{X}|$, (G, X) is the underlying set of $X\Phi$, and that a map $X\Phi \longrightarrow Y\Phi$ in $\mathcal{S}_1^{(U\hat{\mathfrak{G}}^*)}$ may be identified as a map $(G, X) \xrightarrow{f} (G, Y)$ in $\mathcal{S}_1$ which commutes with every n-ary operation $G \xrightarrow{\theta} n \cdot G$ from $\mathcal{X}$. Suppose given such an f. We need to show that $f = (G, \varphi)$ for some $X \xrightarrow{\varphi} Y$ in $\mathcal{X}$. Now there is in $\mathcal{X}$

$$R \underset{\beta}{\overset{\alpha}{\rightrightarrows}} I \cdot G \xrightarrow{p} X$$

such that $p = \alpha E^* \beta$. (For example, we may take $I = (G, X)$ and note that the obvious p is regular since it is onto with respect to G.) Let, for each $i \in I$, e_i denote the i-th injection $G \longrightarrow I \cdot G$. We first define $G \xrightarrow{\varphi_i} Y$, $i \in I$ by

$$\varphi_i = (e_i p) f.$$

Now since $I \cdot G$ is a coproduct, there is a unique $I \cdot G \xrightarrow{\overline{\varphi}} Y$ such that $e_i\overline{\varphi} = \varphi_i$ for $i \in I$. We need to show that $\overline{\varphi}$ coequalizes α, β. Consider any $G \xrightarrow{r} R$. Since G is abstractly finite, there is a finite set n, a map $n \xrightarrow{k} I$, and maps α', β' such that $r\alpha'(k \cdot G) = r\alpha$ and $r\beta'(k \cdot G) = r\beta$.

$$\begin{tikzcd}
G \arrow[r, "r"] & R \arrow[r, shift left, "\alpha'"] \arrow[r, shift right, "\beta'"'] & n \cdot G \arrow[r, "k \cdot G"] \arrow[dr, "((k \cdot G)p)f^n"'] & I \cdot G \arrow[r, "p"] \arrow[d, "\overline{\varphi}"] & X \\
& & & Y &
\end{tikzcd}$$

We show that the triangle is commutative. For each $G \xrightarrow{\pi_1^n} n \cdot G$, $i \in n$, we have

$$\pi_i^n(k \cdot G)\overline{\varphi} = e_{(i)k}\overline{\varphi} = \varphi_{(i)k} = (e_{(i)k}p)f = (\pi_i^n(k \cdot G)p)f = \pi_i^n((k \cdot G)p)f^n,$$

the last since f commutes with the n-ary operation π_i^n. Thus $(k \cdot G)\overline{\varphi} = ((k \cdot G)p)f^n$. Then

$$\begin{aligned}
r\alpha\overline{\varphi} &= r\alpha'(k \cdot G)\overline{\varphi} = r\alpha'((k \cdot G)p)f^n = (r\alpha'(k \cdot G)p)f \\
&= (r\alpha p)f = (r\beta p)f = (r\beta'(k \cdot G)p)f \\
&= r\beta'((k \cdot G)p)f^n = r\beta'(k \cdot G)\overline{\varphi} = r\beta\overline{\varphi}
\end{aligned}$$

since $\alpha p = \beta p$ and f commutes with the n-ary operations $r\alpha'$ and $r\beta'$. Since the foregoing holds for every $G \xrightarrow{r} R$ we have that $\alpha\overline{\varphi} = \beta\overline{\varphi}$, and hence $\exists! X \xrightarrow{\varphi} Y$ such that $p\varphi = \overline{\varphi}$. It remains to show that $(G, \varphi) = f$. But since p is onto with respect to G, for every $G \xrightarrow{x} X$ there is $G \xrightarrow{\overline{x}} I \cdot G$ such that $\overline{x}p = x$.

$$\begin{tikzcd}
n \cdot G \arrow[dr, "h \cdot G"'] & & G \arrow[ll, "\theta"'] \arrow[dl, "\overline{x}"'] \arrow[dr, "x"] & \\
& I \cdot G \arrow[rr, "p"] \arrow[d, "\overline{\varphi}"'] & & X \arrow[dll, "\varphi"] \\
& Y & &
\end{tikzcd}$$

By abstract finiteness there is a finite set n, an n-ary operation θ, and $n \xrightarrow{h} I$ such that $\theta(h \cdot G) = \overline{x}$. Then $x\varphi = \theta(h \cdot G)p\varphi = \theta(h \cdot G)\overline{\varphi}$. On the other hand $(x)f = (\theta(h \cdot G)p)f = \theta((h \cdot G)p)f^n$ since f commutes with n-ary operation θ. Thus to show $x\varphi = (x)f$ reduces to showing $(h \cdot G)p\varphi = ((h \cdot G)p)f^n$. For each $i \in n$, $\pi_i^n(h \cdot G)p\varphi = e_{(i)h}\overline{\varphi} = (e_{(i)h}p)f =$

$(\pi_i^n(h \cdot G)p)f = \pi_i^n((h \cdot G)p)f^n$. Hence $x\varphi = (x)f$. But this is true for all $G \xrightarrow{x} X$, i.e. $f = (G, \varphi)$. Thus Φ is full.

We now must show that Φ is dense. Denoting ${}_1A = (G)\Phi$, we first show that

$$(I \cdot G)\Phi \cong I \cdot {}_1A$$

in $\mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$ for any small set I.

Since there is in any case a $U\hat{\mathfrak{S}}$-map $I \cdot {}_1A \xrightarrow{\lambda} (\ I\ \cdot\ G)\ \Phi$ it suffices to show that λ is one-to-one and onto by Proposition I.3.2; that is, we show that λ induces

$$({}_1A, I \cdot {}_1A) \cong (G, I \cdot G) = ({}_1A, (I \cdot G)\Phi).$$

Now if $I = n$, a finite set, this relation is true since both sides are just the set of n-ary operations of the theory $U\hat{\mathfrak{S}}$. This fact and abstract finiteness enable us to construct an inverse μ to $({}_1A, \lambda)$. For each finite n and $n \longrightarrow I$ we have

$$\begin{tikzcd}
({}_1A, n \cdot {}_1A) \arrow[r, shift left, "\lambda_n"] \arrow[d] & (G, n \cdot G) \arrow[l, shift left, "\mu_n"] \arrow[d] \\
({}_1A, I \cdot {}_1A) \arrow[r, "({}_1A, \lambda)"'] & (G, I \cdot G)
\end{tikzcd} \qquad \mu_n = \lambda_n^{-1}$$

Since $(G, I \cdot G) = \varinjlim(G, n \cdot G)$, the maps $(G, n \cdot G) \longrightarrow ({}_1A, I \cdot {}_1A)$ yield a unique μ, inverse to $({}_1A, \lambda)$.

Now assume that Y is an arbitrary $U\hat{\mathfrak{S}}$-algebra. Then there is a small set I and a regular epimap $I \cdot {}_1A \xrightarrow{p} Y$. The remaining maps in the following diagrams are constructed as described below.

$$\mathcal{X}: \quad \begin{tikzcd}
S \cdot G \arrow[r, shift left, "a"] \arrow[r, shift right, "b"'] \arrow[dr, dotted, "\overline{r}"'] & I \cdot G \arrow[rr, "q"] & & \overline{Y} \\
& K_q \arrow[u, shift left, "\overline{\alpha}"] \arrow[u, shift right, "\overline{\beta}"'] & &
\end{tikzcd}$$

$$\mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}: \quad \begin{tikzcd}
S \cdot {}_1A \arrow[r, "r"] & R \arrow[r, shift left, "\alpha"] \arrow[r, shift right, "\beta"'] & I \cdot {}_1A \arrow[r, "p"] & Y.
\end{tikzcd}$$

First, $\langle \alpha, \beta \rangle$ is the kernel of p, i.e. $\alpha = kp_1$, $\beta = kp_2$ where $k = (p_1p)E(p_2p)$, where $(I \cdot {}_1A) \times (I \cdot {}_1A) \xrightarrow{p_j} I \cdot {}_1A$ are the two projections. Second, there is a regular epimap r from some chosen $S \cdot {}_1A$ to R. Because Φ preserves coproducts of G with itself, and because Φ is full, there are a, b in $\mathcal{X}$ such that $a\Phi = r\alpha$, $b\Phi = r\beta$. Define $q = aE^*b$ in $\mathcal{X}$. We wish to show that $\overline{Y}\Phi \cong Y$ where $\overline{Y}$ is the codomain of q. Let $\langle\overline{\alpha},\overline{\beta}\rangle$ be the kernel of q. Because $\langle a,b\rangle q_1 q = \langle a,b\rangle q_2 q$, where $(I\cdot G)\times(I\cdot G) \xrightarrow{q_j} I\cdot G$ are the projections, there is a unique $S\cdot G \xrightarrow{\overline{r}} K_q$ in $\mathcal{X}$ for which $\overline{r}\bar{\alpha} = a$, $\overline{r}\overline{\beta} = b$. Since $aq = bq$ in $\mathcal{X}$, we have $r\alpha(q\Phi) = (a\Phi)(q\Phi) = (aq)\Phi = (bq)\Phi = (b\Phi)(q\Phi) = r\beta(q\Phi)$. Since r is an epimap, $\alpha(q\Phi) = \beta(q\Phi)$, and there is a unique $Y \xrightarrow{\varphi} \overline{Y}\Phi$ such that $p\varphi = q\Phi$. Because Φ is left exact, $\langle\overline{\alpha}\Phi,\overline{\beta}\Phi\rangle$ is the kernel of $q\Phi$ in $\mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$. Hence there is a unique $R \xrightarrow{\psi} K_q\Phi$ such that $\alpha = \psi(\overline{\alpha}\Phi)$ and $\beta = \psi(\overline{\beta}\Phi)$. Then, because $\langle\overline{\alpha},\overline{\beta}\rangle\Phi$ is a monomorphism, $\overline{r}\Phi = r\psi$, and moreover ψ is a regular monomap since $\langle\alpha,\beta\rangle$ is. Since r is a regular epimap, there are $S'\cdot{}_1A \rightrightarrows S\cdot{}_1A$ with coequalizer r. Let $S\cdot G \xrightarrow{\overline{\overline{r}}} Q$ be the coequalizer in $\mathcal{X}$ of the corresponding $S'\cdot G \rightrightarrows S\cdot G$. The map $\overline{r}$ also coequalizes the last pair since $\langle\overline{\alpha},\overline{\beta}\rangle$ is a monomorphism. Thus there is a unique $\eta \in \mathcal{X}$ such that

$$\begin{tikzcd}
S\cdot G \arrow[rr, "\langle a,b\rangle"] \arrow[d, "\overline{\overline{r}}"'] \arrow[dr, "\overline{r}"] & & (I\cdot G)^2 \\
Q \arrow[r, "\eta"'] & K_q \arrow[ur, "\langle\overline{\alpha},\overline{\beta}\rangle"'] &
\end{tikzcd}$$

is commutative. Since $\overline{\overline{r}}\Phi$ coequalizes $S'\cdot{}_1A \rightrightarrows S\cdot{}_1A$, there is a unique $\xi \in \mathcal{S}_1^{(U\hat{\mathfrak{S}}^*)}$ such that $r\xi = \overline{\overline{r}}\Phi$. Then also $\xi(\eta\Phi) = \psi$ since r is an epimorphism.

$$\begin{tikzcd}
S\cdot{}_1A \arrow[rr, "r"] \arrow[dr, "\overline{\overline{r}}\Phi"] \arrow[ddr, "\overline{r}\Phi"'] & & R \arrow[rr, "\langle\alpha,\beta\rangle"] \arrow[dl, "\xi"'] \arrow[ddl, "\psi"] & & (I\cdot{}_1A)^2 \\
& Q\Phi \arrow[d, "\eta\Phi"] & & & \\
& K_q\Phi \arrow[uurrr, "\langle\overline{\alpha},\overline{\beta}\rangle\Phi"'] & & &
\end{tikzcd}$$

Thus ξ is a monomorphism since ψ is. Now $\overline{\overline{r}}\Phi$ is a regular epimap because $\overline{\overline{r}}$ is and because Φ takes G-onto maps into ${}_1A$-onto maps, and by assumption this is equivalent to taking regular epimaps into regular epimaps. Since r is an epimorphism, ξ is therefore a regular epimap and hence an isomorphism (by Proposition I.3.3 and I.3.1), and without loss we may assume $Q\Phi = \xi = R$, so that in particular $\psi = \eta\Phi$. Because Φ is faithful and ψ is a monomorphism, it follows that η is a monomorphism in $\mathcal{X}$. Now q is also the coequalizer of $\langle\eta\overline{\alpha},\eta\overline{\beta}\rangle$. Thus by assumption (4), in order to show that $Q \cong K_q$, it will be enough to show that $\langle \eta\overline{\alpha}, \eta\overline{\beta}\rangle$ is RST with respect to G. But this is clear since $\langle \alpha, \beta\rangle$ is RST with respect to ${}_1A$ and $\alpha = (\eta\overline{\alpha})\Phi$, $\beta = (\eta\overline{\beta})\Phi$, while Φ is full and faithful. Therefore $\eta : Q \xrightarrow{\approx} K$, so that $\psi = \eta\Phi$ establishes an isomorphism $R \xrightarrow{\approx} K_q\Phi$ in $\mathcal{S}_1^{(U\hat{\mathfrak{G}}^*)}$, and hence $Y = I \cdot {}_1A/R \cong I \cdot {}_1A/K_q\Phi = \overline{Y}\Phi$, the last since $q\Phi$ is a regular epimap with kernel $K_q\Phi$. Thus Φ is dense. ∎

Remark. It will be noted that even in the absence of assumption (4), the correspondence $Y \longrightarrow \overline{Y}$ constructed in the foregoing proof provides an adjoint to Φ. Assumption (4) is needed, for the full category $\mathcal{X}$ of torsion-free abelian groups is complete and has an abstractly finite projective generator Z, yet it is not algebraic since the algebraic structure of $(Z,\ )$ is the theory of abelian groups, and hence the functor Φ has the category $\mathcal{A}_1$ of all abelian groups for its codomain. The adjoint in this case consists of dividing by the torsion part.

Sums (coproducts), as well as products and equalizers, in $\mathcal{X}$ coincide with those in $\mathcal{A}_1$. However, the coequalizer of a pair f_0, f_1 is obtained in $\mathcal{X}$ by first taking the coequalizer in $\mathcal{A}_1$ and then dividing by the torsion part:

$$X \underset{f_1}{\overset{f_0}{\rightrightarrows}} Y \xrightarrow{q} K^* \xrightarrow{\eta_{K^*}} K^*/K^*T$$

$$q = f_0 E^*_{\mathcal{A}_1} f_1\ ,\ q\eta_{K^*} = f_0 E^*_{\mathcal{X}} f_1.$$

This makes it clear that regular epimaps are precisely onto maps, so that Z is projective in $\mathcal{X}$. An explicit counter-example to assumption (4) is obtained by choosing $p > 1$ and considering

$$Z \times Z \underset{f_1}{\overset{f_0}{\rightrightarrows}} Z$$

where $\langle x, y\rangle f_0 = x$, $\langle x, y\rangle f_1 = x + py$. Since the $\mathcal{A}_1$-coequalizer is the torsion group Z_p, the $\mathcal{X}$-coequalizer is 0, so that the kernel of the coequalizer is not $\langle f_0, f_1\rangle$ even though $\langle f_0, f_1\rangle$ is monomorphic and RST.

Corollary. *An abelian algebraic category is equivalent to the category of modules over some associative ring with unity.*

In fact, [Freyd, 1960] has established that any complete abelian category with an abstractly finite projective generator G is equivalent to the category of modules over the endomorphism ring of G.

Corollary. *Suppose $\langle S, E, f\rangle$ is a presentation of an algebraic theory $\mathbb{A}$, and suppose $\mathcal{X}$ is a large category with finite limits. Then $\mathcal{X}$ is equivalent to the category of all $\mathbb{A}$-algebras iff there is an object G in $\mathcal{X}$ satisfying conditions (1)-(4) of the theorem and a sequence of maps of sets $S_n \xrightarrow{q_n} (G, n\cdot G)$ such that every map $n\cdot G \longrightarrow m\cdot G$ in $\mathcal{X}$ can be expressed by composition and $\{\ \}$ in terms of $\mathcal{S}_0$ and S, and such that maps $n\cdot G \rightrightarrows m\cdot G$ are equal if and only if their being so follows from E, equations holding in $\mathcal{S}_0$, and the fact that $n\cdot G$ is an n-fold coproduct,*

Theorem 2. *If $\mathcal{X}$ is any algebraic category and $\mathbb{C}$ is small with $|\mathbb{C}| < \infty$, and if either all $(C,B) \neq 0$ or $\mathcal{X}$ has constant operations, $\mathcal{X}^{\mathbb{C}}$ is also an algebraic category.*

Proof. $\mathcal{X}^{\mathbb{C}}$ is complete, and since limits and monomorphisms are pointwise-characterized in $\mathcal{X}^{\mathbb{C}}$, it follows that every RST monomorphic pair in $\mathcal{X}^{\mathbb{C}}$ is a congruence relation. (Note that RST, monomorphism and congruence are actually intrinsic concepts, though we defined them in term of a given projective generator.) Thus we need only show that $\mathcal{X}^{\mathbb{C}}$ has an abstractly finite projective generator. We do this in such a way that the underlying set of a functor $A : \mathbb{C} \longrightarrow \mathcal{X}$ turns out to be the product $\prod_{C \in |\mathbb{C}|} CAU$, where $U = (G,\ )$ is the underlying set functor determined by some chosen abstractly finite projective generator G for $\mathcal{X}$. Consider the composite functor P

$$\mathcal{X}^{\mathbb{C}} \longrightarrow \mathcal{X}^{|\mathbb{C}|} \xrightarrow{\Pi} \mathcal{X}.$$

By Theorem I.2.5, the first functor has an adjoint, and of course Π has the adjoint $\mathcal{X} \longrightarrow \mathcal{X}^{|\mathbb{C}|}$ induced by $|\mathbb{C}| \longrightarrow \mathfrak{1}$. Thus P has an adjoint, so that in particular there is an object $\overline{G} \in |\mathcal{X}^{\mathbb{C}}|$ such that

$$(\overline{G}, A) \cong (G, AP)$$

for every object $A \in |\mathcal{X}^{\mathbb{C}}|$. Since P is clearly faithful, $\overline{G}$ is a generator.

Since limits are calculated pointwise in $\mathcal{X}^{\mathbb{C}}$, a map $\varphi \in \mathcal{X}^{\mathbb{C}}$ is a regular epimap iff each φ_C, $C \in |\mathbb{C}|$ is a regular epimap in $\mathcal{X}$. Thus to show that $\overline{G}$ is a projective we need to show that φ is $\overline{G}$-onto iff each φ_C is G-onto. If $A \xrightarrow{\varphi} B$ in $\mathcal{X}^{\mathbb{C}}$, we have in $\mathcal{S}_1$

$$\begin{tikzcd}
(\overline{G}, A) \arrow[r, "{(\overline{G}, \varphi)}"] \arrow[d, equal] & (\overline{G}, B) \arrow[d, equal] \\
(G, AP) \arrow[r, "{(G, \varphi P)}"] \arrow[d, equal] & (G, BP) \arrow[d, equal] \\
\prod_{C \in |\mathbb{C}|} (G, CA) \arrow[r, "{\prod_{C \in |\mathbb{C}|} (G, \varphi_C)}"] & \prod_{C \in |\mathbb{C}|} (G, CB)
\end{tikzcd}$$

But in $\mathcal{S}_1$, $\prod_{C \in |\mathbb{C}|} (G, \varphi_C)$ is onto iff each (G, φ_C) is onto. Thus the $\mathcal{X}^{\mathbb{C}}$-regular epimaps are precisely the $\overline{G}$-onto maps, i.e. $\overline{G}$ is projective.

We now show that $\overline{G}$ is abstractly finite if and only if $|\mathbb{C}|$ is finite. For this we need to recall from Theorem I.2.5 the formula for $\overline{G}$:

$$(C)\overline{G} = \varinjlim_{(i,C)}{}^{\mathcal{X}} (d_0^C, \tilde{G})$$

where $|\mathbb{C}| \xrightarrow{i} \mathbb{C}$ is the inclusion and where $\tilde{G}$ is the functor $|\mathbb{C}| \longrightarrow \mathcal{X}$ constantly equal to G. Since (i, C) in this case is a *set* (namely the set of all maps in $\mathbb{C}$ with codomain C) and since the functor $d_0^C \tilde{G}$ is constant, we have

$$(C)\overline{G} = (i, C) \cdot G$$

the (i, C)-fold coproduct of G with itself in $\mathcal{X}$. Now consider any small set I and any map $\overline{G} \xrightarrow{f} I \cdot \overline{G}$ in $\mathcal{X}^{\mathbb{C}}$. We have

$$(\overline{G}, I \cdot \overline{G}) \cong (G, (I \cdot \overline{G})P) \cong \prod_{C \in |\mathbb{C}|} (G, (I \times (i, C)) \cdot G).$$

Now since G is abstractly finite, each of the maps $\overline{f}_C : G \longrightarrow (I \times (i, C)) \cdot G$ corresponding to f factors through a sub-coproduct corresponding to a *finite* $I_C \subset I \times (i, C)$. If $\overline{G}$ is to be abstractly finite, then there must be a finite $J \subset I$ such that $I_C \subset J \times (i, C)$ for all C, because the I_C can be arbitrary. If $|\mathbb{C}|$ is not finite, then the family $I_C = \{\langle C, C \rangle\}$ of singletons admits no such J, showing the necessity. For the sufficiency, suppose $|\mathbb{C}|$ is finite. Then

$$J = \{j \mid \exists C \in |\mathbb{C}| \; \exists u \in (i, C) \; [\langle j, u \rangle \in I_C]\}$$

is a finite subset of I such that f factors

$$\begin{tikzcd} \overline{G} \arrow[rr, "f"] \arrow[dr] & & I \cdot \overline{G} \\ & J \cdot \overline{G} \arrow[ur] & \end{tikzcd}$$

■

Remark. It will be noted that if $\mathbb{C}$ is a monoid, i.e. $|\mathbb{C}| = 1$, then the underlying set functor constructed for $\mathcal{X}^{\mathbb{C}}$ in the above proof matches that of $\mathcal{X}$, i.e.

$$\begin{tikzcd} \mathcal{X}^{\mathbb{C}} \arrow[rr] \arrow[dr] & & \mathcal{X}^{|\mathbb{C}|} = \mathcal{X} \arrow[dl] \\ & \mathcal{S}_1 & \end{tikzcd}$$

is commutative, so that the connecting functor is algebraic of degree one. In Chapter V, we study this case in more detail and in particular determine the algebraic structure of $\mathcal{X}^{\mathbb{C}} \xrightarrow{P} \mathcal{X} \longrightarrow \mathcal{S}_1$ in terms of $\mathbb{C}$ and the algebraic structure of $\mathcal{X} \longrightarrow \mathcal{S}_1$.

In case $\mathcal{X}$ is the category of modules over a ring R, $\mathbb{C}$ arbitrary, I have constructed elsewhere [Lawvere, 1963] a ring $R[\mathbb{C}]$ such that, in case $|\mathbb{C}|$ is finite, the algebraic structure of $\mathcal{X}^{\mathbb{C}} \longrightarrow \mathcal{S}_1$ turns out to be the theory of modules over $R[\mathbb{C}]$. Of course $R[\mathbb{C}]$ reduces to the usual monoid ring in case $|\mathbb{C}| = 1$.

Remark. The generator G of Theorem 2.1 is not uniquely determined by $\mathcal{X}$, i.e. $\mathcal{X}$ may be represented as $\mathcal{S}_1^{(\mathbb{A}^*)}$ for many different non-isomorphic theories $\mathbb{A}$. For example, if $\mathbb{C}$ of Theorem 2.2 is an **equivalence relation**, i.e. for any two $C, C' \in |\mathbb{C}|$ there is exactly one $\mathbb{C} \longrightarrow C'$ in $\mathbb{C}$, then $\mathbb{C} \longrightarrow \mathbb{1}$ is an equivalence, so that

$$\mathcal{X}^{\mathbb{C}} \longrightarrow \mathcal{X}$$

is also an equivalence. However if $|\mathbb{C}| > 1$, then the theory constructed for $\mathcal{X}^{\mathbb{C}}$ is different from the given one for $\mathcal{X}$. In particular, if $\mathcal{X}$ is the category of R-modules and $|\mathbb{C}| = n$, then $R[\mathbb{C}]$ is the ring of $n \times n$ matrices over R. Of course, an equivalence which commutes with underlying set functors does induce an isomorphism of theories.

Chapter IV

Algebraic functors

1. The algebra engendered by a prealgebra

Let $\mathbb{A}$ be an algebraic theory. Recall that a prealgebra X of type $\mathbb{A}$ is any functor $\mathbb{A}^* \xrightarrow{X} \mathcal{S}_1$, and that an algebra of type $\mathbb{A}$ is a prealgebra X of type $\mathbb{A}$ such that $X_n = X_1^n$ for every object n in $\mathbb{A}$.

Theorem 1. *If $\mathbb{A}$ is any algebraic theory, then the inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ of algebras into prealgebras admits an adjoint.*

Proof. Let X be any prealgebra of type $\mathbb{A}$. In the free $\mathbb{A}$-algebra X_1F generated by the set X_1, consider the smallest $\mathbb{A}$-congruence XE containing the following relations:

If $n \in |\mathcal{S}_0|$, if ${}_1A \xrightarrow{\theta} {}_nA$ in $\mathbb{A}$, and if there is $y \in X_n$ such that

$$y\theta_X = x \text{ and } y{\pi_i^n}_X = x_i \text{ for } i \in n$$

then

$$(x_0\kappa, \dots, x_{n-1}\kappa)\theta \underset{XE}{\equiv} x\kappa$$

where $X_1 \xrightarrow{\kappa} X_1FU_{\mathbb{A}}$ is the canonical inclusion. Define a prealgebra $\overline{X}$ as follows. For each $n \in |\mathcal{S}_0|$,

$$(\overline{X}_n) = (X_1F/XE)^n U_{\mathbb{A}}.$$

For each $1 \xrightarrow{i} n$ in $\mathcal{S}_0$, $\overline{X}$ maps i into the i-th projection

$$(\overline{X})_n = (\overline{X})_1^n \xrightarrow{\pi_i^n} \overline{X}_1.$$

For any ${}_1A \xrightarrow{\theta} {}_nA$ in $\mathbb{A}$, $\theta_{\overline{X}}$ is the induced operation on the quotient:

$$\begin{tikzcd}
(X_1F)^nU_{\mathbb{A}} \arrow[r, "\theta_{X_1F}"] \arrow[d, "\eta^n"'] & X_1FU_{\mathbb{A}} \arrow[d, "\eta"] \\
\overline{X}_n = (\overline{X}_1F/XE)^nU_{\mathbb{A}} \arrow[r, "\theta_{\overline{X}}"'] & (X_1F/XE)U_{\mathbb{A}} = \overline{X}_1
\end{tikzcd}$$

where η is the quotient map. It is clear that $\overline{X}$ is an algebra of type $\mathbb{A}$.

We define a map $X \xrightarrow{\varphi} \overline{X}$ as follows. For each $n \in |\mathcal{S}_0| = |\mathbb{A}|$, φ_n is defined by the requirement that

$$\begin{tikzcd}
X_n \arrow[rr, "\varphi_n"] \arrow[d, "\pi^n_{i\,X}"'] & & \overline{X}_n = (X_1F/XE)^nU_{\mathbb{A}} \arrow[d, "\pi^n_{i\,\overline{X}}"] \\
X_1 \arrow[r, "\kappa"'] & X_1FU_{\mathbb{A}} \arrow[r, "\eta"'] & \overline{X}_1 = (X_1F/XE)U_{\mathbb{A}}
\end{tikzcd}$$

is commutative for each $i \in n$, since $\pi^n_{i\,\overline{X}}$ are projections in $\mathcal{S}_1$. We prove that φ is a natural transformation, i.e. a map of prealgebras. Consider any ${}_1A \xrightarrow{\theta} {}_nA$ in $\mathbb{A}$, and let $y \in X_n$. By the definition above we have

$$y\varphi_n\theta_{\overline{X}} = \langle y\pi^n_{0\,X}\kappa\eta, \ldots, y\pi^n_{n-1\,X}\kappa\eta\rangle\theta_{\overline{X}} = y\theta_X\kappa\eta = y\theta_X\varphi_1.$$

Thus for every ${}_nA \in |\mathbb{A}|$ and every ${}_1A \xrightarrow{\theta} {}_nA$, we have that

$$\begin{tikzcd}
X_n \arrow[r, "\varphi_n"] \arrow[d, "\theta_X"'] & \overline{X}_n \arrow[d, "\theta_{\overline{X}}"] \\
X_1 \arrow[r, "\varphi_1"'] & \overline{X}_1
\end{tikzcd}$$

is commutative.

It follows that $\theta_X\varphi_m = \varphi_n\theta_{\overline{X}}$ for any ${}_mA \xrightarrow{\theta} {}_nA$ in $\mathbb{A}$ since $\overline{X}_m$ is a product. Thus $X \xrightarrow{\varphi} \overline{X}$ in $\mathcal{S}_1^{\mathbb{A}^*}$.

We now show that φ is universal. Suppose $X \xrightarrow{f} Y$ in $\mathcal{S}_1^{\mathbb{A}^*}$ where Y is an *algebra.* Define $\overline{X} \xrightarrow{\tilde{\tilde{f}}} Y$ as follows. By the definition of freedom, there is a unique $X_1F \xrightarrow{\tilde{f}} Y$

such that $\kappa(\widetilde{f}U_{\mathbb{A}}) = f_1$. If $y \in X_n$, then for any ${}_1A \xrightarrow{\theta} {}_nA$, $yf_n\theta_Y = y\theta_X f_1$ since f in natural (i.e. a map in $\mathcal{S}_1^{\mathbb{A}^*}$.) But since $Y_n = Y_1^n$, by uniqueness we have

$$yf_n = \langle y\pi^n_{0\,X} f_1, \ldots, y\pi^n_{n-1\,X} f_1\rangle = \langle y\pi^n_{0\,X}, \ldots, y\pi^n_{n-1\,X}\rangle f_1^n.$$

Thus

$$\begin{aligned}
(y\theta_X\kappa)(\widetilde{f}U_{\mathbb{A}}) &= y\theta_X\kappa(\widetilde{f}U_{\mathbb{A}}) = y\theta_X f_1 \\
&= yf_n\theta_Y = \langle y\pi^n_{0\,X}, \ldots, y\pi^n_{n-1\,X}\rangle f_1^n\theta_Y \\
&= \langle y\pi^n_{0\,X}, \ldots, y\pi^n_{n-1\,X}\rangle(\kappa(\widetilde{f}U_{\mathbb{A}}))^n\theta_Y \\
&= \langle y\pi^n_{0\,X}, \ldots y\pi^n_{n-1\,X}\rangle\kappa^n(\widetilde{f}U_{\mathbb{A}})^n\theta_Y \\
&= \langle y\pi^n_{0\,X}\kappa, \ldots, y\pi^n_{n-1\,X}\kappa\rangle(\widetilde{f}U_{\mathbb{A}})^n\theta_Y \\
&= \langle y\pi^n_{0\,X}\kappa, \ldots y\pi^n_{n-1\,X}\kappa\rangle\theta_{X_1F}(\widetilde{f}U_{\mathbb{A}}).
\end{aligned}$$

That is, the map $X_1FU_{\mathbb{A}} \xrightarrow{\widetilde{f}U_{\mathbb{A}}} YU_{\mathbb{A}}$ takes the congruence relation XE into equality, so there is a factorization

$$\begin{tikzcd}
& X_1F \arrow[dl, "\eta"'] \arrow[dr, "\widetilde{f}"] & \\
\overline{X} = X_1F/XE \arrow[rr, "\approx{f}"'] & & Y
\end{tikzcd}$$

By construction $\overset{\approx}{f}$ is the unique map such that $\varphi_1\overset{\approx}{f}_1 = \kappa\eta\overset{\approx}{f}_1 = \kappa\widetilde{f}_1 = f_1$ and hence the unique map such that

$$\varphi\overset{\approx}{f} = f.$$

Therefore $X \xrightarrow{\varphi} \overline{X}$ satisfies the universal mapping property of Theorem I.2.1, so that the inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ has the adjoint $X \longrightarrow \overline{X}$. ∎

Corollary. *An algebraic category has arbitrary small coproducts.*

Proof. As shown above, the full inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ has an adjoint. The category $\mathcal{S}_1^{\mathbb{A}^*}$ clearly has coproducts, by Proposition I.2.2 (dualized); namely, the coproduct of a family of prealgebras is just the prealgebra whose value at n is the sum (in $\mathcal{S}_1$) of the family of values at n. Now consider any small set Λ and any $\Lambda \longrightarrow \mathcal{S}_1^{(\mathbb{A}^*)}$. Let X be the direct limit in $\mathcal{S}_1^{\mathbb{A}^*}$ of this functor, and let $\overline{X}$ be the value at X of the adjoint, $X \xrightarrow{\varphi} \overline{X}$

the canonical map. Then if Y is any algebra, $X_\lambda \xrightarrow{y_\lambda} Y$ any family of maps, $\lambda \in \Lambda$, we have unique maps

$$\begin{tikzcd}
X_\lambda \arrow[rr, "e_\lambda"] \arrow[drrr, "y_\lambda"'] & & X \arrow[rr, "\varphi"] \arrow[dr, dotted] & & \overline{X} \arrow[dl, dotted] \\
& & & Y &
\end{tikzcd}$$

the first since $X = \underset{\lambda\in\Lambda}{\star} X_\lambda$ in $\mathcal{S}_1^{\mathbb{A}^*}$, the second by adjointness. But this shows that $\overline{X} = \underset{\lambda\in\Lambda}{\star} X_\lambda$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$. ∎

Remark. It is clear that the above proof is much more general than the statement of the corollary. We can actually state the following: If a full subcategory of a right complete category is such that the inclusion admits an adjoint, then the subcategory itself is right complete. Of course, the direct limits in the subcategory will not usually agree with those in the big category, i.e., the inclusion is ordinarily not right continuous, even though it is as left continuous as is possible.

2. Algebraic functors and their adjoints

Definition. *Let $\mathbb{A}'$, $\mathbb{A}$ be algebraic theories, considered as small categories. Let $\mathbb{A}' \xrightarrow{f} \mathbb{A}$ be any functor which commutes with finite coproducts. Then there is a unique factorization*

$$\begin{tikzcd}
\mathcal{S}_1^{(\mathbb{A}^*)} \arrow[r, "T"] \arrow[d] & \mathcal{S}_1^{(\mathbb{A}'^*)} \arrow[d] \\
\mathcal{S}_1^{\mathbb{A}^*} \arrow[r, "\mathcal{S}_1^f"'] & \mathcal{S}_1^{\mathbb{A}'^*}
\end{tikzcd}$$

Any functor between algebraic categories which is equivalent to some T constructed as above will be called an **algebraic functor.**

It is clear that every algebraic functor has a **degree**, namely the unique k such that $(1)f = k$, which is also the unique k such that $U_{\mathbb{A}}^k = TU_{\mathbb{A}'}$, where $\mathbb{A}' \xrightarrow{f} \mathbb{A}$, T are as above. The algebraic functors of degree one are precisely those functors equivalent to some value of the functor

$$\mathcal{T}^* \xrightarrow{\mathfrak{S}} \mathcal{K} \longrightarrow \mathcal{C}_2.$$

Theorem 1. *Every algebraic functor has an adjoint.*

Proof. Consider the square above. $\mathcal{S}_1^f$ has an adjoint by Theorem I.2.5. The inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ has an adjoint by Theorem 1.1. The composite provides an adjoint for T since the inclusion $\mathcal{S}_1^{(\mathbb{A}'^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}'^*}$ is full. ∎

Example. If $\mathbb{A}' = \mathcal{S}_0$ (the identity functor considered as an object in $\mathcal{T}$), then the unique $\mathcal{S}_0 \longrightarrow \mathbb{A}$ induces an algebraic functor T (of degree one)

$$\begin{tikzcd}
\mathcal{S}_1^{(\mathbb{A}^*)} \arrow[r, "T"] \arrow[d] & \mathcal{S}_1^{(\mathcal{S}_0^*)} \cong \mathcal{S}_1 \arrow[d] \\
\mathcal{S}_1^{\mathbb{A}^*} \arrow[r] & \mathcal{S}_1^{\mathcal{S}_0^*}
\end{tikzcd}$$

whose adjoint T assigns to each $S \in |\mathcal{S}_1|$ the free $\mathbb{A}$-algebra over S. We have actually analyzed the construction of the free algebra into two steps, as follows. T is equivalent to the composite

$$\begin{tikzcd}
\mathcal{S}_1^{(\mathbb{A}^*)} \arrow[r] & \mathcal{S}_1^{\mathbb{A}^*} \arrow[rr, "E_1"] \arrow[dr] & & S_1 \\
& & \mathcal{S}_1^{\mathcal{S}_0^*} \arrow[ur] &
\end{tikzcd}$$

where the last is evaluation at 1. By a corollary to Theorem I.2.5 we have an explicit formula for the value of the adjoint $\hat{E}_1$ of E_1 at S:

$$\begin{aligned}
({}_nA)(S)\hat{E}_1 &= ({}_1A, {}_nA)^* \cdot S &= \quad ({}_nA, {}_1A) \cdot S \\
&= ({}_1A, {}_1A)^n \cdot S &= ({}_1A, {}_1A)^n \times S.
\end{aligned}$$

Here $({}_1A, {}_1A)^n$ is the set of all n-tuples of unary operations of $\mathbb{A}$. The value at $(S)\hat{E}_1$ of the adjoint to the inclusion $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1^{\mathbb{A}^*}$ is $(S)F$ where $F = \hat{T}$ is the free algebra functor, and the natural transformation φ constructed in Section 1 gives in particular a sequence of maps

$$({}_1A, {}_1A)^n \times S \xrightarrow{\varphi_n} (SFU_{\mathbb{A}})^n$$

in $\mathcal{S}_1$. Here

$$\langle\langle\theta_0, \ldots, \theta_{n-1}\rangle, s\rangle\varphi_n = \langle s\kappa\theta_0, \ldots, s\kappa\theta_{n-1}\rangle$$

where $s \in S$ and where $\langle\theta_0, \ldots, \theta_{n-1}\rangle$ is an n-tuple of unary operations of $\mathbb{A}$.

In particular, if $\mathbb{A}$ is the theory of monoids, then it is known that the underlying set of free monoid is

$$(S)FU_{\mathbb{A}} = \sum_{k \in N} S^k$$

where N is the set of non-negative integers. Also the set of unary operations $({}_1A, {}_1A) \cong N$. Taking ${}_1A \xrightarrow{\theta} {}_2A$ as the multiplication operation, we have

$$\begin{tikzcd}
N^2 \times S \arrow[r, "\varphi_2"] \arrow[d, "\theta \times S"'] & (\sum_{k=0}^{\infty} S^k)^2 \arrow[d, "\theta_{SF}"] \\
N \times S \arrow[r, "\varphi_1"'] & \sum_{k=0}^{\infty} S^k
\end{tikzcd}$$

where

$$\begin{tikzcd}
\langle i, j, s\rangle \arrow[r, "\varphi_2"] \arrow[d] & \langle \underbrace{sss\ldots}_{i}, \underbrace{sss\ldots}_{j}\rangle \arrow[d] \\
\langle i+j, s\rangle \arrow[r, "\varphi_1"'] & \langle \underbrace{sss\ldots}_{i+j}\rangle.
\end{tikzcd}$$

Example. Let $\mathcal{A}$ be the category of abelian groups, $\mathcal{G}$ the category of all (small) groups, $\mathcal{M}$ the category of monoids, $\mathcal{R}$ the category of rings, $\mathcal{M}_c$ and $\mathcal{R}_c$ those of commutative monoids and rings respectively. Let $\mathbb{A}$ be the theory presented:

n	S	E
0	ϵ	$\epsilon \equiv \nu\epsilon$
1	ν	$\pi_0^1 \equiv \nu\nu$

$S_n = E_n = 0$ for $n > 1$.
Then there are obvious algebraic functors of degree one

$$\begin{tikzcd}
\mathcal{A} \arrow[r] \arrow[d] & \mathcal{M}_c \arrow[d] & \mathcal{R}_c \arrow[l] \arrow[d] \\
\mathcal{G} \arrow[r] \arrow[d] & \mathcal{M} \arrow[d] & \mathcal{R} \arrow[l] \arrow[d] \\
\mathcal{S}_1^{(\mathbb{A}^*)} \arrow[r] & \mathcal{S}_1 & \mathcal{A} \arrow[l]
\end{tikzcd}$$

forming a commutative diagram. (Actually all these functors are of the type known to logicians as 'reducts', i.e. they are induced by maps of theories which are induced by inclusions on the usual presentations.) We now describe their adjoints (the description are mostly immediate from [Bourbaki, (1950–1959)] or [Chevalley, 1956]). First, the adjoint of any composite functor ending at $\mathcal{S}_1$ is the free-algebra functor of the appropriate type, e.g. the adjoint to $\mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1$ assigns to each small set S the $\mathbb{A}$-algebra X such that $X_1 = \{\epsilon\}+S+S$, and ν_X switches the last two summands. The adjoint to any of the three vertical functors from the first row to the second consists of 'dividing by the commutator' in the appropriate sense. The adjoint to the left vertical arrow from the second row to the third assigns to each $\mathbb{A}$-algebra X the group obtained by reducing the free group over X_1 modulo the relations $\epsilon \equiv e$, $x \cdot \nu(x) \equiv \nu(x) \cdot x \equiv e$. The adjoint to the vertical functor from rings to abelian groups assigns to each abelian group A the tensor ring $Z+A+A\otimes A+A\otimes A\otimes A+\ldots$. The adjoints to the two horizontal functors from groups to monoids are described in [Chevalley, 1956], pages 41-42. Finally, the adjoints to the two horizontal functors from rings to monoids assign to a given monoid M the monoid ring $Z[M]$ with integer coefficients.

The fact that the adjoint functors also form a commutative diagram (with arrows reversed) implies, for example, the well-known fact that a free ring may be constructed either as the tensor ring of a free abelian group or as the monoid ring of a free monoid.

The above diagram and discussion have an obvious modification by applying a fixed $\Lambda \in \mathcal{R}$ to each category in the right hand column.

The only functors in the above diagram which also have *coadjoints* are the two from groups to monoids, whose coadjoints assign to a monoid or commutative monoid, respectively, its group of units.

Example. For an example which is not a reduct, consider the functor from rings to Lie rings induced by the $\mathcal{T}$-map defined at the end of Chapter II. The adjoint to this functor assigns to each Lie ring its associative **enveloping ring**. (See [Cartan & Eilenberg, 1956].)

Example. Let $\mathbb{A}$ be the theory of commutative associative rings with unity, and $\mathbb{A}'$ be the theory presented like $\mathbb{A}$ with the exception that $S_0' = \{i\} + S_0$ where i satisfies the identity $i^2 = -1$. The adjoint to the obvious reduct is itself an algebraic functor, but of degree two.

Chapter V

Certain 0-ary and unary extensions of algebraic theories

1. Presentations of algebras: polynomial algebras

Consider the functor $\mathcal{S}_1 \xrightarrow{T_0} \mathcal{T}$ such that $ST_0 = S \cdot \mathbb{I}_0$; that is, T_0 assigns to each small set S the free algebraic theory over the sequence of sets $S_0 = S$, $S_n = 0$, $n > 0$. Since T_0 is right continuous and full, the subcategory of theories which arise as values of T_0 is closed under direct limits, in particular coproducts and coequalizers (quotients). The corresponding categories of algebras are not much more complicated than the category of sets itself. For example, we have the

Proposition. *If S is any small set, then coproducts in the category $\mathcal{S}_1^{(ST_0^*)}$ are 'wedge products'. That is, given any two ST_0-algebras X, Y, their coproduct is the* comeet *(of sets)*

$$
\begin{tikzcd}
S \arrow[r] \arrow[d] & X_1 \arrow[d] \\
Y_1 \arrow[r] & (X \star_{ST_0} Y)_1
\end{tikzcd}
$$

where $S \longrightarrow X_1$, $S \longrightarrow Y_1$ are the unique maps defining the structure of X, Y.

Definition. *Let S be any small set, $\mathbb{A}$ any algebraic theory, $N \xrightarrow{E} S_1$ any sequence of sets, and $E \xrightarrow{r} ((\mathbb{A} \star ST_0)T)^2$ any map in $\mathcal{S}_1^N$ (where T is the functor discussed in II.2). Write $\mathbb{A}' = (\mathbb{A} \star ST_0)/E$ for the coequalizer of the corresponding pair of maps $\langle r_0F, r_1F\rangle$ in $\mathcal{T}$, and denote by f the composite $\mathcal{T}$-map*

$$\mathbb{A} \longrightarrow \mathbb{A} \star ST_0 \longrightarrow (\mathbb{A} \star ST_0)/E = \mathbb{A}'.$$

Then by the **algebra presented by** $\langle \mathbb{A}, S, E, r\rangle$ *is meant the $\mathbb{A}$-algebra $({}_0A')\mathcal{S}_1^{(f^*)}$, where ${}_0A' = \lim_{\to\circ}^{\mathcal{S}_1^{(\mathbb{A}'^*)}}$. $\mathbb{A}'$ is the* **theory of the presentation.**

Theorem 1. *If X is the algebra presented by $\langle \mathbb{A}, S, E, r\rangle$, then there is an equivalence*

$$(X, \mathcal{S}_1^{(\mathbb{A}^*)}) \cong \mathcal{S}_1^{(\mathbb{A}'^*)}$$

where $\mathbb{A}'$ is the theory of the presentation. If X is any $\mathbb{A}$-algebra, then $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$ is algebraic; in fact, there is a presentation of X such that the above relation holds.

Recall that the objects in the category $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$ are maps $X \longrightarrow Y$ of $\mathbb{A}$-algebras, and that maps in this category are commutative triangles

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Y \arrow[rr] & & Y' \end{tikzcd}$$

is $\mathcal{S}_1^{(\mathbb{A}^*)}$. The Theorem follows from Lemmas 1.1, 1.2, 1.3 below.

Lemma 1. *If X is any $\mathbb{A}$-algebra, then the functor*

$$(X, \mathcal{S}_1^{(\mathbb{A}^*)}) \xrightarrow{V} \mathcal{S}_1^{(\mathbb{A}^*)}$$

has an adjoint, which assigns to each $\mathbb{A}$-algebra Y the injection $X \longrightarrow Y \star_{\mathbb{A}} X$, considered as an object $Y\hat{V}$ in $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$.

Proof. Since $(X \longrightarrow Y \star X)V = Y \star X$, there is an obvious map $Y \xrightarrow{\varphi} Y\hat{V}V$, namely the injection $Y \longrightarrow Y \star X$. If $X \xrightarrow{y} Y'$ is any object in $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$, and if $Y \xrightarrow{\psi} yV = Y'$ is any map in $\mathcal{S}_1^{(\mathbb{A}^*)}$, then there is a unique map λ in $\mathcal{S}_1^{(\mathbb{A}^*)}$ such that

$$\begin{tikzcd} X \arrow[r] \arrow[dr, "y"'] & Y \star_{\mathbb{A}} X \arrow[d, dashed, "\lambda"] & Y \arrow[l, "\varphi"'] \arrow[dl, "\psi"] \\ & Y' & \end{tikzcd}$$

is commutative. Because the left hand triangle is commutative, λ defines a map $Y\hat{V} \xrightarrow{\lambda} y$ in $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$, which is the unique map such that $\varphi(\lambda V) = \psi$. ■

Lemma 2. *Consider the composite functor U:*

$$(X, \mathcal{S}_1^{(\mathbb{A}^*)}) \longrightarrow \mathcal{S}_1^{(\mathbb{A}^*)} \xrightarrow{U_{\mathbb{A}}} \mathcal{S}_1.$$

The value at $X \longrightarrow Y$ of U is the set of maps ${}_0A \star X \longrightarrow Y$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$ such that

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ {}_0A \star X \arrow[rr] & & Y \end{tikzcd}$$

is commutative. If $\mathbb{A}' = U\hat{\mathfrak{S}}$ is the algebraic structure of U, then the map $U \xrightarrow{\Phi} U\hat{\mathfrak{S}}\mathfrak{S}$ in $\mathcal{K}$ is an equivalence, i.e.

$$\begin{tikzcd} (X, \mathcal{S}_1^{(\mathbb{A}^*)}) \arrow[r, "\Phi", "\approx"'] \arrow[d] \arrow[dr, "U"] & \mathcal{S}_1^{(\mathbb{A}'^*)} \arrow[d, "U_{\mathbb{A}'}"] \\ \mathcal{S}_1^{(\mathbb{A}^*)} \arrow[r, "U_{\mathbb{A}}"'] & \mathcal{S}_1 \end{tikzcd}$$

is commutative in $\mathcal{C}_2$ and Φ is an equivalence of categories.

Proof. The first assertion is immediate by Lemma 1.1. The commutativity in $\mathcal{C}_2$ follows from our work in Chapter III (where the definition of Φ was given). We need to show that Φ is an equivalence. Now the n-ary operations ${}_1A' \xrightarrow{\theta} {}_nA'$ of $\mathbb{A}'$ are in one-to-one correspondence with commutative triangles

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ {}_1A \star_{\mathbb{A}} X \arrow[rr] & & {}_nA \star_{\mathbb{A}} X \end{tikzcd}$$

in $\mathcal{S}_1^{(\mathbb{A}^*)}$ (where the legs are the injections). In particular, every n-ary operation of $\mathbb{A}$ determines such a triangle, so that there is a map $\mathbb{A} \xrightarrow{f} \mathbb{A}'$ of theories and a corresponding map $\mathbb{A}'\mathfrak{S} \xrightarrow{f\mathfrak{S}} \mathbb{A}\mathfrak{S}$ in $\mathcal{K}$. We also have, since ${}_0A \star_{\mathbb{A}} X \cong X$, that the 0-ary operations of $\mathbb{A}'$ are in one-to-one correspondence with maps ${}_1A \star_{\mathbb{A}} X \longrightarrow X$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$ such that $X \longrightarrow {}_1A \star_{\mathbb{A}} X \longrightarrow X$ is the identity, which in turn are in one-to-one correspondence with maps ${}_1A \longrightarrow X$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$ Therefore $({}_0A')\mathcal{S}_1^{(f^*)} \cong X$. Since ${}_0A' = \varinjlim_{\circ}$, there is, for each $\mathbb{A}'$-algebra Y, a unique map ${}_0A' \longrightarrow Y$ in $\mathcal{S}_1^{(\mathbb{A}'^*)}$ which gives a map $Y\Psi : X = ({}_0A')\mathcal{S}_1^{(f^*)} \longrightarrow Y\mathcal{S}_1^{(f^*)}$ in $\mathcal{S}_1^{(\mathbb{A}^*)}$. There thus results a functor $\mathcal{S}_1^{(\mathbb{A}'^*)} \xrightarrow{\Psi} (X, \mathcal{S}_1^{(\mathbb{A}^*)})$ such that $\Psi\Phi \cong \mathcal{S}_1^{(\mathbb{A}'^*)}$ and $\Phi\Psi \cong (X, \mathcal{S}_1^{(\mathbb{A}*)})$. ■

Lemma 3. *If X is any $\mathbb{A}$-algebra, then the algebraic structure $\mathbb{A}' = U\hat{\mathfrak{S}}$ of the functor $(X, \mathcal{S}_1^{(\mathbb{A}^*)}) \longrightarrow \mathcal{S}_1^{(\mathbb{A}^*)} \longrightarrow \mathcal{S}_1$ is the theory of a presentation $\langle \mathbb{A}, S, E, r\rangle$ of which X is the algebra presented.*

Proof. It was pointed out in the proof of Lemma 1.2 that $X = ({}_0A')\mathcal{S}_1^{(f^*)}$. A map $\mathbb{A} \xrightarrow{f} \mathbb{A}'$ of theories was constructed and it was pointed out that $({}_1A, X) \cong ({}_1A', {}_0A')$. Thus setting $S = ({}_1A, X)$, we have a map g of theories defined by

$$\begin{tikzcd}
\mathbb{A} \arrow[r] \arrow[dr, "f"'] & \mathbb{A} \star ST_0 \arrow[d, "g"] & ST_0 \arrow[l] \arrow[dl, "h"] \\
& \mathbb{A}' &
\end{tikzcd}$$

where h corresponds to the isomorphism $S \cong (\mathbb{A}'T)_0$ under the isomorphism

$$(ST_0, \mathbb{A}') \cong (S, (\mathbb{A}'T)_0)$$

determined by the definition of T_0. Letting $\mathbb{K}$ be the equalizer of

$$(\mathbb{A} \star ST_0)^2 \rightrightarrows \mathbb{A} \star ST_0 \xrightarrow{g} \mathbb{A}'$$

and defining E, r by

$$\mathbb{K}T = E \xrightarrow{r} ((\mathbb{A} \star ST_0)T)^2$$

it follows that $\langle \mathbb{A}, S, E, r\rangle$ has all the correct properties, as it is clear from Chapter II that g is the coequalizer of

$$EF \longrightarrow \mathbb{K} \longrightarrow (\mathbb{A} \star ST_0)^2 \rightrightarrows \mathbb{A} \star ST_0$$

where F is the free theory functor. ■

Example. In particular, if $S \in |\mathcal{S}_1|$, then

$$(S, \mathcal{S}_1) \cong \mathcal{S}_1^{(ST_0^*)}.$$

Example. If $\Lambda \in \mathcal{R}_c$, then $(\Lambda, \mathcal{R}_c)$ is equivalent to the usual category of commutative associative Λ-algebras. However, if $\Lambda \in \mathcal{R}$, then the usual (algebraic) category of associative Λ-algebras is the full subcategory of $(\Lambda, \mathcal{R})$ determined by objects $\Lambda \xrightarrow{x} X$ such that the image of x lies in the center of X.

The structure of a category of the form $(X, \mathcal{S}_1^{(\mathbb{A}^*)})$ can of course be studied entirely within $\mathcal{S}_1^{(\mathbb{A}^*)}$. This may be considered a partial motivation for the introduction of the following

Definition. *If $\mathbb{A}$ is an algebraic theory and if X is an $\mathbb{A}$-algebra, then by the* **algebra of polynomials in n variables with coefficients in** X *is meant the $\mathbb{A}$-algebra ${}_nA \star_{\mathbb{A}} X$.*

Example. If $\Lambda \in \mathcal{R}$, then the algebra of polynomials in n variables with coefficients in Λ in the ring $\Lambda[\pi_0^n, \ldots, \pi_{n-1}^n]$ of polynomials in n *noncommuting* variables.

Proposition 1. *If X, $\mathbb{A}$, $\mathbb{A}'$ are as in the preceding Theorem, the members of the algebra of polynomials in n variables with coefficients in X are in one-to-one correspondence with the n-ary operations of $\mathbb{A}'$.*

PROOF. Obvious from the foregoing. Here by 'members' of ${}_nA \star_{\mathbb{A}} X$ we mean of course maps ${}_1A \longrightarrow {}_nA \star_{\mathbb{A}} X$. ∎

Proposition 2. *If X is an $\mathbb{A}$-algebra, then every n-tuple ${}_1A \xrightarrow{x} X^n$ of members of X determines an evaluation homomorphism ${}_nA \star_{\mathbb{A}} X \xrightarrow{(x)} X$.*

PROOF. ${}_1A \xrightarrow{x} X^n$ is equivalent to a map ${}_nA \xrightarrow{\overline{x}} X$, which together with the identity map X yields

$$\begin{tikzcd}
X \arrow[r] \arrow[dr, "X"'] & {}_nA \star_{\mathbb{A}} X \arrow[d, "(x)"] & {}_nA \arrow[l] \arrow[dl, "\overline{x}"] \\
& X &
\end{tikzcd}$$

∎

Definition. *If θ is a polynomial in n variables with coefficients in X (i.e. a member ${}_1A \xrightarrow{\theta} {}_nA \star_{\mathbb{A}} X$ of ${}_nA \star_{\mathbb{A}} X$) and if x is an n-tuple of members of X, then the composite $\theta(x)$ is a member of X, the* **value at x of θ.**

Remark. This shows that it is consistent to write $\theta(x)$ for the evaluation (or composition) of polynomials, and xf for the evaluation of homomorphisms f. This may be regarded as another manifestation of the duality between structure and maps as expressed by our Theorem III.1.2.

Example. Let $M \in \mathcal{M}$ be a monoid, $n = 1$. Because of the 'interlacing' description of the coproduct of monoids, and since ${}_1A \cong N$ (the additive monoid of non-negative integers) in this case, we see that any unary polynomial θ with coefficients in M can be represented as a string

$$m_0\ n_0\ m_1\ n_1 \ldots m_{k-1}n_{k-1}$$

where $m_i \in M$, $n_i \in \mathbb{N}$ for $i \in k$. The value of θ at a member x of M is the product

$$\theta(x) = m_0 x^{n_0} m_1 x^{n_1} m_2 x^{n_2} \ldots m_{k-1} x^{n_{k-1}}.$$

A similar remark holds for groups, except that the n_i may have negative values in that case.

2. Monoids of operators

If M is any small monoid, then there are unique functors $\mathtt{1} \rightleftarrows M$. If $\mathcal{B}$ is any complete category, these induce functors

$$\mathcal{B}^M \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathcal{B}$$

and we have $\beta\alpha = \mathcal{B}$. By Theorem I.2.5 and its dual, α and β have adjoints $\hat{\alpha}$, $\hat{\beta}$ and coadjoints $\check{\alpha}$, $\check{\beta}$.

Proposition 1.

$$\begin{array}{rclcrcl} \hat{\alpha}\hat{\beta} & = & \mathcal{B} & , & \check{\alpha}\check{\beta} & = & \mathcal{B} \\ \beta\hat{\beta} & = & \mathcal{B} & , & \beta\check{\beta} & = & \mathcal{B} \end{array}$$

PROOF. The first two equations follow from the equation $\beta\alpha = \mathcal{B}$. The other two are immediate since β is full. ∎

Proposition 2. *If $X \in |\mathcal{B}^M|$, then $X\hat{\beta} = X/M$, the 'orbit object', and there is a regular epimap $X \longrightarrow (X/M)\beta$. $X\check{\beta} = M\backslash X$, the 'fixed object', and there is a regular monomap $(M\backslash X)\beta \longrightarrow X$.*

PROOF. Immediate from the proof of Theorem I.2.2 and the fact that $\hat{\beta} = \lim\limits_{\rightarrow M}^{\mathcal{B}}$ while $\check{\beta} = \lim\limits_{\leftarrow M}^{\mathcal{B}}$. ∎

Remark. If M is not a group, and if e.g. $\mathcal{B} = \mathcal{S}_1$, then the 'orbit object' does not necessarily consist of 'orbits' in the usual sense, since these need not form a partition if M is not a group. However, by X/M we mean the quotient by the smallest equivalence (or more generally congruence) relation which requires that any two points on the same orbit are equivalent. In case M is commutative, this equivalence relation is simply

$$x \equiv y \text{ iff } \exists m\ \exists m'\ [xm = ym'].$$

Proposition 3. *If $B \in |\mathcal{B}|$, then $B\hat{\alpha}$ is the functor $M \longrightarrow \mathcal{B}$ whose value at the object $\mathbf{1} \xrightarrow{e} M$ is the object $(e,e) \cdot B$ in $\mathcal{B}$. For any $m' \in M$, $m'(B\hat{\alpha})$ is the map $(e,e) \cdot B \longrightarrow (e,e) \cdot B$ determined by the commutativity of*

$$\begin{tikzcd}
B \arrow[r, "B"] \arrow[d, "j_m"'] & B \arrow[d, "j_{mm'}"] \\
(e,e) \cdot B \arrow[r, "m'(B\hat{\alpha})"'] & (e,e) \cdot B
\end{tikzcd}$$

for every $m \in (e,e) = |M^{\mathbf{2}}|$ =set of members of M (the j_m being the injections into the (e,e)-fold coproduct). $B\check{\alpha}$ is the functor $M \longrightarrow \mathcal{B}$ whose value at e is the (e,e)-fold product $B^{(e,e)}$ and whose value at any $m' \in M$ is the map $B^{(e,e)} \longrightarrow B^{(e,e)}$ determined by the commutativity of

$$\begin{tikzcd}
B^{(e,e)} \arrow[r, "m'(B\check{\alpha})"] \arrow[d, "p_{m'm}"'] & B^{(e,e)} \arrow[d, "p_m"] \\
B \arrow[r, "B"'] & B
\end{tikzcd}$$

for every $m \in (e,e)$ (the p_m being the projections).

Proof. The formula of the corollary to Theorem I.2.5 specialized to this case. ∎

Remark. In the case $M = \mathbb{N}$, the additive monoid of non-negative integers, $B^\infty = B\check{\alpha}\alpha$ (the 'object of sequences of B') is characterized in $\mathcal{B}$ by the generalized and dualized Peano's postulate:

$$\forall X\ \forall t\ \forall x\ \exists! f \qquad \begin{tikzcd}
B^\infty \arrow[r, "s"] & B^\infty \arrow[r, "e_0"] & B \\
X \arrow[r, "t"'] \arrow[u, dashed, "f"] & X \arrow[r, "x"'] \arrow[u, dashed, "f"] & B \arrow[u, "B"']
\end{tikzcd}$$

This shows that the N-fold product (and dually coproduct) of an object with itself is a concept which is definable within the first-order theory of a category, whereas infinite products and coproducts in general are of course not first-order definable (without passing to the theory of the category of categories).

By Theorem III.2.2, if $\mathcal{B}$ is an algebraic category and M is a small monoid (i.e. category with one object), then $\mathcal{B}^M$ is also algebraic. Our aim now will be to describe explicitly the algebraic structure of $\mathcal{B}^M \longrightarrow \mathcal{B} \longrightarrow \mathcal{S}$, and to show that the functors α and β discussed above are algebraic functors of degree one.

For any small monoid M, let MT_1 be the algebraic theory whose n-ary operations are all of the form

$$m\pi_i^n,\ m \in M,\ i \in n$$

and which satisfies the relations

$$m'(m\pi_i^n) = (mm')\pi_i^n$$

where mm' on the right hand side is composition in M.
This defines a functor

$$\mathcal{M}_1 \xrightarrow{T_1} \mathcal{T}.$$

In particular $(\mathbf{1})T_1 = \mathcal{S}_0$.
Consider the functor

$$\mathcal{T} \times \mathcal{M} \longrightarrow \mathcal{T}$$

which assigns to $\langle \mathbb{A}, M\rangle$ the algebraic theory

$$\mathbb{A}[M] = (\mathbb{A} \star MT_1)/R(\mathbb{A}, M)$$

where $R(\mathbb{A}, M)$ is the smallest congruence (in the sense of $\mathcal{T}$) containing all relations of the form

$$\theta\{m\pi_0^n, m\pi_1^n, \ldots, m\pi_{n-1}^n\} = m\theta\{\pi_0^n, \ldots \pi_{n-1}^n\}$$

where θ is an n-ary operation of $\mathbb{A}$ and $m \in M$. In particular $\mathbb{A}[\mathbf{1}] \cong \mathbb{A}$ for each algebraic theory, and for each monoid M, the maps $\mathbf{1} \rightleftarrows M$ induce maps $\mathbb{A} \underset{b}{\overset{a}{\rightleftarrows}} \mathbb{A}[M]$ in $\mathcal{T}$.

Theorem 1. *For any algebraic theory $\mathbb{A}$ and for any small monoid M,*

$$\mathcal{S}_1^{(\mathbb{A}^*)^M} \cong \mathcal{S}_1^{(\mathbb{A}[M]^*)}.$$

Also, for the functors

$$\mathcal{S}_1^{(\mathbb{A}^*)^M} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathcal{S}_1^{(\mathbb{A}^*)}$$

induced by $\mathbf{1} \rightleftarrows M$ we have

$$\begin{aligned} \alpha &\cong \mathcal{S}_1^{(a^*)} \\ \beta &\cong \mathcal{S}_1^{(b^*)}. \end{aligned}$$

Proof. Both $\mathcal{S}_1^{(\mathbb{A}^*)^M}$ and $\mathcal{S}_1^{(\mathbb{A}[M]^*)}$ are equivalent to the full subcategory of $\mathcal{S}_1^{\mathbb{A}^* \times M}$ determined by functors X such that $\langle {}_nA, e\rangle X = \langle {}_1A, e\rangle X^n$ for all $n \in |\mathcal{S}_0|$. By the results of I.2, the functor $\mathcal{S}_1^{\mathbb{A}^*} \longrightarrow \mathcal{S}_1^{\mathbb{A}^* \times M}$ induced by $\mathbb{A}^* \times M \longrightarrow \mathbb{A}^*$ takes $\mathcal{S}_1^{(\mathbb{A}^*)}$ into this subcategory. But the restriction of this functor to $\mathbb{A}$-algebras is β. It follows easily that $\beta = \mathcal{S}_1^{(b^*)}$, and similarly $\alpha = \mathcal{S}_1^{(a^*)}$. ∎

3. Rings of operators (Theories of categories of modules)

Let $\mathcal{R}$ be the category of (small) rings and define a functor

$$\mathcal{R} \xrightarrow{T_1'} \mathcal{T}$$

as follows. For each $R \in |\mathcal{R}|$, RT_1' is the algebraic theory presented as follows. (Note that $(Z[\pi_0^1], R)$ is the set of members of R.)

n	S	E		
0	$\odot$	empty		
1	λ	$0 + \pi_0^1$	$\equiv$	π_0^1
	for $\lambda \in (Z[\pi_0^1], R)$	$\pi_0^1 + 0$	$\equiv$	π_0^1
		1	$\equiv$	π_0^1
		$(\odot)(0 \longrightarrow 1)$	$\equiv$	0
		$(\lambda + \lambda')\pi_0^1$	$\equiv$	$\lambda\pi_0^1 + \lambda'\pi_0^1$
		$(\lambda \cdot \lambda')\pi_0^1$	$\equiv$	$\lambda'(\lambda\pi_0^1)$
		for $\lambda, \lambda' \in (Z[\pi_0^1], R)$		
2	$+$	$\lambda(\pi_0^2 + \pi_1^2)$	$\equiv$	$\lambda\pi_0^2 + \lambda\pi_1^2$
		for $\lambda \in (Z[\pi_0^1], R)$		
3	empty	empty		
4	empty	$(\pi_0^4 + \pi_1^4) + (\pi_2^4 + \pi_3^4)$	$\equiv$	$(\pi_0^4 + \pi_2^4) + (\pi_1^4 + \pi_3^4)$

$S_n = E_n = 0$ for $n > 4$.

Thus ZT_1' is the theory of abelian groups, $Z[\pi_0^1]T_1'$ is the theory of abelian groups with a distinguished endomorphism, and in general $\mathcal{S}_1^{(RT_1'^{*})}$ is the category of (right) R-modules.

Proposition 1. *For any map $\varphi : \Lambda \longrightarrow \Gamma$ in $\mathcal{R}$, the functor $\mathcal{S}_1^{(\varphi T'^{*}_1)}$ has a coadjoint (as well as an adjoint).*

Proof. Well known, see e.g. [Cartan & Eilenberg, 1956]. The adjoint is

$$X \longrightarrow X \otimes {}_{\Lambda}\Gamma$$

and the coadjoint is

$$X \longrightarrow Hom_{\Lambda}(\Gamma, X).$$

■

Proposition 2. *The diagram*

$$\begin{tikzcd}
\mathcal{R} \times \mathcal{M} \arrow[rr] \arrow[d, shift right, "T_1'"'] \arrow[d, shift left, "\mathcal{M}"] & & \mathcal{R} \arrow[d, "T_1'"] \\
\mathcal{T} \times \mathcal{M} \arrow[rr] & & \mathcal{T}
\end{tikzcd}$$

is commutative (up to equivalence) where the bottom row is the functor $\langle \mathbb{A}, M\rangle \longrightarrow \mathbb{A}[M]$ *of Section 2, and where the top row is* $\langle R, M\rangle \longrightarrow R[M] = R \otimes Z[M]$.

Proof. It is well known that $\mathcal{S}_1^{(R[M]T_1'^*)} \cong \mathcal{S}_1^{(RT_1'^*)^M}$. By the Theorem of Section 2, $\mathcal{S}_1^{(RT_1'^*)^M} \cong \mathcal{S}_1^{(RT_1'[M]^*)}$. Since these equivalences preserve underlying sets, the algebraic structures are also equivalent by III.1. That is

$$R[M]T_1' \cong RT_1'[M].$$

■

References

Bourbaki, N. (1950–1959), *Algèbre,* II. Eléments de Mathématique, Actualités Sci. Ind., Hermann, Paris. Nos. 1144, 1958; 1032–1236, 1955; 1044, 1958; 1102, 1959; 1179, 1952; 1261, 1958; 1272, 1959.

Cartan, H. and Eilenberg, S. (1956), *Homological algebra.* Princeton University Press, Princeton, N.J.

Chevalley, C. (1956), *Fundamental concepts of algebra.* Academic Press, New York.

Eilenberg, S. and Mac Lane, S. (1945), *General theory of natural equivalences*, Trans. Amer. Math. Soc. **58**, 231–294.

Freyd, P.J. (1960), *Functor theory.* Thesis, Princeton University.

Gaifman, H. (1961), *Free complete boolean algebras.* Notices Amer. Math. Soc. **8**, 510.

Godement, R. (1958), *Topologie algébrique et théorie des faisceaux.* XIII Publications de l'Institut de Mathématique de l'Université de Strasbourg, Hermann, Paris,

Gray, J.W. (1962), *Sheaves with values in a category.* National Science Foundation Report CU-5-62 G 19002.

Kan, D.M. (1958), *Adjoint functors*, Trans. Amer. Math. Soc. **87**, 294–329.

Lawvere, F.W. (1963), *The "group ring" of a small category.* Abstract in Notices Amer. Math. Soc.

Mac Lane, S. (1950), *Duality for groups*, Bull. Amer. Math. Soc. **56**, 485–516.

Suppes, P. (1960), *Axiomatic set theory*, Van Nostrand, Princeton, N.J.

Part C

Some Algebraic Problems in the context of Functorial Semantics of Algebraic Theories

Introduction

The categorical approach to universal algebra, initiated in [Lawvere, 1963] has been extended from finitary to infinitary operations in [Linton, 1966a], from sets to arbitrary base categories through the use of triples (monads) in [Eilenberg & Moore, 1965] and [Barr & Beck, 1966] and from one-sorted theories over l-dimensional categories to Γ-sorted theories over 2-dimensional categories in [Bénabou, 1966]. But despite this generality, there is still enough information in the machinery of algebraic categories, algebraic functors, adjoints to algebraic functors, the semantics and structure superfunctors, etc. to allow consideration of specific problems analogous to those arising in group theory, ring theory, and other parts of classical algebra. The approach also suggests new problems. As examples of the latter we may mention Linton's considerations of general "commutative" theories [Linton, 1966b], Barr's discussion of general "distributive" laws [Barr, 1969], and Freyd's construction of Kronecker products of arbitrary theories and tensor products of arbitrary algebras [Freyd, 1966]. It is our purpose here to indicate some of the "specific" aspects of the approach, and also to mention some of the representative problems which seem to be open. We restrict ourselves to the case of finitary single-sorted theories over sets.

1. Basic concepts

An elegant exposition of part of the basic machinery appears in [Eilenberg & Wright, 1967] – we content ourselves here with a brief summary. An *algebraic theory* is a category $\mathbb{A}$ having as objects

$$1, A, A^2, A^3, \ldots$$

and, for each $n = 0, 1, 2, 3, \ldots, n$, morphisms

$$A^n \xrightarrow{\Pi_i^{(n)}} A, \qquad i = 0, 1, \ldots, n-1$$

such that for any n morphisms

$$A^m \xrightarrow{\theta_i} A, \qquad i = 0, 1, \ldots, n-1$$

in $\mathbb{A}$ there is exactly one morphism

$$A^m \xrightarrow{\langle\theta_0,\theta_1,\ldots,\theta_{n-1}\rangle} A^n$$

in $\mathbb{A}$ so that

$$\langle\theta_0, \theta_1, \ldots, \theta_{n-1}\rangle\Pi_i^{(n)} = \theta_i, \qquad i = 0, 1, \ldots, n-1.$$

The arbitrary morphisms $A^n \xrightarrow{\varphi} A$ are called the *n-ary operations* of $\mathbb{A}$. The *algebraic category* associated with $\mathbb{A}$ is the *full* subcategory

$$\mathbb{A}^\flat \subset \mathcal{S}^{\mathbb{A}}$$

consisting of those covariant set-valued functors which are product-preserving; its objects are called $\mathbb{A}$-algebras and its morphisms $\mathbb{A}$-homomorphisms. Clearly there is a full embedding $\mathbb{A}^{\mathrm{op}} \underset{\rightarrow}{\subset} \mathbb{A}^\flat$ which preserves coproducts; its values are the *finitely-generated free* $\mathbb{A}$-algebras, where "free" refers to the left adjoint of the functor "underlying"

$$\mathbb{A}^\flat \xrightarrow{U_{\mathbb{A}}} \mathcal{S}$$

whose value at the algebra X is the value of X at A:

$$XU_{\mathbb{A}} = AX.$$

The underlying functor is a particular *algebraic functor*, where the latter means a functor

$$\mathbb{A}^\flat \xrightarrow{f^\flat} \mathbb{B}^\flat$$

induced by composition of functors from a *theory morphism* $\mathbb{B} \xrightarrow{f} \mathbb{A}$, where a theory morphism is just a functor f such that

$$\left(\Pi_i^{(n)}\right) f = \Pi_i^{(n)}, \quad \text{for all } i \in n \in \omega.$$

Clearly all the theory morphisms determine a category $\mathcal{T}$, and every algebraic functor preserves the underlying functors. Hence $f \rightsquigarrow f^\flat$ determines a *semantics* functor

$$\mathcal{T}^{\mathrm{op}} \longrightarrow (\mathrm{Cat}, \mathcal{S})$$

where the category on the right has as morphisms all commutative triangles

$$\begin{tikzcd}
\mathcal{X} \arrow[rr, "\Phi"] \arrow[dr, "U"'] & & \mathcal{X}' \arrow[dl, "U'"] \\
& \mathcal{S} &
\end{tikzcd}$$

of functors. Switching functor categories a bit shows that *structure*, the left adjoint of semantics, may be calculated as follows: Given a set-valued functor U, the n-ary operations of its algebraic structure are just the *natural transformations* $U^n \xrightarrow{\varphi} U$, where U^n is the n-th cartesian power of U in the functor category $\mathcal{S}^{\mathcal{X}}$, i.e. φ is a way of assigning an operation to every value of U in such a way that all morphisms of $\mathcal{X}$ are homomorphisms with respect to it. Several applications of Yoneda's Lemma show that if in fact $U = U_{\mathbb{A}}, \mathcal{X} = \mathbb{A}^\flat$ for some theory $\mathbb{A}$, then the algebraic structure of U is isomorphic to $\mathbb{A}$. As a corollary every functor $\mathbb{A}^\flat \longrightarrow \mathbb{B}^\flat$ which preserves underlying sets is induced by one and only one theory morphism $\mathbb{B} \longrightarrow \mathbb{A}$. More generally, if we denote by $\mathbb{I}_n$ the *free theory* generated by one n-ary operation, then the n-ary operations of the algebraic structure of any $\mathcal{X} \xrightarrow{U} \mathcal{S}$ are in one-to-one correspondence with the *functors*

$$\mathcal{X} \xrightarrow{\Phi} \mathbb{I}_n^\flat$$

for which $U = \Phi U_{\mathbb{I}_n}$.

Algebraic functors are faithful and possess left adjoints. In fact (as pointed out by M. André and H.Volger), if $\mathbb{B} \xrightarrow{f} \mathbb{A}$ is a morphism of theories then the usual (left) Kan adjoint

$$\mathcal{S}^{\mathbb{B}} \dashrightarrow \mathcal{S}^{\mathbb{A}}$$

corresponding to f actually takes product-preserving functors into product-preserving functors, and so restricts to a functor f_* with

$$f_* \dashv f^\flat.$$

Thus we have the commutative diagram of functors

$$\begin{tikzcd}
\mathbb{B}^{\mathrm{op}} \arrow[r, "f^{\mathrm{op}}"] \arrow[d, "\cap"'] & \mathbb{A}^{\mathrm{op}} \arrow[d, "\cap"] \\
\mathbb{B}^\flat \arrow[r, "f_*"] \arrow[d, "\cap"'] & \mathbb{A}^\flat \arrow[d, "\cap"] \\
\mathcal{S}^{\mathbb{B}} \arrow[r, "\mathrm{Kan}"] & \mathcal{S}^{\mathbb{A}}
\end{tikzcd}$$

Explicitly, for any $\mathbb{B}$-algebra Y, the underlying set of the "relatively free" $\mathbb{A}$-algebra Yf_* is the colimit of

$$(f, A) \longrightarrow \mathbb{B} \xrightarrow{Y} \mathcal{S}$$

where the first factor of this composite is the obvious forgetful functor from the category whose morphisms are triples θ, φ, θ' with θ, θ' operations in $\mathbb{A}$ and φ a morphism in $\mathbb{B}$ such that

$$\begin{tikzcd}
\bullet \arrow[rr, "\varphi f"] \arrow[dr, "\theta"'] & & \bullet \arrow[dl, "\theta'"] \\
& A &
\end{tikzcd}$$

is commutative in $\mathbb{A}$. In particular, free algebras can be computed by such a direct limit by taking $\mathbb{B}$ = the initial theory $\simeq$ the dual of the category of finite sets and maps. For the unique f in this case we also write $f_* = F_{\mathbb{A}}$.

Given two theories $\mathbb{A}$ and $\mathbb{B}$, the category of all product-preserving functors $\mathbb{B} \longrightarrow \mathbb{A}^\flat$ has an obvious underlying set functor, whose algebraic structure is denoted by $\mathbb{A}\otimes\mathbb{B}$, the *Kronecker product* of $\mathbb{A}$ with $\mathbb{B}$. The Kronecker product is a coherently associative functor $\mathcal{T}\times\mathcal{T} \longrightarrow \mathcal{T}$ having the initial theory as unit object; it also satisfies $\mathbb{A}\otimes\mathbb{B} \cong \mathbb{B}\otimes\mathbb{A}$. The foregoing semantical definition of $\mathbb{A}\otimes\mathbb{B}$ is equivalent to the following wholly "theoretical" definition.

$$\mathbb{A} \otimes \mathbb{B} = (\mathbb{A} * \mathbb{B})/R$$

where $\mathbb{A} * \mathbb{B}$ denotes the coproduct in $\mathcal{T}$ and R is the congruence relation generated by the conditions that each $A^n \xrightarrow{\varphi} A$ in $\mathbb{A}$ should be a "homomorphism" with respect to each $A^m \xrightarrow{\psi} A$ in $\mathbb{B}$ [$(A^n)^m \xrightarrow{\sigma} (A^m)^n \xrightarrow{\psi^n} A^n$ being defined as the operation of ψ on A^n, σ being the transpose isomorphism] and that, symmetrically, each $\mathbb{B}$-operation is an "$\mathbb{A}$-homomorphism". A famous example is: if $\mathbb{G}$ is the theory of groups, $\mathbb{G}\otimes\mathbb{G}$ is the theory of abelian groups.

2. Methodological remarks and examples

Having briefly described some of the main tools of the functorial semantics point of view in general algebra, we now make some methodological remarks which this point of view suggests. First, many problems will take the forms: Characterize, in terms of $\mathcal{T}$, those $\mathbb{A}$ for which $\mathbb{A}^\flat$ has a given property stated in terms of $(\mathrm{Cat}, \mathcal{S})$, or characterize those $f \in \mathcal{T}$ for which $f^\flat$ has a given property; or for which f_* has a given property. Properties of $\mathbb{A}$ may be viewed as properties of $U_{\mathbb{A}}$ or of $F_{\mathbb{A}}$ and as such may have natural "relativizations" to properties of $f^\flat$ or f_*. Properties of diagrams in $\mathcal{T}$ may be "semantically" defined via arbitrary "mixtures" of the processes $f \rightsquigarrow f^\flat$, $g \rightsquigarrow g_*$, and algebraic structure from properties in $(\mathrm{Cat}, \mathcal{S})$, and direct descriptions in $\mathcal{T}$ of such properties of diagrams may be sought. Most of the solved and unsolved problems mentioned below are of this general sort. For example, light would be shed on many situations in algebra if one could give a computation entirely in terms of $\mathcal{T}$ of the algebraic structure of

$$\mathbb{G}^\flat \xrightarrow{g^\flat} \mathbb{M}^\flat \xrightarrow{f_*} \mathbb{R}^\flat \xrightarrow{U_{\mathbb{R}}} \mathcal{S}$$

for any given diagram

$$\mathbb{G} \xleftarrow{g} \mathbb{M} \xrightarrow{f} \mathbb{R}$$

in $\mathcal{T}$. A case in point is that where $\mathbb{G}$ = theory of groups, $\mathbb{M}$ = theory of monoids, $\mathbb{R}$ = theory of rings with g and f the obvious inclusions; what is sought in the example is in this case the full algebraic structure of group rings - this is a very "rich" theory, having *linear* "p-th power" unary operations for all p and more generally an n-ary *multilinear* operation for every element of the free group on n letters (e.g. convolution corresponds to the binary operation of group multiplication). Are these multilinear operations a generating set for the theory in question? Probably this case is simpler than the example in general, since it is equivalent to the structure of

$$\mathbb{G}^\flat \xrightarrow{U_\mathbb{G}} \mathcal{S} \xrightarrow{F_\mathbb{A}} \mathbb{A}^\flat \longrightarrow \mathcal{S}$$

where $\mathbb{A}$ is the theory of abelian groups, and $\mathbb{A}^\flat$ has a convenient tensor product.

Sometimes the problem is in the other direction: for example, the product $\mathbb{A} \times \mathbb{B}$ of course has an easy description in terms of $\mathcal{T}$, but a bit of computation is needed to deduce from general principles that $(\mathbb{A} \times \mathbb{B})^\flat$ consists of algebras which canonically split as sets into a product $X \times Y$, where X carries the structure of an $\mathbb{A}$-algebra and Y the structure of a $\mathbb{B}$-algebra).

A second general methodological remark is that the structure functor often yields much more information than the usual Galois connection of Birkhoff between classes of algebras of a given type and sets of equations, precisely because in many situations it is natural to change the type. Namely, a subcategory $\mathcal{X} \subseteq \mathbb{B}^\flat$ of an algebraic category (even a full one) may have an algebraic structure with more operations (as well as more equations) than $\mathbb{B}$, i.e. the induced morphism $\mathbb{B} \longrightarrow \mathbb{A}_\mathcal{X}$ may be non-surjective, where $\mathbb{A}_\mathcal{X}$ denotes the algebraic structure of $\mathcal{X} \longrightarrow \mathbb{B}^\flat \xrightarrow{U_\mathbb{B}} \mathcal{S}$. An obvious example is that in which $\mathbb{B}$ is the theory of monoids and $\mathcal{X}$ is the full subcategory consisting of those monoids in which every element has a two-sided inverse. Two other examples arise from subcategories of the algebraic category of commutative rings: the algebraic structure of the full category of fields includes the theory $\mathbb{R}_\theta$ generated by an additional unary operation θ subject to

$$\begin{aligned} 1^\theta &= 1 \\ (x \cdot y)^\theta &= x^\theta \cdot y^\theta \\ x^2 \cdot x^\theta &= x \\ (x^\theta)^\theta &= x \end{aligned}$$

and similarly the algebraic structure of the category of integral domains and monomorphisms includes the theory $\mathbb{R}_e$ generated by an additional operation e subject to

$$\begin{aligned} 0^e &= 0 \\ (x \cdot y)^e &= x^e \cdot y^e \\ (x^e)^e &= x^e \\ x^e \cdot x &= x \end{aligned}$$

The inclusion of fields in integral domains corresponds to the morphism

$$\mathbb{R}_e \longrightarrow \mathbb{R}_\theta$$

which, while the identity on the common subtheory $\mathbb{R}_c$ (= theory of commutative rings), takes e into the operation of $\mathbb{R}_\theta$ defined as follows

$$x^e \stackrel{\text{def}}{=} x \cdot x^\theta.$$

The third general methodological remark is that, within the doctrine of universal algebra, the "natural" domain of a construction used in some classical theorem may be in fact much larger than the domain for which the theorem itself can be proved. For example, the only $\mathbb{R}_e$-algebras which can be embedded in fields are integral domains, but the usual "field of fractions" construction is just the restriction of the adjoint functor $(\mathbb{R}_e \longrightarrow \mathbb{R}_\theta)_*$ whose domain is all of $\mathbb{R}_e^\flat$. To the same point, the usual construction of Clifford algebras is defined only for K-modules V equipped with a quadratic form $V \xrightarrow{q} K$; these pairs $\langle V, q\rangle$ do not form an algebraic category. But if we allow ourselves to consider quadratic forms $V \longrightarrow S$ with values in arbitrary commutative K-algebras S, we can

(1) define the underlying set to be $V \times S$ and find that these generalized quadratic forms do constitute an algebraic category,

(2) extend the Clifford algebra construction to this domain and find that there it is entirely a matter if algebraic functors and their adjoints (for this certain idempotent operations have to be introduced, as below).

Certain constructions which have the form of algebraic functors composed with adjoints to algebraic functors may also be interpretable along the line of the foregoing remark. For example, the "natural" domain of the group ring construction might be said to be the larger category of all monoids, for there it becomes simply the adjoint of an algebraic functor. Similar in this respect is the construction of the exterior algebra of a module, whose usual universal property is not that of a single left adjoint, but does allow interpretation in terms of the composition of algebraic functors and the adjoint of an algebraic functor:

$$\mathbb{A}^\flat \xrightarrow{f^\flat} \mathbb{A}_P^\flat \overset{g_*}{\dashrightarrow} \mathbb{R}_P^\flat \xrightarrow{h^\flat} \mathbb{R}^\flat$$

where $\mathbb{A}$ is the theory of K-modules, $\mathbb{A}_P$ is the theory of modules with an idempotent K-linear operator P, $\mathbb{R}$ is the theory of K-algebras, and $\mathbb{R}_P$ is the theory of K-algebras with an idempotent K-linear unary operation P satisfying the equation

$$(x^P)^2 = 0$$

(f, g, h being the obvious inclusions). Thus one might claim that the natural domain of the exterior algebra functor consists really of modules with given split submodules whose elements are destined to have square zero.

The problem mentioned earlier, of computing the structure of a composition: algebraic functor followed by an adjoint of an algebraic functor, is of relevance also in the above examples, since e.g. the natural anti-automorphism of Clifford algebras is an element of the structure theory of that functor, while composing the exterior algebra functor with the forgetful functor from Lie algebras or $K[x]$-modules and then taking algebraic structure should yield exterior differentiation and determinant, respectively, as operations in appropriate algebraic theories.

It is obvious and well-known that the constructions of tensor algebras, symmetric algebras, universal enveloping algebras of Lie algebras, abelianization of groups, and of the group engendered by a monoid are all of the form f_* for a suitable morphism $f \in \mathcal{T}$. Perhaps less well-known is the theory $\mathbb{M}_{(-)}$ of monoids equipped with a unary operation "minus" satisfying

$$\begin{aligned} -(-x) &= x \\ (-x)\cdot(-y) &= x\cdot y \end{aligned}$$

and the functor $\mathbb{M}^\flat_{(-)} \overset{f_*}{\dashrightarrow} \mathbb{R}^\flat$ associated to the obvious inclusion f of $\mathbb{M}_{(-)}$ into the theory of rings; this functor has the quaternions as one of its values, the eight quaternions $\{\pm 1, \pm i, \pm j, \pm k\}$ forming an $\mathbb{M}_{(-)}$-algebra. The quaternions also appear in another way, namely as a value of the Cayley-Dickson monad (triple) which is the composition of a certain algebraic functor with its adjoint and is defined on an appropriate algebraic category of non-associative (not even all alternative) algebras with involution.

An algebraic functor whose adjoint does not seem to have been investigated is the *Wronskian*, which assigns to each commutative algebra equipped with a derivation $x \rightsquigarrow x'$ the Lie algebra consisting of the same module with

$$[a,b] \overset{\text{def}}{=} a\cdot b' - a'\cdot b.$$

For example, is the adjunction always an embedding, giving an entirely different sort of "universal enveloping algebra" for a Lie algebra?

3. Solved problems

For the remainder of this paper we wish to discuss some problems exemplifying the canonical sort of the first methodological remark. Some semantically-defined subcategories of $\mathcal{T}$ admit not only simple descriptions entirely in terms of $\mathcal{T}$, but also can themselves be parameterized by single algebraic subcategories. Consider the full subcategory of $\mathcal{T}$ determined by those $\mathbb{A}$ for which $U_\mathbb{A}$ has a *right* adjoint (as well as the usual left adjoint $F_\mathbb{A}$). These $\mathbb{A}$ are easily seen to be characterized by the property that for each $n = 0, 1, 2, \ldots$ each $\mathbb{A}$-operation $A^n \longrightarrow A$ factors uniquely through one of the projections $\Pi_i^{(n)}$. Such *unary theories* are in fact parameterized by the full and faithful left adjoint of the "unary core" functor

$$\mathcal{T} \xrightarrow{Un} \mathbb{M}^\flat$$

where $\mathbb{M}$ is the theory of monoids with

$$\mathbb{M}(A^n, A) \cong \sum_{k=0}^{\infty} n^k$$

and where $(\mathbb{A})Un = \mathbb{A}(A,A)$ as a monoid. Thus we may also say that a theory "is a monoid" iff it is unary. Note that if we denote the left adjoint

$$\mathbb{M}^\flat \longrightarrow \mathcal{T}$$

to Un by $M \rightsquigarrow \overline{M}$, then we have

$$\overline{M_1 \times M_2} = \overline{M_1} \otimes \overline{M_2}$$

for any two monoids M_1, M_2.

Another algebraically parameterized subcategory of $\mathcal{T}$ consists of all $\mathbb{A}$ for which $\mathbb{A}^\flat$ is *abelian*. We often say that such a theory "is a ring", for it must necessarily be isomorphic to a value of the full and faithful functor

$$\mathbb{R}^\flat \xrightarrow{\text{Mat}} \mathcal{T}$$

which assigns to each ring R the category Mat_R whose morphisms are all the finite rectangular matrices with entries from R (i.e. the algebraic theory of R-modules). Here $\mathbb{R}(A^n, A) \cong Z[x_1, \ldots, x_n]$ = the set of polynomials with integer coefficients in n noncommuting indeterminates. The functor Mat commutes with the Kronecker product operations defined in the two categories, and has a *left* adjoint given by $\mathbb{A} \rightsquigarrow Z \otimes \mathbb{A}$ where we now mean by Z the theory corresponding to the ring Z (i.e. the theory of abelian groups). Note that while a *quotient theory* of a ring is always a ring, e.g. the theory of convex sets (consisting of all stochastic matrices) is a *subtheory* of a ring which is not a ring.

Since $\mathbb{A} \longrightarrow Z \otimes \mathbb{A}$ canonically, we have the adjoint functor

$$\mathbb{A}^\flat \dashrightarrow (Z \otimes \mathbb{A})^\flat$$

from the category of $\mathbb{A}$-algebras to the canonically associated abelian category, and for each $\mathbb{A}$-algebra X an adjunction morphism $X \longrightarrow \overline{X}$ if we denote by $\overline{X}$ the associated $Z \otimes \mathbb{A}$-module. The kernel of this adjunction morphism may be denoted by $[X, X]$, suggesting notions of *solvability* for algebras over any theory $\mathbb{A}$, which do in fact agree with the usual notions for $\mathbb{A}$ = theory of groups, theory of Lie algebras, or theory of unitless associative algebras. Sometimes $[X, X]$ may actually be the empty set; for example, if $\mathbb{A}$ is a monoid, X is a set on which the monoid acts, then $X \longrightarrow \overline{X}$ is the embedding of X into the free abelian group generated by X (equipped with the induced action of $\mathbb{A}$).

The composition

$$\mathbb{M}^\flat \xrightarrow{\subset} \mathcal{T} \xrightarrow{Z \otimes ()} \mathbb{R}^\flat$$

is another way of defining the monoid ring; more generally, for any theory $\mathbb{A}$ and monoid M, $\mathbb{A} \otimes M$ is the theory of $\mathbb{A}$-algebras which are equipped with an action of M by $\mathbb{A}$-endomorphisms. In fact, thinking of theories as generalized rings often suggests a natural extension of concepts or constructions ordinarily defined only for rings to arbitrary theories. For example consider fractions: the category whose objects are theory-morphisms $M \longrightarrow \mathbb{A}$, M any monoid, $\mathbb{A}$ any theory, admits a reflection to the subcategory in which M is a group, constructed by first ignoring $\mathbb{A}$ and forming the algebraic adjoint, and then taking a pushout in $\mathcal{T}$.

Part of the intrinsic characterization of those $\mathbb{A}$ which are rings is of course the condition that for each n, A^n is the n-fold *coproduct* (as well as product) of A in $\mathbb{A}$ (in fact this alone is characteristic of semi-rings). Another condition which some theories $\mathbb{A}$ satisfy is that A^n is the 2^n*-fold* coproduct of A; such theories turn out to be parameterized by the algebraic category of Boolean algebras.

One of the famous solved problems of our canonical type is: Which theories $\mathbb{A}$ are such that in $\mathbb{A}^\flat$, every *reflexive* subalgebra $Y \subseteq X \times X$ is actually a congruence relation? The answer is: those for which there exists at least one $\mathcal{T}$-morphism $\mathbb{B}_3 \longrightarrow \mathbb{A}$, where

$$\mathbb{B}_3 = \mathbb{I}_3/E$$

is the theory generated by one ternary operation θ satisfying the two equations E:

$$\begin{aligned} \langle x, x, z\rangle\theta &= z \\ \langle x, z, z\rangle\theta &= x. \end{aligned}$$

For example, if $\mathbb{A} = \mathbb{G}$, the theory of groups, one could define such a morphism by

$$\langle x, y, z\rangle\theta \stackrel{\text{def}}{=} x \cdot y^{-1} \cdot z.$$

Also $\mathbb{R}$, $\mathrm{Mat}(R)$ for any ring R, the theory of Lie algebras, as well as certain theories of loops or lattices, share with $\mathbb{G}$ the property described.

Also by now well-known, but apparently more recently considered, is the problem: For which $\mathbb{A}$ does $\mathbb{A}^\flat$ have a closed (autonomous) structure with respect to the standard underlying set functor $U_{\mathbb{A}}$? The answer is: the *commutative* $\mathbb{A}$, meaning those for which every operation is also a homomorphism. Since a monoid or ring is commutative as a monoid or ring iff it is commutative as a theory, one is not surprised to note that in the category of commutative theories, the coproduct is the Kronecker product.

Less classical, but more trivial, is the question: for which $\mathbb{A}$ is the trivial algebra 1 a good generator for $\mathbb{A}^\flat$? The answer is: the *affine* $\mathbb{A}$, meaning those for which

$$A \xrightarrow{\text{diag}} A^n \xrightarrow{\varphi} A$$

is the identity morphism for every n-ary $\mathbb{A}$-operation φ and for every $n = 0, 1, 2, \dots$. Being "equationally defined", the inclusion (of affine theories into all) clearly has a left adjoint, but more interesting seems to be the right adjoint which happens to exist; this assigns to any $\mathbb{A}$ the *subtheory* $\mathrm{Aff}(\mathbb{A})$ consisting of all (tuples of) those φ which do satisfy the above condition. Noting the first four letters of the word "coreflection", we call $\mathrm{Aff}(\mathbb{A})$ the *affine core* of $\mathbb{A}$. The term "affine" was suggested by the fact that

$$\mathbb{R}^\flat \xrightarrow{\text{Mat}} \mathcal{T} \xrightarrow{\text{Aff}} \mathcal{T}_{\text{Aff}} \xrightarrow{\subset} \mathcal{T}$$

assigns to each ring its *theory of affine modules.*

4. Unsolved problems

We now list some semantically-defined subcategories $\mathcal{C}$ of $\mathcal{T}$ for which good characterizations in terms of $\mathcal{T}$ alone do not seem to be known. They will be presented in relativized form, so that *none* of them are full subcategories of $\mathcal{T}$ but *all* of them contain all the isomorphisms of $\mathcal{T}$. With each such relativized problem $\mathcal{C}$ there is a corresponding "absolute" problem: namely to find those $\mathbb{A}$ such that the morphism f from the initial theory to $\mathbb{A}$ belongs to the class $\mathcal{C}$. We simply list the condition that arbitrary $\mathbb{B} \xrightarrow{f} \mathbb{A}$ belong to $\mathcal{C}$ in each case:

(1) $f^\flat$ takes epimorphisms in $\mathbb{A}^\flat$ into epimorphisms in $\mathbb{B}^\flat$. The corresponding absolute question is: for which $\mathbb{A}^\flat$ are epimorphisms surjective? so that for example $\mathbb{G}$ has the property while $\mathbb{M}$ does not.

(2) $f^\flat$ has a right adjoint (as well as the usual left adjoint). Note that this second category (2) is included in the category (1) defined above, and that the corresponding absolute question was answered with "unary theories". However, the present relative question is definitely more general than just morphisms of unary theories since every morphism between rings is included in category (2) as is the inclusion $\mathbb{M} \longrightarrow \mathbb{G}$ (recall the "group of units"). Since the right adjoint of $f^\flat$ would have to be represented by $f \cdot X_1$, X_1 being the free $\mathbb{A}$-algebra on one generator, the question is related to the more general one of computing, for any f, the algebraic structure of the set-valued functor $\mathbb{B} \longrightarrow \mathcal{S}$ so represented.

(3) f_* is right adjoint to $f^\flat$. This very strong condition obviously implies (2). We call the f satisfying (3) *Frobenius* morphisms since a typical example is a morphism in $\mathcal{T}$ of the form $K \xrightarrow{f} R$ where K is a commutative ring, R is a ring, and f makes R a Frobenius K-algebra. It does not seem to be known if there are any examples in $\mathcal{T}$ of Frobenius morphisms which are not ring morphisms. In the context of triples in arbitrary categories, a characterization in terms of the existence of a "nonsingular associative quadratic form" can be given, but it is not clear what the abstract form of this condition means when restricted back to theories (unless they are rings).

(4) $f^\flat$ takes finitely generated $\mathbb{A}$-algebras into finitely generated $\mathbb{B}$-algebras. A thorough understanding of this category would imply the solution of the restricted Kurosh and restricted Burnside problems as special cases. In fact the restricted Burnside problem belongs to the absolute case of the question, taking $\mathbb{A} = \mathbb{G}_r$ = theory of groups of exponent r, and the restricted Kurosh problem to the case relative to $\mathbb{B}$ = ground field, taking $\mathbb{A}$ = theory of algebras satisfying a given polynomial identity.

(5) The adjunction morphism $Y \longrightarrow f \cdot (Y f_*)$ is monomorphic for all $\mathbb{B}$-algebras Y. This category includes the f defined by the Lie bracket, but not that defined by the Jordan bracket, into the theory of associative algebras over a field. Since when applied to finitely generated free algebras, the adjunction reduces to f itself, it is clear that all f in category (5) are necessarily monomorphisms themselves. But this is not

sufficient, as the morphism $Z \xrightarrow{f} Q$ from the ring of integers to the ring of rationals shows (apply f_* to an abelian group with torsion). Linton has suggested that the universally mono-morphic f in $\mathcal{T}$ may coincide with category (5).

(6) $f^\flat$ reflects the existence of quasi-sections; i.e. for any $\mathbb{A}$-homomorphism h, if there is a $\mathbb{B}$-homomorphism g with $(h)f \cdot g \cdot (h)f = (h)f$, there is an $\mathbb{A}$-homomorphism $\overline{g}$ with $h \cdot \overline{g} \cdot h = h$. The absolute form of this condition applies to a ring $\mathbb{A}$ if it is semi-simple Artinian. Since simplicity, chain conditions, etc. have sense in the category $\mathcal{T}$, it would be interesting if subcategory (6) could be characterized in these terms.

5. Completion problems

Finally, various completion processes on the category of theories are suggested by the adjointness of the structure functor. For example, consider the inclusion $\mathcal{S}_{\text{fin}} \longrightarrow \mathcal{S}$ of finite sets into all sets. Pulling back and composing with this functor yields an adjoint pair

$$(\mathrm{Cat}, \mathcal{S}) \rightleftarrows (\mathrm{Cat}, \mathcal{S}_{\text{fin}})$$

which, when composed with the semantics-structure adjoint pair, yields a triple (monad) on the category $\mathcal{T}$. This triple assigns to each theory $\mathbb{A}$ the algebraic theory $\overline{\mathbb{A}}$ consisting of all operations naturally definable on the *finite* $\mathbb{A}$ algebras. For example $\overline{\mathbb{G}}$ is the (finitary part of) the theory of profinite groups.

Burnside's general problem suggests a different, "unary" completion $\tilde{\mathbb{A}}$ for a theory $\mathbb{A}$, namely let $\tilde{\mathbb{A}}$ be the structure of (the underlying set functor of) the category of those $\mathbb{A}$-algebras which are finitely generated and in which each single element x generates a finite sub-algebra F_x. Since this category is a union and semantics is an adjoint we have

$$\tilde{\mathbb{A}} \cong \varprojlim_F \mathbb{A}_F$$

where F ranges over finite sets of finite cyclic A-algebras, since structure is an adjoint. Note that this completion is not functorial unless we restrict ourselves to category (4). Since every finite A-algebra satisfies the two finiteness conditions above, one obtains a morphism.

$$\tilde{\mathbb{A}} \longrightarrow \overline{\mathbb{A}}$$

the study of which reflects one form of a generalized Burnside problem.

The functorial completion can also be done relative to a given theory $\mathbb{B}_\circ$ by using finitely generated or finitely presented $\mathbb{B}_\circ$-algebras, and considering theories $\mathbb{A}$ equipped with $\mathbb{B}_\circ \longrightarrow \mathbb{A}$. For example, with $B_\circ$ = a field K, the completion of $\mathbb{A} = K[x]$ is the full natural operational calculus $\overline{K[x]}$ for arbitrary operators on finite-dimensional spaces; explicitly this ring consists of all functions θ assigning to every square matrix a over K another a^θ of the same size, such that for every suitable rectangular matrix b and square a_1, a_2

$$a_1 b = b a_2 \Rightarrow a_1^\theta b = b a_2^\theta.$$

If K is the field of complex numbers, one has

$$\begin{tikzcd}
 & & & K[[x]] \\
K[x] \arrow[r] & \mathcal{E}(K) \arrow[r] & \overline{K[x]} \arrow[ur] \arrow[dr] & \\
 & & & K^K
\end{tikzcd}$$

where $\mathcal{E}(K)$ is the ring of entire functions and $K[[x]]$ the ring of formal power series. (Formal power series also arise as algebraic structure, by restricting to the subcategory where the action of x is nilpotent.) The ring $\overline{K[x]}$ would seem to have a possible role in "formal analytic geometry"; it has over formal power series the considerable advantage that substitution is always defined, so that formal endomorphisms of the formal line would be composable. This monoid is extended to $\overline{\mathbb{A}}$, (the dual of) a category of formal maps of formal spaces of all dimensions by applying the structure-semantical completion process over finite-dimensional K-vector spaces to the theory $\mathbb{A}$ of commutative K-algebras.

References

M. Barr (1969), Composite cotriples and derived functors. In B. Eckmann, ed. Seminar on Triples and Categorical Homology Theory, Lecture Notes in Mathematicas **80**, 336–356, Springer-Verlag.

M. Barr and J.M. Beck (1966), Acyclic Models and Triples. Proceedings of the 1965 La Jolla Conference on Categorical Algebra, Springer-Verlag.

J. Bénabou (1966), Structures algébriques dans les catégories. Thèse de l'Université de Paris.

S. Eilenberg and J.C. Moore (1965), Adjoint Functors and Triples. Ill. J. Math. **9**, 381-389.

S. Eilenberg and Wright (1967), Automata in General Algebras. Notes distributed in conjunction with the Colloquium Lectures given at the Seventy-Second Summer Meeting of AMS.

P.J. Freyd (1966), Algebra-valued functors in general and tensor products in particular. Colloq. Math. **14**, 89-106.

F.W. Lawvere (1963), Functorial Semantics of Algebraic Theories. Proc. Nat. Acad. Sci. (USA) **50**, 869-872.

F.E.J. Linton (1966a), Some Aspects of Equational Categories. Proceedings of the 1965 La Jolla Conference on Categorical Algebra, Springer-Verlag.

F.E.J. Linton (1966b) Autonomous Equational Categories. J. Math. & Mechanics **15**, 637-642.

University at Buffalo, Department of Mathematics
244 Mathematics Building, Buffalo, New York 14260-2900

Email: `wlawvere@buffalo.edu`

This article may be accessed from `http://www.tac.mta.ca/tac/reprints` or by anonymous ftp at `ftp://ftp.tac.mta.ca/pub/tac/html/tac/reprints/articles/5/tr5.{dvi,ps}`

Reprints in Theory and Applications of Categories, No. 1, 2002, pp. 1–37.

METRIC SPACES, GENERALIZED LOGIC, AND CLOSED CATEGORIES

F. WILLIAM LAWVERE

Author Commentary: Enriched Categories in the Logic of Geometry and Analysis

Because parts of the following 1973 article have been suggestive to workers in several areas, the editors of TAC have kindly proposed to make it available in the present form. The idea on which it is based can be developed considerably further, as initiated in the 1986 article [1]. In the second part of this brief introduction I will summarize, for those familiar with the theory of enriched categories, some of the more promising of these further developments and possibilities, including suggestions coming from the modern theory of metric spaces which have not yet been elaborated categorically. (The 1973 and 1986 articles had also a didactic purpose, and so include a detailed introduction to the theory of enriched categories itself.)

While listening to a 1967 lecture of Richard Swan, which included a discussion of the relative codimension of pairs of subvarieties, I noticed the analogy between the triangle inequality and a categorical composition law. Later I saw that Hausdorff had mentioned the analogy between metric spaces and posets. The poset analogy is by itself perhaps not sufficient to suggest a whole system of constructions and theorems appropriate for metric spaces, but the categorical connection is! This connection is more fruitful than a mere analogy, because it provides a sequence of mathematical theorems, so that enriched category theory can suggest new directions of research in metric space theory and conversely, unusual for two subjects so old (1966 and 1906 respectively).

The closed interval $[0, \infty]$ of real numbers as objects, $\geq$ as maps, $+$ as "tensor" and truncated subtraction as adjoint "hom", constitute a bona fide example of a complete, symmetric, monoidal closed category V. For any such V there is the rich system of constructions and theorems (worked out by Eilenberg and Kelly, Day, and others) involving

- V-valued categories;

- V-strong functors;

Originally published as: Metric spaces, generalized logic, and closed categories, *Rendiconti del seminario matématico e fisico di Milano, XLIII (1973), 135-166*

Received by the editors 2002-04-01 and, in revised form, 2002-06-24.

Transmitted by Michael Barr. Reprint published on 2002-09-1.

2000 Mathematics Subject Classification: 18D20.

Key words and phrases: Metric spaces, enriched categories, logic.

- V-natural transformations as an object of V and hence
- V-functor categories, λ-transformation and double dualization;
- the Cayley-Hausdorff-Yoneda lemma;
- free V-categories (generated by V-graphs) whose adjointness expresses Dedekind-Peano recursion via an objective geometric series;
- V-valued "relations" or bimodules and their convolution;
- Kan quantifiers which give extensions in particular well-defined situations.

All of these turn out to specialize, for the stated example V, to important results and constructions for metric spaces:

- V-functors are Lip_1 maps;
- the V-natural hom of two such turns out to be their sup-distance;
- some embedding and extension theorems of the Polish school and of MacShane follow from the general Yoneda-Kan lore;
- profunctor composition is Bellman-Fenchel convolution.

It is important that, in general, metric spaces satisfy only the two axioms for a V-category; the evidence is compelling that the usually-given more restrictive definition was too hastily fixed. Note that our V itself is quite non-symmetric (from now on we use "symmetry" of A to mean that $A^{\mathrm{op}} = A$ in an object-preserving way, rather than to mean that the tensor is commutative). A metric space can always be symmetrized (by one of two methods, $+$ and max), but it is often better to delay that until the last stage of a calculation, because the natural asymmetry carries considerable information and also because the main rules for passing from one stage of a calculation to the next are adjointness relations. Even though examples from pure geometry are symmetric, many constructions arising in dynamics as well as many constructions in analysis lead naturally to non-symmetric metric spaces; for example, the Hausdorff metric on subsets of a metric space, or the usual distance between subsets of a probability space (usually discussed only in their symmetrized form) yield in particular an "approximate inclusion" partial order upon applying the standard monoidal functor (represented by the unit 0) from V to the cartesian-closed poset V_0 of truth-values.

Likewise, metric spaces need not have all distances finite, but one can (when appropriate) restrict consideration to those points which have finite distance to a given part. The coproducts in V-cat naturally have infinite distance between points in different summands; infinite distance corresponds to a vacuous hom-set in the case of ordinary categories.

The relation between truth-values V_0 and distances V may be understood, informally, in terms of the cost or work required to transform or move one point to another, and formally in terms of three adjoint monoidal functors. The inclusion of V_0 into V interprets "true" as zero distance or "already achieved", but interprets "false" as infinite distance or "unattainably expensive". This inclusion has a right adjoint which transforms any V-category into the underlying V_0-category (or poset) as mentioned above; but it also has a left-adjoint π_0 which is also monoidal and hence transforms a V-category into a different V_0-category ordered this time by finiteness of cost.

The symbol π_0 for the "finiteness" monoidal truth functor was chosen by strict analogy with the relation between simplicial sets and abstract sets, where the connected components concept is indeed the left adjoint of a left adjoint and moreover monoidal (with respect to cartesian product). Following the Hurewicz tradition we can define for any V-category A a corresponding homotopy category $\pi_0 A$ and in particular for V-functors f_1 and f_2 from A to B a corresponding homotopy value $\pi_0(B^A)(f_1, f_2)$.

The content of the resulting "homotopy theory" is largely about rotations: Defining translations to mean automorphisms at finite distance from the identity, one sees that these form a normal sub-group with a recognizable quotient group in the case of Euclidian space, where the "search light effect" shows that these are indeed only the translations.

Closed subsets of a metric space have been identified with certain Lip_1 functions on the whole space in both constructive analysis and variational calculus. More precisely, every V presheaf is a V colimit of representables, but among those are the mere V_0 colimits (infima) and indeed between those the sup metric is the same as the (non-symmetric) Hausdorff metric. Every presheaf has in particular its zero set (and more generally sub-level sets).

The interpretation of presheaves as refined subsets suggests the following further construction: By definition, representables A are V-adequate in presheaves $V^{A^{\mathrm{op}}}$, but how co-adequate are they? That is measured by the monad which is the composite of the Isbell conjugacies to and from $(V^A)^{\mathrm{op}}$, i.e. double dualization into the identity bi-module. The action of this monad on subsets is the formation of the closed convex hull (at least in case A itself is a closed convex subset of a suitably reflexive Banach space).

Although habitually the diameter is used as a measure of the size of a subset, for many purposes a more appropriate (because more functorial) quantity is the radius, defined as follows: The direct limit functor from V^A to V exists and in fact is just inf; given any presheaf F on A

$$\mathrm{rad}(F) = \mathrm{inf}^{\mathrm{op}}(F^*)$$

where $(\ )^*$ is Isbell conjugation.

V-cat is itself a monoidal closed category and moreover the monoidal endo-functors of V act on it, giving rise to a fibered category whose maps include Lip_λ functions for various λ. But these monoidal functors are considerably more general than multiplication by a constant so that Lipschitz continuity, as well as Hausdorff dimension, admit much more refined measurements. Note that the square root, but not squaring, is monoidal. This suggests a whole family of monoidal structures on V-cat interpolating between the standard one given by V and the cartesian product at the other extreme (probably it already occurred to analysts that an equation like

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$$

indicates that the parameterization chosen may not have been the most natural.) Thus, contrary to the apology in the introduction of the 1973 paper, it appears that the unique role of the Pythagorean tensor does indeed have expression strictly in terms of the enriched category structure.

The geodesic re-metrization G is the co-monad on V-cat, resulting from a general categorical idea: Namely, it measures the adequacy of a particular family of objects, in this case a family of intervals parameterized by V itself.

Recent work of Gromov and others suggests that V-cat itself has a useful structure as a V-category. Presumably the Gromov distance between two metric spaces A and B is the symmetrization of a more refined invariant obtained as their Hausdorff distance in an extremal metric on $A + B$; but the latter metrics are determined by bimodules, which is a standard V notion!

References

1 Lawvere, F.W., Taking Categories Seriously, Revista Colombiana de Matemáticas **XX**, 147-178, 1986.

2 Gromov, M., Metric Structures for Riemannian and non-Riemannian Spaces, Progress in Mathematics, **152**, Birkhäuser, 1999.

Department of Mathematics, SUNY at Buffalo, Buffalo, NY
Email: wlawvere@buffalo.edu

This article may be accessed from http://www.tac.mta.ca/tac/reprints or by anonymous ftp at
ftp://ftp.tac.mta.ca/pub/tac/html/reprints/articles/1/tr1.pdf

F. WILLIAM LAWVERE
dell'Università di Perugia

METRIC SPACES, GENERALIZED LOGIC, AND CLOSED CATEGORIES

(Conferenza tenuta il 30 marzo 1973) *

SUNTO. — In questo articolo viene rigorosamente sviluppata l'analogia fra $\mathrm{dist}(a, b) + \mathrm{dist}(b, c) \geq \mathrm{dist}(a, c)$ e $\hom(A, B) \otimes \hom(B, C) \rightarrow \hom(A, C)$, giungendo a numerosi risultati generali sugli spazi metrici, come conseguenza di una « logica pura generalizzata » i cui « valori di verità » sono scelti in una arbitraria categoria chiusa.

INTRODUCTION.

It is a banality that all mathematical structures of a given kind constitute the objects of a category; the sequence: elements/structures/categories thus has led some people to attempt to characterize the philosophical significance of the theory of categories as that of a « third level of abstraction ». But the theory of categories actually penetrates much more deeply than that attempted characterizaion would suggest toward summing up the essence of mathematics. The kinds of structures which actually arise in the practice of geometry and analysis are far from being « arbitrary », and indeed in this paper we will investigate a particular case of the way in which logic should be specialized to take account of this experience of non-arbitrariness, as concentrated in the thesis that *fundamental* structures are themselves categories. Two cases of this thesis have been known for 30 years; an ordered set (often called poset) is a category in which for any ordered pair of objects there is at most one morphism from the first to the second, while a group is a category in which there is just one object and in which every morphism is a isomorphism. That the thesis has non-vacuous implications for these two cases follows from the facts that when the general idea

* Pervenuta in Redazione il 16 novembre 1973.

of functor between categories is restricted to the special categories (that is to the posets, respectively to the groups) it agrees with the correct idea of morphism between these fundamental structures (i.e. with order-preserving map, respectively group homomorphism) and that a functor from one of the special categories to some category (e.g. that of vector spaces) is itself an important structure (a direct system of vector spaces, respectively a linear group-representation; moreover the correct morphism between *these* structures are then just the natural transformations). Two further cases of the thesis have been developed in the past 10 years: functorial semantics, in which categories with special properties are identified as theories and special functors as interpretations or models, and the theory of topoi, in which certain categories correspond to (usefully generalized) topological spaces.

By taking account of a certain natural generalization of category theory within itself, namely the consideration of strong categories whose hom-functors take their values in a given « closed category » $\mathcal{V}$ (not necessarily in the category $\mathcal{S}$ of abstract sets), we will show below that it is possible to regard a metric space as a (strong) category and that moreover by specializing the constructions and theorems of general category theory we can deduce a large part of *general* metric space theory. The theory of closed categories (and of strong categories valued in them) was originally developed to deal with more complicated examples such as compactly generated topological spaces, Banach spaces, differential graded modules, etc., and moreover some of the publications on the subject seem forbiddingly technical to the beginner. I hope that this article can also be read as an introduction to closed categories on the basis of the guiding example of metric spaces considered as strong categories valued in the closed category of nonnegative real quantities.

Since closed categories are just what is suficient to have a reasonable theory of strong categories, we consider some examples of the latter first in order to bring out the elementary nature of the analogy on which the present work is based. In a *metric space* X, we will denote by $X(a, b)$ the non-negative real quantity (we will allow the value ∞) of X-distance from the point a to the point b. Then the laws satisfied by X are greater-than relations

$$\begin{aligned} X(a,b) + X(b,c) &\geq X(a,c) \\ 0 &\geq X(a,a)\,. \end{aligned}$$

(see below for remarks on the possible non-symmetry of the metric).

In a *category* X, we will denote by $X(a, b)$ the abstract set of X-morphisms from the object a to the object b. Then the composition law and specification of identity morphisms for X are mappings

$$\begin{array}{rcl} X(a,b) \times X(b,c) & \to & X(a,c) \\ 1 & \to & X(a,a) \end{array}$$

(which are subject to associativity and unity axioms which may be expressed as commutative diagrams of mappings betwen abstract sets, using elementary properties of the cartesian product $\times$ of abstract sets and of the one-element set 1). In a *poset* X, we will denote by $X(a, b)$ the truth-value of the X-dominance of the element a over the element b. Then the transitivity and reflexivity laws for X are entailments

$$\begin{array}{rcl} X(a,b) \wedge X(b,c) & \vdash & X(a,c) \\ \text{true} & \vdash & X(a,a)\,. \end{array}$$

If K is a commutative ring, then in a K-*additive category* X we will denote by $X(a, b)$ the K-module of X-morphisms from the object a to the object b. Then the composition and identity laws for X are K-linear mappings

$$\begin{array}{rcl} X(a,b) \otimes X(b,c) & \to & X(a,c) \\ K & \to & X(a,a) \end{array}$$

(again subject to associativity and unity axioms which may be expressed by commutative diagrams of K-linear mappings of K-modules, using elementary properties of the K-tensor product $\otimes$ of K-modules and of the K-module K). Thus we are led to consider that a greater-than-or-equal-to relation between nonnegative real quantities is analogous to a K-linear mapping between K-modules, since both are morphisms of possible hom-values for categories X, but in two different closed categories $\mathcal{V}$. Similarly, the sum of quantities is analogous to the tensor product of modules, both because they play the same role in the structure of a $\mathcal{V}$-valued category and also because they satisfy the same « elementary properties » (within $\mathcal{V}$ itself) of functoriality (i.e. monotonicity in the case of quantities), of associativity and commutativity up to $\mathcal{V}$-isomorphism (i.e. up to equality in the case of quantities), and of having a unit object K (the zero quantity). We have the table

X	hom-values for X	composition law and identity law for X	domain of composition law for X	domain of identity law for X
metric space	nonnegative real quantities	$\geqq$	sum	zero
category	abstract sets	mapping	cartesian product	one element set
poset	truth values	entailment	conjunction	true
$\mathcal{V}$-valued category	objects in $\mathcal{V}$	morphism in $\mathcal{V}$	« tensor » product in $\mathcal{V}$	unit object K for tensor product in $\mathcal{V}$

The associativity and unity axioms which the composition law and identity law of a $\mathcal{V}$-based category X must satisfy are automatic in the case of metric spaces or posets since *all* diagrams in $\mathcal{V}$ commute if $\mathcal{V} =$ reals or $\mathcal{V} =$ truth-values.

None of our results in this paper will depend on the additional Frechet axioms:

$$\begin{array}{lccl} \text{if} \quad X(a, b) & = & 0 & \text{then} \quad a = b \\ X(a, b) & < & \infty & \\ X(a, b) & = & X(b, a)\,. & \end{array}$$

The first of these is not very natural from the categorical viewpoint since it corresponds to requiring that isomorphic objects are equal; passage to the quotient can be avoided (as it *must* be for ordinary categories) by employing *equivalence* (broader than isomorphism) of strong categories. Allowing ∞ among the quantities is precisely analogous to including the empty set among abstract sets, and it is done for similar reasons of completeness; a metric space can be analyzed as a structured system of metric spaces with finite distances by considering the equivalence relation defined by « $X(a, b) < \infty$ and $X(b, a) < \infty$ ». The non symmetry is the more serious generalization, and moreover occurs in many naturally arising examples, such as $X(a, b) =$ work required to get from a to b in mountainous region X. Also within analysis itself a naturally arising metric is often non-symmetric but traditionally symmetrized by one of the two procedures

$$X(a,b) \quad + \quad X(b,a)$$
$$\max\,(X(a,b) \quad , \quad X(b,a))$$

which in fact could be applied to any metric; we mention three common examples where this nonsymmetry exists. If M is any Boolean algebra equipped with an outer measure, then

$$M(a,b) \underset{\text{def}}{=} M(b-a)$$

(where $b - a = b \cap a'$ in the Boolean algebra and $M(c)$ denotes the measure of c) defines a metric space in which

$$0 \geq M(a,b) \quad \text{iff} \quad a \supseteq b$$

almost everywhere.

(For more about this example see the section below on closed functors).

If K is any convex set we may define a metric on it by

$$K(a,b) \quad = \quad \inf_{f:a \to b} \{-\log(\alpha_f)\}$$

where $f: a \to b$ means that $f \in K$ with a on the open segment from f to b, with α_f then denoting the $0 < \alpha < 1$ with $a = (1-\alpha)f + \alpha b$. (The proof of the triangle inequality follows from the fact that K is actually a « normed category », since if $g: b \to c$ we can define $fg: a \to c$ by

$$fg = \frac{1 - \alpha_f}{1 - \alpha_f\,\alpha_g}\, f + \frac{\alpha_f(1-\alpha_g)}{1 - \alpha_f\,\alpha_g}\, g,$$

note that this is associative and that $\alpha_{fg} = \alpha_f\,\alpha_g$, and adjoin identities formally). The notion of « normed category » can also be related to the (nonsymmetric) Hausdorff metric

$$2^X(A,B) \quad = \quad \sup_{a \in A} \quad \inf_{b \in B} X(a,b)$$

for subsets A, B of a metric space X: let $f: A \to B$ mean that f is any mapping from A to B and define

$$|f| \quad = \quad \sup_{a \in A} X(a, af)\,;$$

then by the axiom of choice

$$2^X(A, B) = \inf_{f: A \to B} |f|$$

and the fundamental property of a normed category

$$|f| + |g| \geq |fg|$$

leads to a proof of the triangle inequality. We will leave as an exercise for the reader to define a closed category $\mathcal{S}(\mathbf{R})$ such that « normed categories » are just $\mathcal{S}(\mathbf{R})$-valued categories and a « closed functor » $\inf: \mathcal{S}(\mathbf{R}) \to \mathbf{R}$ which induces the passage from any « normed category » to a metric space with the same objects. Another approach to the Hausdorff metric will be evident from the discussion below of the comprehension scheme. The canonical symmetrization procedure actually applies to categories valued in any given closed category $\mathcal{V}$; in general, symmetry of a $\mathcal{V}$-valued category X has to mean (not a property but) a given structure consisting of $\mathcal{V}$-isomorphisms

$$\sigma_{ab}: X(a, b) \longrightarrow X(b, a)$$

subject to suitable coherence axioms, and the first canonical procedure (specializing to the sum in the case of a metric space) is

$$\operatorname{sym}(X)(a, b) = X(a, b) \otimes X(b, a).$$

We still have not, in this introduction, touched on the property of closed categories for which they are called « closed ». Basically, closed categories are closed with respect to the operation of forming the hom of two objects, so that $\mathcal{V}$ itself is a fundamental example of a $\mathcal{V}$-category. The internal Hom (a, c) in $\mathcal{V}$ is related to the internal tensor product in $\mathcal{V}$ by adjointness so that there is a natural one-to-one correspondence between the $\mathcal{V}$-morphisms

$$b \longrightarrow \operatorname{Hom}(a, c)$$

and the $\mathcal{V}$-morphisms

$$a \otimes b \longrightarrow c.$$

Thus in the closed category $\mathbf{R} = [0; \infty]$ in which metric spaces are valued

$$\operatorname{Hom}(a, c) = \begin{cases} c - a & \textit{if} \quad c \geq a \\ 0 & \textit{if} \quad a \geq c \end{cases}$$

so that the Hom-$\otimes$ adjointness reduces (if we denote by « minus » this truncated subtraction) to the if-and-only-if

$$\frac{b \geq c - a}{a + b \geq c},$$

whereas in the closed category $\mathcal{S}$ in which ordinary categories are valued, it reduces to the rule of lambda conversion

$$\frac{B \rightarrow C^A}{A \times B \rightarrow C}$$

and in the closed category **2** in which posets are valued, it reduces essentially to modus ponens and the « deduction theorem »

$$\frac{\beta \vdash \alpha \Rightarrow \gamma}{\alpha \wedge \beta \vdash \gamma}$$

i.e. the internal Hom for truth-values is implication. We will also assume that our closed categories $\mathcal{V}$ have cartesian products and coproducts over arbitrary index sets as well as equalizers and co-equalizers, i.e. that they are complete and cocomplete. From general properties of adjoint functors it follows that $\otimes$ preserves direct limits in each variable separately, while

$$\begin{aligned} \mathrm{Hom}\ (\varinjlim a_i, c) &\cong \varprojlim \mathrm{Hom}\ (a_i, c) \\ \mathrm{Hom}\ (a, \varprojlim c_j) &\cong \varprojlim \mathrm{Hom}\ (a, c_j) \end{aligned}$$

For example in $\mathbf{R}$, $\varprojlim$ means sup and $\varinjlim$ means inf; in particular 0 is the empty $\varprojlim$ and ∞ is the empty $\varinjlim$ so that

$$\begin{aligned} a + \infty &= \infty \\ c - \infty &= 0 \\ 0 - a &= 0\,. \end{aligned}$$

This completeness of $\mathcal{V}$ itself will be necessary for most of our general constructions. We will not consider (categorical) completeness of $\mathcal{V}$-valued categories; but on the other hand we will see that completeness in the *Cauchy* sense does have a meaning for categories valued in any closed $\mathcal{V}$.

The general constructions of functor categories, free categories, left and right Kan extensions, and « discrete » fibrations reduce in the very special case $\mathcal{V} = \mathbf{2}$ to higher types, transitive closure of a relation, existential and universal quantification, and the principle of set abstraction. Like the operation of implication mentioned above, the position of all these constructions in the general scheme as well as their fundamental properties of transformation are uniquely determined by adjointness. Since logic signifies formal relationships which are general in character, we may more precisely identify logic with this scheme of interlocking adjoints and then observe that all of logic applies *directly* to structures valued in an arbitrary closed category $\mathcal{V}$ (not only to structures valued in truth-values). For example, in quantitative logic (the case $\mathcal{V} = \mathbf{R}$ with which we will be mainly concerned in this paper) the isometric embedding of a metric space into a space of functions with sup metric is the application of the exactly same principle of logic which in the case $\mathcal{V} = \mathbf{2}$ gives Dedekind's representation of a poset by order ideals, and the upper and lower integrals of a real function are precisely cases of the generalized universal and existential quantification. This vast generalization is quite compatible with the specialization of logic called for in the first paragraph: although for any given $\mathcal{V}$ we could consider « arbitrary » $\mathcal{V}$-valued structures, there is one type of such structure which is of first importance, namely for $\mathcal{V}$ respectively truth-values, quantities, abstract sets, abelian groups, the structure of respectively poset, metric space, category, additive category (a very natural generalization of ring) is the generally useful first approximation possible with $\mathcal{V}$-valued logic for analyzing various problems; it even seems that there is a natural second approximation, namely the structure of a « rigidly $\mathcal{V}$-closed $\mathcal{V}$-category » which in the four cases mentioned specializes *roughly* to partially ordered abelian group, normed abelian group, rigidly closed category, and (in the additive case) to a common generalization of the category of locally free modules on an algebraic space and the category of finite-dimensional representations of an algebraic group. Detailed discussion of this second approximation awaits further investigation, as does the extension of the results of this paper to an arbitrary base topos, i.e. the extension from the *constant* abstract sets which we consider here to continuously *variable* sets as «sets» of points for metric spaces, as index «sets» for products, etc.

Of course the really deep results in a subject depend very much on the particularity of that subject and the results we offer here in the field of metric spaces, taken individually, will justly appear shallow to those with any experience. Indeed for me the surprising aspect was that methods originally devised to deal with quite different fields of algebra and geometry could yield any significant known theorems at all (for example the known theorem on extension of Lipschitz maps in section three). But there are many particularities, for example the special role of quadratic metrics, which I do not see how could be a result of « generalized logic ».

1. - Closed categories, strong categories, strong functors, closed functors.

In this paper we use the term « closed category » as short for « bicomplete symmetric monoidal closed category ». That is, a closed category $\mathcal{V}$ has equalizers, coequalizers, set-indexed products and coproducts plus a « monoidal » structure, which is symmetric and closed. A monoidal structure is a given functor

$$\mathcal{V} \times \mathcal{V} \xrightarrow{\otimes} \mathcal{V}$$

which is symmetric

$$u \otimes v \cong v \otimes u$$

and associative

$$u \otimes (v \otimes w) \cong (u \otimes v) \otimes w$$

and has a unit object k satisfying

$$k \otimes v \cong v \cong v \otimes k \ ;$$

more precisely, the symmetry, associativity, and unitary isomorphisms are in general required to be given, natural, and «coherent», but since these conditions are automatic in the particular examples we consider, we do not emphasize them (MacLane). That the monoidal structure is closed means that we are further given a functor

$$\mathcal{V}^{op} \times \mathcal{V} \xrightarrow{\mathrm{Hom}} \mathcal{V}$$

and two natural transformations

$$u \xrightarrow{\lambda} \mathrm{Hom}\,(a, a \otimes u)$$

$$a \otimes \mathrm{Hom}\,(a, v) \xrightarrow{\epsilon} v$$

such that the two processses

$$a \otimes u \xrightarrow{f} v \rightsquigarrow u \xrightarrow{\lambda} \mathrm{Hom}\,(a, a \otimes u) \xrightarrow{\mathrm{Hom}\,(a,f)} \mathrm{Hom}\,(a, v)$$

$$u \xrightarrow{g} \mathrm{Hom}\,(a, v) \rightsquigarrow a \otimes u \xrightarrow{a \otimes g} a \otimes \mathrm{Hom}\,(a, v) \xrightarrow{\in} v$$

are inverse bijections,

$$\downarrow\uparrow \ \frac{a \otimes u \to v}{u \to \mathrm{Hom}\,(a, v)}\,.$$

Thus in particular there is a natural bijection

$$\frac{a \to v}{k \to \mathrm{Hom}\,(a, v)}$$

between $\mathcal{V}$-morphism $a \to v$ and $\mathcal{V}$-morphisms from the unit object to the object $\mathrm{Hom}(a, v)$ for any two objects a, v of $\mathcal{V}$, in terms of which the natural transformation $\in$ behaves as an « evaluation » morphism, and two successive applications of appropriate instances of evaluation corresponds to a $\mathcal{V}$-morphism

$$\mathrm{Hom}\,(a, b) \otimes \mathrm{Hom}\,(b, c) \to \mathrm{Hom}\,(a, c)$$

which represents a « strong » form of composition. There is moreover a natural double-dualization morphism

$$u \to \mathrm{Hom}\,(\mathrm{Hom}\,(u, b), b)$$

for any fixed object b.

The simplest non-trivial example **2** of a closed category, which serves as the values for classical logic, has two objects false and $k =$ true and three morphisms

$$\begin{array}{lcl} \text{false} & \vdash & \text{false} \\ \text{false} & \vdash & \text{true} \\ \text{true} & \vdash & \text{true} \end{array}$$

with conjunction and implication as the tensor and Hom. Conjunction is of course symmetric and associative and since

$$u \vdash a \Longrightarrow (a \wedge u)$$

$$a \wedge (a \Longrightarrow v) \vdash v$$

always, we have the bijection

$$\frac{a \wedge u \vdash v}{u \vdash (a \Longrightarrow v)}$$

i.e. for any assignments of false, true to the three variables a, u, v either both or neither of these entailments exist. In particular

$$\frac{a \vdash v}{\text{true} \vdash a \Longrightarrow v}$$

and

$$(a \Longrightarrow b) \wedge (b \Longrightarrow c) \vdash a \Longrightarrow c$$

and

$$u \vdash (u \Longrightarrow b) \Longrightarrow b \, .$$

A somewhat more general and « dual » example is as follows: let $\mathcal{M}$ be any system of subsets of a given finite set which contains the empty subset $k = \emptyset$ and which is closed with respect to finite unions and with respect to set-theoretic difference; for $a, v \in \mathcal{M}$ let $a \to v$ mean that $a \supseteq v$, and let $a \otimes u$ mean $a \cup u$. Then $\mathcal{M}$ is a closed category since

$$\frac{a \cup u \supseteq v}{u \supseteq v - a}$$

(Even if $\mathcal{M}$ is not finite, our condition of bicompleteness can often be achieved by considering e.g. measurable sets *modulo* null sets).

Our central example **R** has as objects all non-negative real quantities (including ∞), as morphisms $a \to v$ the greater-than-or-equal-to relations $a \geq v$, and as « tensor » $a \otimes u = a + u$ the sum of quantities. Then Hom is forced to be truncated subtraction

$$\operatorname{Hom}(a, v) = \begin{cases} v - a & v \geq a \\ 0 & a \geq v \end{cases}$$

(denoted simply as subtraction) where in particular

$$\begin{aligned} \infty - \infty &= 0 \\ \infty - a &= \infty \quad \mathit{if} \quad a \neq \infty \\ v - \infty &= 0 \end{aligned}$$

(bicompleteness holds since

$$\prod_{i \in I} v_i = \sup_{i \in I} v_i$$

$$\sum_{i \in I} a_i = \inf_{i \in I} a_i$$

and equalizers and coequalizers are trivial). We have in particular

$$\frac{a \geq v}{0 \geq v - a}$$

$$(b - a) + (c - b) \geq c - a$$

$$u \geq b - (b - u) .$$

The « original » example of a closed category is the category $\mathcal{S}$ of abstract sets and mappings with the (unique) closed structure in which tensor means cartesian product, $k = 1$ is a one-element set, and $\mathrm{Hom}(a, v)$ is the (abstract) set of (indices for) all the mappings $a \to v$.

Given a closed category $\mathcal{V}$, a strong category valued in $\mathcal{V}$, or simply a $\mathcal{V}$-category X is any structure consisting of a specified set of X-objects $a, b, c \ldots$ together with the assignment of an object $X(a, b)$ of $\mathcal{V}$ to every ordered pair of X-objects $\langle a, b \rangle$, the assignment of a $\mathcal{V}$-morphism

$$X(a, b) \otimes X(b, c) \xrightarrow{\mu_{abc}} X(a, c)$$

to every ordered triple $\langle a, b, c \rangle$ of X-objects, and the assignment of a $\mathcal{V}$-morphism

$$k \xrightarrow{\eta_a} X(a, a)$$

to every X-object a, subject to the conditions that the following diagrams in $\mathcal{V}$ always commute

$$\begin{tikzcd}
X(a, b) \otimes (X(b, c) \otimes X(c, d)) \arrow[rr, "X(a,b) \otimes \mu_{bcd}"] \arrow[d, "\mathcal{V} - \text{assoc} \cong"'] & & X(a,b) \otimes X(b, d) \arrow[dd, "\mu_{abd}"] \\
(X(a, b) \otimes X(b, c)) \otimes X(c, d) \arrow[d, "\mu_{abc} \otimes X(c, d)"'] & & \\
X(b, c) \otimes X(c, d) \arrow[rr, "\mu_{bcd}"'] & & X(a, d)
\end{tikzcd}$$

$$\begin{tikzcd}
X(a,b) \arrow[r, "\mathcal{V}-\text{unitary} \cong"] \arrow[d, "\mathcal{V}-\text{unitary} \cong"'] \arrow[ddrr, "\mathrm{id}"] & k \otimes X(a,b) \arrow[r, "\eta_a \otimes X(a,b)"] & X(a,a) \otimes X(a,b) \arrow[dd, "\mu_{abd}"] \\
X(a,b) \otimes k \arrow[d, "X(a,b) \otimes \eta_b"'] & & \\
X(a,b) \otimes X(b,b) \arrow[rr, "\mu_{abb}"'] & & X(a,b)
\end{tikzcd}$$

Thus an $\mathcal{S}$-category is just an ordinary (small) category-whereas if $\mathcal{V} = Ab$ the category of abelian groups then an Ab-category with one object is just an ordinary ring (and more general Ab-categories arise in linear algebra just as naturally as do rings, e.g. *all* finite rectangular matrices over a given field form an Ab-category with the natural numbers as objects). The associativity and identity axioms (i.e. the above commutative diagrams) are (like coherence) automatic in case $\mathcal{V}$ itself is a poset, and hence, as claimed in the introduction, a $\mathbf{2}$ -category is an arbitrary poset while an $\mathbf{R}$-category is an arbitrary (generalized) metric space.

Every $\mathcal{V}$-category X has an opposite X^{op} with the same objects and units but with

$$X^{op}(a, b) = X(b, a)$$

and

$$\begin{tikzcd}
X^{op}(a, b) \otimes X^{op}(b, c) \arrow[rr, "\mu^{op}_{abc}"] \arrow[d, equal] & & X^{op}(a, c) \arrow[d, equal] \\
X(b, a) \otimes X(c, b) \cong X(c, b) \otimes X(b, a) \arrow[rr, "\mu_{cba}"'] & & X(c, a)
\end{tikzcd}$$

If we define also

$$\mathrm{sym}\,(X)(a, b) = X(a, b) \otimes X(b, a)$$

with

$$\begin{tikzcd}
\mathrm{sym}\,(X)(a, b) \otimes \mathrm{sym}\,(X)(b, c) \arrow[r, "\mu^{\mathrm{sym}}_{abc}"] \arrow[d, equal] & \mathrm{sym}\,(X)(a, c) \arrow[dd, equal] \\
(X(a, b) \otimes X(b, a)) \otimes (X(b, c) \otimes X(c, b)) \arrow[d, "\cong"'] & \\
(X(a, b) \otimes X(b, c)) \otimes (X(c, b) \otimes X(b, a)) \arrow[r, "\mu_{abc} \otimes \mu_{cba}"'] & X(a, c) \otimes X(c, a)
\end{tikzcd}$$

and

$$\begin{tikzcd} k \arrow[r, "\eta_a^{\mathrm{sym}}"] \arrow[d, "\cong"'] & \mathrm{sym}(X)(a,a) \arrow[d, equal] \\ k \otimes k \arrow[r, "\eta_a \otimes \eta_a"] & X(a,a) \otimes X(a,a) \end{tikzcd}$$

then sym(X) is a $\mathcal{V}$-category isomorphic with its opposite in an object-preserving manner.

In case the objects of $\mathcal{V}$ can be indexed by a set (which by bicompletness actually forces $\mathcal{V}$ to have a poset as its underlying category) then $\mathcal{V}$ itself is an example of a $\mathcal{V}$-category (more generally any small part of $\mathcal{V}$ becomes a $\mathcal{V}$-category) by setting

$$\mathcal{V}(a, b) = \mathrm{Hom}(a, b)$$

Thus **2** is an example of a poset, while **R** is a (highly non-symmetric) example of a metric space. But sym(**R**) is the usual metric on the reals.

If X and Y are two $\mathcal{V}$-categories, then by a $\mathcal{V}$-functor $X \xrightarrow{f} Y$ is meant any structure consisting of mapping of the objects of X into the objects of Y together with an assignment of a $\mathcal{V}$-morphism

$$X(a, b) \xrightarrow{f_{ab}} Y(fa, fb)$$

to every ordered pair $\langle a, b \rangle$ of X-objects, subject to the commutativity of the following diagrams in $\mathcal{V}$.

$$\begin{tikzcd} X(a,b) \otimes X(b,c) \arrow[r, "\mu^X_{abc}"] \arrow[d, "f_{ab} \otimes f_{bc}"'] & X(a,c) \arrow[d, "f_{ac}"] \\ Y(fa,fb) \otimes Y(fb,fc) \arrow[r, "\mu^Y_{fa\,fb\,fc}"'] & Y(fa,fc) \end{tikzcd}$$

$$\begin{tikzcd} & X(a,a) \arrow[d] \\ k \arrow[ur, "\eta^X_a"] \arrow[dr, "\eta^Y_{fa}"'] & \\ & Y(fa,fa) \end{tikzcd}$$

Thus an $\mathcal{S}$-functor is just an ordinary functor between (small) categories, an Ab-functor is just an additive functor (e.g. a ring homomorphism in case X has only one object) while a **2** functor (satisfying

$$X(a, b) \vdash Y(fa, fb)\ )$$

is an arbitrary order-preserving mapping from the poset X to the poset Y, and an **R**-functor (satisfying

$$X(a, b) \geq Y(fa, fb))$$

is an arbitrary distance-decreasing map (Lipschitz map of Lipschitz constant ≤ 1) from the metric space X to the metric space Y.

If $X \xrightarrow{f} Y$ and $Y \xrightarrow{g} Z$ are $\mathcal{V}$-functors we can define $(gf)_{ab}$ as the composition

$$X(a, b) \xrightarrow{f_{ab}} Y(fa, fb) \xrightarrow{g_{fa,fb}}. Z(gfa, gfb)$$

in $\mathcal{V}$ to obtain a $\mathcal{V}$-functor $X \xrightarrow{gf} Z$ and thus a category $\mathcal{V}$-Cat whose objects and morphisms are all the $\mathcal{V}$-categories and $\mathcal{V}$-functors. In the next section we will see that $\mathcal{V}$-Cat (e.g. the category of metric spaces and distance-decreasing maps) is itself a closed category.

A morphism between closed categories $\mathcal{V} \xrightarrow{\Phi} \mathcal{W}$ is basically the concentrated expression of a process which takes every $\mathcal{V}$-category X into a $\mathcal{W}$-category ΦX with the same objects, every $\mathcal{V}$-functor f into a $\mathcal{W}$-functor Φf and in general interprets $\mathcal{V}$-valued category-theory as $\mathcal{W}$-valued category-theory; in view of many examples it is too restrictive to require that Φ « strictly » preserves the tensor product and Hom, so we adopt the following definition. A closed functor is a triple consisting of a functor $\mathcal{V} \xrightarrow{\Phi} \mathcal{W}$, a $\mathcal{W}$-morphism

$$k_{\mathcal{W}} \xrightarrow{\varphi_0} \Phi\,(k_{\mathcal{V}})$$

and a natural transformation

$$\Phi\,(u) \underset{\mathcal{W}}{\otimes} \Phi\,(v) \xrightarrow{\varphi_{uv}} \Phi\,(u \underset{\mathcal{V}}{\otimes} v)$$

subject to compatibility between themselves and with the given symmetry, associativity and unitary isomorphisms in $\mathcal{V}$ and $\mathcal{W}$ (Eilenberg-Kelly).

For example, any locally small closed category $\mathcal{V}$ has a canonical closed functor

$$\mathcal{V} \xrightarrow{V} \mathcal{S}$$

to the category of abstract sets defined by

$$V(u) = \text{set of all } \mathcal{V}\text{-morphisms } k \to u$$
$$1 \xrightarrow{\varphi_0} V(k) = \text{the identity } \mathcal{V}\text{-morphism } k \to k$$

with

$$V(u_1) \times V(u_2) \xrightarrow{\varphi_{u_1 u_2}} V(u_1 \otimes u_2)$$

the mapping taking any ordered pair $k \xrightarrow{f_1} u_1$, $k \xrightarrow{f_2} u_2$ into

$$k \xrightarrow{\cong} k \otimes k \xrightarrow{f_1 \otimes f_2} u_1 \otimes u_2 .$$

If $\mathcal{M}$ is a Boolean ring of sets made into a closed category with union as tensor as above then a closed functor

$$\mathcal{M} \xrightarrow{M} \mathbf{R}$$

is any order-preserving real-valued function on $\mathcal{M}$ satisfying

$$0 = M(\emptyset)$$
$$M(a) + M(b) \geq M(a \cup b)$$

i.e. an « outer measure ». The closed functors $\mathbf{R} \xrightarrow{\lambda} \mathbf{R}$ are just the « subadditive » order-preserving functions, i.e.

$$u_1 \geq u_2 \Rightarrow \lambda u_1 \geq \lambda u_2$$
$$0 = \lambda 0$$
$$\lambda u_1 + \lambda u_2 \geq \lambda (u_1 + u_2) .$$

If $\mathcal{V} \xrightarrow{\Phi} \mathcal{W}$ is any closed functor and X any $\mathcal{V}$-category, then

$$(\Phi X)(a, b) = \Phi(X(a, b))$$

$$\begin{array}{ccc}
(\Phi X)(a,b) \underset{\mathcal{W}}{\otimes} (\Phi X)(b,c) & \xrightarrow{(\Phi\,\mu)_{abc}} & (\Phi X)(a,c) \\
\downarrow \varphi & & \| \\
\Phi\,(X(a,b) \underset{\mathcal{V}}{\otimes} X(b,c)) & \xrightarrow[\Phi(\mu_{abc})]{} & \Phi\,(X(a,c))
\end{array}$$

$$\begin{tikzcd}
k_{\mathcal{W}} \arrow[rr, dashed, "(\Phi\eta)_a"] \arrow[dr, "\varphi_0"'] & & (\Phi X)(a,a) \\
& \Phi\,(k_{\mathcal{V}}) \arrow[ur, "\Phi\,(\eta_a)"'] &
\end{tikzcd}$$

defines a $\mathcal{W}$-category ΦX with the same set of objects as X. Moreover if $X \xrightarrow{f} Y$ is any $\mathcal{V}$-functor by defining Φf to be the same mapping as f on objects and $(\Phi f)_{ab} = \Phi(f_{ab})$ we obtain a $\mathcal{W}$ — functor $\Phi X \to \Phi Y$, and moreover this process is compatible with composition of $\mathcal{V}$-functors (and more). For example, any $\mathcal{V}$-category X has an underlying $\mathcal{S}$-category VX because of the closed functor $\mathcal{V} \xrightarrow{V} \mathcal{S}$.

In particular if $\mathcal{V}$ is small and $\mathcal{V} \xrightarrow{\Phi} \mathcal{W}$ is a closed functor, then $\Phi\ \mathcal{V}$ is a $\mathcal{W}$-category. As an example, any Boolean ring of sets $\mathcal{M}$ equipped with an outer measure $\mathcal{M} \xrightarrow{M} \mathbf{R}$ becomes a metric space.

If $\mathbf{R} \xrightarrow{\lambda} \mathbf{R}$ is a closed functor X and Y are metric spaces, then an $\mathbf{R}$-functor

$$\lambda\ X \xrightarrow{f} Y$$

is just a function satisfying

$$\lambda\,(X(a,b)) \geq Y(fa, fb)\,.$$

Since multiplication by a given constant is subadditive and thus a closed functor, we see that the study of (a very natural generalization of) arbitrary Lipschitz mappings $X \to Y$ is naturally incorporated into the functorial set-up.

A special class of closed categories are those in which the tensor product is actually the categorical product, which amounts to

$$\frac{u \longrightarrow v_1 \otimes v_2}{u \longrightarrow v_1\,,\ u \longrightarrow v_2}$$

further « rule of inference of propositional Logic »; such closed categories are usually called cartesian closed. $\mathcal{S}$ and **2** are cartesian closed, while Ab and **R** are not. If a category admits a cartesian closed structure, then that structure is essentially unique. Ab does not admit a cartesian closed structure, but the underlying category of **R** does admit the cartesian closed structure $\mathbf{R}_{cart}$ in which

$$u \otimes v \quad = \quad \max\,(u\,,\,v)$$

$$\mathrm{Hom}\,(a, v) \; = \; \begin{cases} v & v > a \\ 0 & a \geq v \end{cases}$$

An $\mathbf{R}_{cart}$-category is just an *ultrametric space*, so that all of our general results may be reinterpreted as applying to ultrametric spaces. Moreover, the identity mapping $\mathbf{R}_{cart} \to \mathbf{R}$ is a closed functor (i.e. $a + v \geq \max(a, v)$) which induces the inclusion

$$\mathbf{R}_{cart}\text{-Cat} \to \mathbf{R}\text{-Cat}$$

of ultrametric spaces into all metric spaces.

2. - Functor Categories, Yoneda embedding, adequacy, comprehension scheme.

If A and X are two $\mathcal{V}$-categories then there is a $\mathcal{V}$-category $A \times X$ whose objects are the ordered pairs $\langle a, x \rangle$ of objects and whose hom-values are

$$(A \times X)\,(\langle a, x\rangle\,,\,\langle a', x'\rangle) \; = \; A\,(a\,,\,a') \times X\,(x,\,x')$$

the last being the cartesian product (e.g. max. in the case of **R**) in $\mathcal{V}$. But if $\mathcal{V}$ is not cartesian closed, there is another more important $\mathcal{V}$-category structure on the same objects, defined by

$$(A \otimes X)\,(\langle a, x\rangle, \langle a', x'\rangle) \; = \; A\,(a\,,\,a') \otimes X\,(x, x')\,.$$

This gives, for example, the l_1-style metric on the product of two metric spaces. The unit k for this tensor product has one object $*$ with $k(*\,,*) = k$.

This tensor product of $\mathcal{V}$-categories always has a « Hom » adjoint to it, making $\mathcal{V}$-cat itself into a closed category. This Hom always has a concrete interpretation in terms of « strong natural transformations »; we are going to denote it by Y^A, and its objects are just all the $\mathcal{V}$-functors $A \to Y$. Recall that in the usual $\mathcal{S}$-valued case, a natural transformation $f_1 \to f_2$ (where $A \xrightarrow{f_1, f_2} Y$) is a family of Y-morphisms indexed by the objects of A, more exactly, an element of the set $\prod_{a \in A} Y(f_1 a, f_2 a)$ subject to equations indexed by the elements of the $A(a, b)$. In the $\mathcal{V}$-valued case, we define directly (not the set of but) the $\mathcal{V}$-object of natural transformations $Y^A(f_1, f_2)$ by the equalizer in $\mathcal{V}$ of the two morphisms

$$\prod_a \rightrightarrows \prod_{a,b}$$

in

$$Y^A(f_1, f_2) \to \prod_{a \in A} Y(f_1 a, f_2 a) \rightrightarrows \prod_{a,b \in A} \mathrm{Hom}\,(A(a, b), Y(f_1 a, f_2 b))$$

whose constructions we leave to the reader. A $\mathcal{V}$-morphism $k \to Y^A(f_1, f_2)$ may then be identified with a strong natural transformation $f_1 \to f_2$. In case $\mathcal{V}$ is itself a poset, as with $\mathcal{V} = \mathbf{R}$, the two morphisms are already equal so that $Y^A(f_1, f_2) = \prod_{a \in A} Y(f_1 a, f_2 a)$ in such a case. Since $\prod$ in $\mathbf{R}$ means sup, we deduce that

$$Y^A(f_1, f_2) = \sup_{a \in A} Y(f_1 a, f_2 a)$$

in the case of metric spaces, i.e. that the sup metric on the space of 1-Lipschitz maps is a special case of the general notion of $\mathcal{V}$-natural transformations. The reader should be able to verify that there is a bijection

$$\frac{A \otimes X \to Y}{X \to Y^A}$$

between the two indicated sets of $\mathcal{V}$-functors at least in the case $\mathcal{V} = \mathbf{R}$.

Of special interest is the case $\mathcal{V}\, A^{op}$, because of the existence of the basic Yoneda embedding

$$A \to \mathcal{V}^{A^{op}}$$

which is the $\mathcal{V}$-functor whose value at an object a of A is the $\mathcal{V}$-functor $A^{op} \to \mathcal{V}$ defined by

$$a' \rightsquigarrow A(a', a)$$

and which works on A-morphism objects precisely by use of A-composition. Saying that a $\mathcal{V}$-functor $A \xrightarrow{f} Y$ is $\mathcal{V}$-full-and-faithful iff each of the $\mathcal{V}$-morphisms

$$A(a, b) \xrightarrow{f_{ab}} Y(fa, fb)$$

is actually an isomorphism, we have the important *Yoneda Lemma* (Eilenberg-Kelley): The Yoneda embedding is, for any closed $\mathcal{V}$ and any $\mathcal{V}$-category A, a $\mathcal{V}$-full-and-faithful $\mathcal{V}$-functor.

As a simple example, note that for $\mathcal{V} = \mathbf{2}$, an order-preserving map $A^{op} \to \mathbf{2}$ is equivalent to an order-ideal in A, and the Yoneda embedding is simply Dedekind's representation of a poset by its principal ideals. In the case $\mathcal{V} = \mathcal{S}$ or $\mathcal{V} = k$-Modules, the Yoneda embedding is often (especially when A has one object) called the regular representation of A, and Yoneda's Lemma includes Cayley's theorem on representing an abstract group by transformations. For a metric space A and for each point a, the function assigning to any point a' its distance to a is a distance decreasing function, and Yoneda's Lemma states that assigning to each a the just-described function is an *isometric* embedding of A into the space $\mathbf{R}\ A^{op}$ where the last is equipped with the sup metric.

More generally, given any $\mathcal{V}$-functor $A \xrightarrow{i} X$, we can consider the Yoneda representation of X restricted to A, i.e. the composite $\mathcal{V}$-functor

$$X \to \mathcal{V}^{X^{op}} \xrightarrow{\mathcal{V}^{i^{op}}} \mathcal{V}^{A^{op}}$$

which assigns to each x the functor $A^{op} \to \mathcal{V}$ defined by $a' \leadsto X(ia', x)$. In case this restricted representation is still $\mathcal{V}$-full--and-faithful on all X, we say after Isbell that i is *$\mathcal{V}$-adequate,* or in case i is an inclusion that A is a $\mathcal{V}$-adequate subcategory of X. This concept is the basis of much representation theory of categories, especially for algebraic categories and topoi, since it often happens that a quite small category A is adequate in a quite large one X. (We hasten to point out that this paragraph is meaningful even when $\mathcal{V}$ is not small; for example $\mathcal{V}\ A^{op}$ can be interpreted in terms of « modules » as in the following section). For example, if $\mathcal{V}$ = abelian groups and A is any ring, then A (considered as a $\mathcal{V}$-category with only one object) is adequate in any category X of A-modules, no matter how large.

In particular, to say that a subspace A of a metric space X is adequate is to say that for any two points x_1, x_2 of X, the inequality

$$X(x_1, x_2) \geq \sup_{a \in A} [X(a, x_2) - X(a, x_1)]$$

is actually an equality, i.e. that for any $d > 0$ there exists $a \in A$ with

$$X(a, x_1) + X(x_1, x_2) \leq X(a, x_2) + d .$$

For example, the unit circle is adequate in the unit disc (this simple example was pointed out to me by Prof. Isbell). A more restricted notion is that of $\mathcal{V}$-density, by which we mean (here differing in terminology with some authors) that $i_* o i^* \cong 1_X$ in the sense of the next section on bimodules; this reduces in the case of metric spaces to the requirement that

$$X(x_1, x_2) = \inf_{a \in A} [X(x_1, ia) + X(ia, x_2)] .$$

Proposition. - If a distance-decreasing map $A \xrightarrow{i} X$ of metric spaces is $\mathbf{R}$-dense, then it is $\mathbf{R}$-adequate.

Proof. - We always have

$$X(a, x_1) + X(x_1, x_2) \leq X(a, x_1) + X(x_1, a) + X(a, x_2) .$$

But by taking $x_2 = x_1$ in the definition of density, we see that $X(a, x_1) + X(x_1, a) \leq d$ for suitable $a \in A$, as required.

Define a particular metric space A whose points are the natural numbers and for which

$$A(n, m) = \begin{cases} \infty & n \neq m \\ 0 & n = m \end{cases} .$$

Then $\mathbf{R}^{A^{op}}$ is just the space of all sequences of nonnegative reals, with the (nonsymmetric) sup metric. Say that a metric space X is *separable* iff there exists an $\mathbf{R}$-dense map $A \to X$. Then

Corollary. - Any separable metric space X can be isometrically embedded in the space of all sequences of non-negative reals with the sup metric.

To discuss the « comprehension scheme » we will limit ourselves to those closed categories $\mathcal{V}$ in which $K = 1$, i.e. in which the unit object for the tensor product is also the terminal object of $\mathcal{V}$; thus for this form of the comprehension scheme, $\mathcal{V}$ may be cartesian closed, e.g. $\mathcal{S}$ or **2**, but more generally e.g. **R** satisfies this condition, as does the category of so-called « affine modules » over a given commutative ring (but not the usual category of modules). This has the effect that for any $\mathcal{V}$-category E, there is a canonical « augmentation » $E(e, e') \to K$. Thus we may simply consider the category $\mathcal{V}$-Cat/B of all $\mathcal{V}$-categories equipped with with a $\mathcal{V}$--functor with codomain B, and compare it with the category $\mathcal{V}^B$ of all $\mathcal{V}$-functors with domain B and the fixed codomain $\mathcal{V}$. Namely, given any $E \xrightarrow{p} B$, we define $\varphi_p : B \to \mathcal{V}$ by the coequalizer diagram in $\mathcal{V}$

$$\sum_{e,e'} E(e, e') \otimes B(p(e'), b) \rightrightarrows \sum_{e \in E} B(p(e), b) \longrightarrow \varphi_p(b)$$

where one of the two morphisms in induced by the functor p followed by composition in B, while the other is induced by the augmentation. In the case $\mathcal{V} = \mathbf{R}$, this simply means that for any distance-decreasing map p from a metric space E to the fixed metric space B, we define a real-valued function on B by

$$\varphi_p(b) = \inf_e B(p(e), b)$$

i.e. the distance from the image of p to the variable point b.

In the case $\mathcal{V} = \mathbf{2}$, φ_p is the order-preserving map $B \to \mathbf{2}$ defined by

$$\varphi_p(b) = \text{true} \quad \text{iff } \exists\, e\, [p(e) \text{ dominates } b \text{ in } B]$$

In the case $\mathcal{V} = \mathcal{S}$, $\varphi_p(b)$ may be interpreted as the set of components of the category p/b whose objects are pairs e, β with $p(e) \xrightarrow{\beta} b$ in B and whose morphisms are morphisms $e \xrightarrow{\xi} e'$ in E such that $\beta = p(\xi)\, \beta'$ in B.

The « comprehension scheme » then refers to the right adjoint of the functor

$$\mathcal{V}\text{—}Cat/B \longrightarrow \mathcal{V}^B$$

defined by $p \rightsquigarrow \varphi_p$. Namely, given any $B\varphi \to \mathcal{V}$, define a category $\{B|\varphi\}$ whose objects are pairs $\langle b, x \rangle$ such that $k \xrightarrow{x} \varphi(b)$ in $\mathcal{V}$

and for which $\{B|\varphi\}$ $(\langle b, x\rangle, \langle b', x'\rangle)$ is defined by the following pullback diagram in $\mathcal{V}$

$$\begin{array}{ccc}
\{B \mid \varphi\}\ (\langle b, x\rangle, \langle b', x'\rangle) & \xrightarrow{\quad n \quad} & k \\
\pi\ xx' \downarrow & & \downarrow x' \\
B(b, b') \cong k \otimes B(b, b') & \xrightarrow[x \otimes id]{} \varphi(b) \otimes B(b, b') \longrightarrow & \varphi(b')
\end{array}$$

where $\varphi(b) \otimes B(b, b') \to \varphi(b')$ is the « action » of B on φ adjoint to the functoriality of φ [note that $\mathcal{V}$-functors with codomain $\mathcal{V}$ always have an equivalent interpretation as « right modules » over the domain category]. The morphism n is unique since we have assumed $k = 1$, but without that assumption it would provide the $\mathcal{V}$-category $\{B|\varphi\}$ with an « augmentation » structure in addition to the structural functor π. In the case $\mathcal{V} = \mathcal{S}$, $\{B|\varphi\}$ is sometimes called the category of all elements of φ, and the functor $\{B|\varphi\} \xrightarrow{\pi} B$ is the discrete fibration corresponding to the set-valued functor φ. In case $\mathcal{V} = \mathbf{2}$

$$\{B \mid \varphi\} = \{ b \in B \mid \varphi(b) = \text{true} \}$$

(accounting for the terminology « comprehension scheme ») and we have of course

$$\frac{Im(p) \subseteq \{B \mid \varphi\}}{\varphi_p \vdash \varphi}$$

But in the case $\mathcal{V} = \mathbf{R}$, we have for any « quantity-valued propositional function » φ on the metric space B that

$$\{ B \mid \varphi \} = \{ b \in B \mid 0 = \varphi(b)\}$$

and for any $E \xrightarrow{p} B$,

$$\frac{Im(p) \subseteq \{B \mid \varphi\}}{\inf B(p(e), b) \leq \varphi(b), \text{ all } b}.$$

To what extent are objects in one of the categories $\mathcal{V}$-Cat$/B$, $\mathcal{V}^B$ « equivalent » to objects in the other via this adjoint pair? For $\mathcal{V} = \mathcal{S}$, $\mathbf{2}$, $\mathbf{R}$ respectively, $E \xrightarrow{p} B$ must be respectively a discrete fibration, the inclusion of an order-ideal, the inclusion of a *closed* subset in order to have $p \cong \{|\varphi_p\}$. On the other side, for $\mathcal{V} = \mathbf{R}$ there are distance decreasing maps $B \to \mathbf{R}$ *not* of the form « distance from a certain closed set », but for $\mathcal{V} = \mathcal{S}$ or $\mathcal{V} = \mathbf{2}$ *every* $\varphi \in \mathcal{V}^B$ is of the form φ_p.

3. - Bimodules, Kan quantification, Cauchy completeness.

For any $\mathcal{V}$-category Y, we have the canonical Yoneda embedding $Y \to \mathcal{V}^{Y^{op}}$ which means in particular that we can consider that the concept of a $\mathcal{V}$-functor $X \to \mathcal{V}^{Y^{op}}$ is a generalization of the concept of a $\mathcal{V}$-functor $X \to Y$. Such a generalized $\mathcal{V}$-functor $X \mapsto Y$ is equivalent to a $\mathcal{V}$-functor $Y^{op} \otimes X \xrightarrow{\varphi} \mathcal{V}$, and may be considered as a « $\mathcal{V}$-valued relation » from X to Y, $\varphi(y, x)$ being the « truth-value of the φ-relatedness of y to x ». In particular, every $\mathcal{V}$-functor $X \xrightarrow{f} Y$ thus yields a $X \overset{f_*}{\mapsto} Y$ defined by

$$f_*(y, x) = Y(y, f(x))$$

but also yields $\quad Y \overset{f^*}{\mapsto} X$ defined by

$$f^*(x, y) = Y(f(x), y) .$$

We first give an alternate description, without recourse to the notion of $\mathcal{V}$-functor, of such $\mathcal{V}$-valued relations as *bimodules*. We are using, by the way, the notational convention that inside a $\mathcal{V}$-category composition is written from left to right, while composition of $\mathcal{V}$-functors and of bimodules (= « relations ») is written from right to left.

If X, Y are $\mathcal{V}$-categories, a *bimodule* $X \overset{\varphi}{\mapsto} Y$ (also called a right X, left Y-module) consists of a family $\varphi(y, x)$ of objects of $\mathcal{V}$ indexed by the objects of Y and X together with morphisms

$$Y(y', y) \otimes \varphi(y, x) \longrightarrow \varphi(y', x)$$

$$\varphi(y, x) \otimes X(x, x') \longrightarrow \varphi(y, x')$$

in $\mathcal{V}$ which behave as « actions » in the sense that the five axioms (commutative diagrams in $\mathcal{V}$) of X-unity, X-associativity, Y-unity, Y-associativity, and mixed associativity (= commutativity of X-action with Y-action) hold. Thus in case $\mathcal{V} = Ab$, this is just the usual notion of bimodule suitably extended to « rings » X and Y with more than one object. In case $\mathcal{V} = \mathcal{S}$, bimodules are sometimes called « profunctors » or « distributeurs ». In case $\mathcal{V} = \mathbf{R}$, a bimodule $X \overset{\varphi}{\mapsto} Y$ between two metric spaces is just a real-valued function on the product $Y \times X$ satisfying

$$\inf_y \, [Y(y', y) + \varphi(y, x)] \geq \varphi(y', x)$$

$$\inf_x \, [\varphi(y, x) + X(x, x')] \geq \varphi(y, x') .$$

The latter is clearly equivalent to putting a metric on the sum $X + Y$ extending the given metrics and having $d(x, y) = \infty$ for $x \in X$, $y \in Y$, and indeed this alternate mode of description works generally, recalling that the infinite quantity corresponds to the empty coproduct in $\mathcal{V}$.

If $X \overset{\varphi}{\mapsto} Y \overset{\psi}{\mapsto} Z$ are bimodules, then composition $X \overset{\psi \circ \varphi}{\mapsto} Z$ is defined by the following coequalizer diagram in $\mathcal{V}$

$$\sum_{y_1, y_2} \psi(z, y_1) \otimes Y(y_1, y_2) \otimes \varphi(y_2, x) \rightrightarrows \sum_{y \in Y} \psi(z, y) \otimes \varphi(y, x) \longrightarrow (\psi \circ \varphi)(z, x)$$

Thus in case $\mathcal{V}$ is itself a poset, $\psi \circ \varphi$ reduces to a « matrix product »

$$\sum_{y \in Y} \psi(z, y) \otimes \varphi(y, x)$$

where the sigma denotes coproduct in $\mathcal{V}$, but in general we take the quotient of the matrix product modulo the discrepancy between the two actions of Y in the middle, just as in the familiar case $\mathcal{V} = Ab$ where one often writes more explicitly

$$\psi \circ \varphi = \psi \underset{Y}{\otimes} \varphi .$$

In particular for $\mathcal{V} = \mathbf{2}$

$$(\psi \circ \varphi)(z, x) = \exists \, y \, [\psi(z, y) \wedge \varphi(y, x)]$$

is the usual relational product, while for $\mathcal{V} = \mathbf{R}$ we have

$$(\psi \circ \varphi)(z, x) = \inf \, [\psi(z, y) + \varphi(y, x)] .$$

It can be verified in general that if $X \xrightarrow{f} Y \xrightarrow{g} Z$ are $\mathcal{V}$-functors, then

$$(gf)_* \cong g_* \circ f_*$$

where isomorphism of bimodules has an obvious sense; also 1_X defined by

$$1_X(x, x') = X(x, x')$$

plays the role of identity with respect to composition with any bimodules (not only those of the form f_*). Further, the composition of bimodules is associative up to (coherent) isomorphism, but in general not commutative. Note that the endobimodules of the one-object category k are in correspondence with the objects of $\mathcal{V}$, with the composition of these reducing to the tensor product in $\mathcal{V}$.
The concept Hom, right adjoint to $\otimes$ in $\mathcal{V}$ has two extensions to mixed Hom of bimodules, right adjoint on opposite sides to the composition of bimodules. That is, given any three $\mathcal{V}$-categories X, Y, Z and a bimodule $X \overset{\varphi}{\mapsto} Z$, there are two universal problems: given any $X \overset{\alpha}{\mapsto} Y$, there is a $Y \mapsto Z$ denoted by $\mathrm{Hom}^X(\alpha, \varphi)$ which is « best » in the sense that for any bimodule $Y \overset{\beta}{\mapsto} Z$ there is a natural bijection of morphisms of bimodules

$$\frac{\beta \longrightarrow \mathrm{Hom}^X(\alpha, \varphi)}{\beta \circ \alpha \longrightarrow \varphi}$$

likewise for given β there is a « best α », denoted by $\mathrm{Hom}_Z(\beta, \varphi)$ satisfying

$$\frac{\alpha \longrightarrow \mathrm{Hom}_Z(\beta, \varphi)}{\beta \circ \alpha \longrightarrow \varphi}$$

for any α.

Explicitly, $\mathrm{Hom}^X(\alpha, \varphi)$ is defined by an equalizer diagram in $\mathcal{V}$

$$\mathrm{Hom}^X(\alpha, \varphi)(z, y) \longrightarrow \prod_{x \in X} \mathrm{Hom}(\alpha(y, x)\varphi, (z, x) \rightrightarrows \prod_{x_1, x_2} \mathrm{Hom}\,(X(x_1, x_2), \mathrm{Hom}\,(\alpha(y, x_1), \varphi(z, x_2)))$$

and similarly $\mathrm{Hom}_Z(\beta, \varphi)(y, x)$ is defined by an equalizer condition as a subobject of

$$\prod_{z \in Z} \mathrm{Hom}\,(\beta(z, y), \varphi(z, x)).$$

For *some* bimodules $X \overset{\alpha}{\mapsto} Y$, the operation of composing with α, $\beta \rightsquigarrow \beta \circ \alpha$, has not only the right adjoint $\varphi \rightsquigarrow \mathrm{Hom}^X(\alpha, \varphi)$ but has also a *left* adjoint. The typical such bimodule α is one of the form

$\alpha = f_*$, where $X \xrightarrow{f} Y$ is a $\mathcal{V}$-functor; presently we will also discuss the extent to which « all » such bimodules α are induced by functors f.

Lemma. - For any $\mathcal{V}$-functor $X \xrightarrow{f} Y$ and any $\mathcal{V}$-bimodule $Y \overset{\beta}{\mapsto} Z$, there is a natural isomorphism

$$\beta \circ f_* = \mathrm{Hom}^Y(f^*, \beta)$$

of bimodules $X \mapsto Z$.

Proof. - One verifies that both sides are naturally isomorphic to the bimodule whose typical component is

$$\beta(z, f(x)) .$$

Theorem. - For $\alpha = f_*$, composition $\beta \rightsquigarrow \beta \circ \alpha$ has as left adjoint the operation of composing with f^*

$$\varphi \rightsquigarrow \varphi \circ f^*$$

Proof. -

$$\frac{\dfrac{\varphi \circ f^* \to \beta}{\varphi \to \mathrm{Hom}^Y(f^*, \beta)}}{\varphi \to \beta \circ f_*}$$

Corollary. - (Kan quantification). For any $\mathcal{V}$-functor $X \xrightarrow{f} Y$, the functor « composition with f »

$$\mathcal{V}^Y \xrightarrow{\mathcal{V}^f} \mathcal{V}^X$$

has both left and right adjoints. On $\varphi \in \mathcal{V}^X$, the left adjoint gives as a result the functor $\varphi \circ f^*$, whose value at y is the quotient of

$$\sum_X \varphi(x) \otimes Y(f(x), y)$$

« modulo the distinction between the two actions of X in the middle », while the right adjoint gives as a result the functor $\mathrm{Hom}^X(f_*, \varphi)$ whose value at y is that subobject of

$$\prod_X \mathrm{Hom}(Y(y, f(x)), \varphi(x))$$

« on which the two actions of X agree ». In the special case that f

is $\mathcal{V}$-full-and-faithful, $\mathcal{V}^f$ is also full and faithful, or equivalently the two Kan quantifications are actually *extensions*, in the sense that

$$(\varphi \circ f^*) \circ f_* \cong \varphi$$
$$\mathrm{Hom}^X(f_*, \varphi) \circ f_* \cong \varphi$$

for any φ.

Proof. - The corollary is for the most part just a summary of the preceding discussion for the special case $Z = k$ (Indeed the corollary remains valid if $\mathcal{V}$ is replaced by $\mathcal{V}^{Z^{op}}$ for any (small) $\mathcal{V}$-category Z).

The assertion that, under the assumption that f is $\mathcal{V}$-full-and-faithful, first « extending » φ along f (in either of the two adjoint ways) and then « restricting » along f gives again φ, is proved in very much the same way as the Lemma above.

Corollary. - Let $X \xrightarrow{f} Y$ be an isometric embedding of a metric space X into a metric space Y. Then any Lipschitz function $X \xrightarrow{\varphi} \mathbf{R}$ can be extended to Y, with the same Lipschitz constant λ. In fact, among all extensions there is a largest one and a smallest one, given respectively by

$$\overline{\varphi}(y) = \inf_X \left[\varphi(x) + \lambda \; Y(f(x), y)\right]$$

$$\underline{\varphi}(y) = \sup_X \left[\varphi(x) - \lambda \; Y(y, f(x))\right].$$

Proof. - Apply the preceding discussion to the diagram

$$\begin{tikzcd}
\lambda X \arrow[rr, "\lambda f"] \arrow[dr, "\varphi"'] & & \lambda Y \arrow[dl, dashed, shift left] \arrow[dl, dashed, shift right] \\
& \mathcal{V} &
\end{tikzcd}$$

noting that λf as a function is the same as the inclusion f.

As an example of the last corollary, we could take for X the space of nonnegative step functions on a probability space S and for Y the space of all nonnegative functions with the natural sup metric, (so that $Y(y_2, y_1) = 0$ iff $y_1 \geq y_2$), and consider as φ the elementary integral; thus in general we might call y φ-integrable if $\overline{\varphi}(y) \cong \underline{\varphi}(y)$.

The (non standard) name Kan quantification was suggested by the case $\mathcal{V} = \mathbf{2}$, in which the adjointness rules reduce to the usual rules of inference for quantification, φ being thought of as property or relation, i.e.

$$(\varphi \circ f^*)(z, y) \equiv \exists\, x\, [\varphi(z, x) \wedge f(x) \geq y]$$

$$\mathrm{Hom}^X(f_*, \varphi)(z, y) \equiv \forall\, x\, [y \geq f(x) \Longrightarrow \varphi(z, x)]$$

But for other choices of $\mathcal{V}$, induced representations and relatively free universal algebras can be shown to arise as special cases of these two constructions.

The essential property of a bimodule of the form f_* is that there exists another bimodule f^* which is right adjoint to it, in the sense that there are morphisms

$$1_X \longrightarrow f^* \circ f_*$$

$$f_* \circ f^* \longrightarrow 1_Y$$

satisfying the usual two adjunction equations.

Proposition. - In order that a metric space Y be Cauchy-complete, it is necessary and sufficient that every adjoint pair of bimodules

$$X \underset{f^*}{\overset{f^*}{\rightleftarrows}} Y$$

be induced by a $\mathcal{V}$-functor $X \xrightarrow{f} Y$.

Proof. - It suffices to consider $X = 1$ and show that an adjoint pair of $\mathbf{R}$-valued bimodules $1 \leftrightarrows Y$ is essentially just a point in the completion of Y. But adjointness just means that

$$X(x, x') \geq \inf_y \; [f^*(x, y) + f^*(y, x')]$$

$$\inf_X \; [f_*(y, x) + f^*(x, y')] \geq Y(y, y')$$

in addition to the bimodule property for each of f_*, f^*. In case $X = 1$, we have then

$$0 = \inf_y \; [f^*(y) + f_*(y)]$$

$$f_*(y) + f^*(y') \geq Y(y, y')\,.$$

Thus for each n we can choose y_n satisfying, for example

$$f^*(y_n) + f_*(y_n) \leq \frac{1}{n}$$

and then

$$Y(y_n; y_m) \leq \frac{1}{n} + \frac{1}{m}$$

so that we have a Cauchy sequence, and any other choice y'_n satisfying the same condition would have

$$Y(y_n\, y'_n) \leq \frac{2}{n}$$

i.e. would be an equivalent Cauchy sequence. Conversely, any equivalence class of Cauchy sequences yields an adjoint pair of bimodules by the definition

$$f^*(y) = \lim_{n\to\infty} Y(y_n, y)$$

$$f_*(y) = \lim_{n\to\infty} Y(y, y_n)\,.$$

It can be shown that the suggested definition of « $\mathcal{V}$-Cauchy-completeness » means in case $\mathcal{V} = Ab$ that a « point of the completion » of an additive category Y is simply any finitely generated projective module over Y, while in case $\mathcal{V} = \mathcal{S}$, it means that Y is Cauchy-complete iff all idempotents in Y split in Y.

4. - Free $\mathcal{V}$-categories.

Defining a $\mathcal{V}$-graph to be a pair consisting of any set X and any $X \times X$-indexed family of objects of $\mathcal{V}$, and a morphism $\langle X, \gamma\rangle \xrightarrow{f} \langle Y, \delta\rangle$ to be any pair consisting of a mapping $X \to Y$ and any family

$$\gamma(x, x') \xrightarrow{f_{x,x'}} \delta(f(x), f(x'))$$

of morphisms in $\mathcal{V}$, we have an obvious forgetful functor

$$\mathcal{V}\text{-Cat} \to \mathcal{V}\text{-Graph}.$$

This forgetful functor has a left-adjoint called taking the free $\mathcal{V}$-category generated by a $\mathcal{V}$-graph. Explicitly, the free category has as objects the vertices of the generating graph, and as hom from x to x' has the coproduct in $\mathcal{V}$

$$\sum_{x_1 \ldots, x_n} \gamma(x, x_1) \otimes \gamma(x_1, x_2) \otimes \ldots \otimes (x_n, x')$$

over all finite sequences $x_1 \ldots x_n$ of vertices. For example with $\mathcal{V} = k$-modules, this formula contains the construction of the tensor algebra over a vector space, or with $\mathcal{V} = \mathcal{S}$, we have as a special case the word algebra (= free monoid) over a given set. For the case $\mathcal{V} = \mathbf{R}$, an $\mathbf{R}$-graph is just an arbitrary assignment γ of quantities to pairs of points, and in the « free metric space » over such, the distance from x to x' is

$$\inf_{x_1 \ldots, x_n} [\gamma(x, x_1) + \gamma(x_1, x_2) + \ldots\ldots + \gamma(x_n, x')]$$

the well-known « least-cost » distance.

The adjointness of the free $\mathcal{V}$-category construction contains the essence of the notion of recursion, especially when one considers it in relation with bimodules, where it leads for example to the iteration of endobimodules.

5. - Further remarks.

We already remarked that the composition of bimodules is similar to matrix multiplication, and indeed the analogy with linear algebra goes further. For example, if bimodules rather than $\mathcal{V}$ functors are considered as the « morphisms » between $\mathcal{V}$-categories, then $A^{op} \otimes Y$ plays the role of « Hom ». Either further developing that remark in the case $A = Y$, or proceding directly, it is natural to define, for any endobimodule $A \xrightarrow{\alpha} A$, $Tr(\alpha)$ to be the object of $\mathcal{V}$ defined by coequalizer

$$\sum_{a, b \in A} A(b, a) \otimes \alpha(a, b) \rightrightarrows \sum_{a \in A} \alpha(a, a) \rightarrow Tr(\alpha) .$$

The injections $\alpha(a, a) \rightarrow Tr(\alpha)$, denoted by tr_a, are a natural generalization of many examples of classical constructions such as trace of endomorphisms or Lefschitz numbers (when $\alpha = 1_A$) as has been verified by A. Kock. In case A is a metric space and $\alpha = f_*$ where f is a distance decreasing endomap,

$$Tr(\alpha) = \inf_a A(a, fa)$$

showing that the vanishing of the trace is related to the existence of fixed points. It seems likely that there may be theorems holding for more general $\mathcal{V}$, relating the trace of an endobimodule to an infinite iteration of it, which would extend the Banach fixed-point theorem.

SUMMARY. — The analogy between $\operatorname{dist}(a, b) + \operatorname{dist}(b, c) \geq \operatorname{dist}(a, c)$ and $\hom(A, B) \otimes \hom(B, C) \to \hom(A, C)$ is rigorously developed to display many general results about metric spaces as consequences of a « generalized pure logic » whose « truth-values » are taken in an arbitrary closed category.

REFERENCE

MACLANE S., *Categories for the Working Mathematician*. Springer 1972.

EILENBERG S. & KELLY G. M., *Closed Categories, in Proceedings of La Jolla Conference on Categorical Algebra*. Springer 1966.

BENABOU JEAN, *Les Distributeurs*, Rapport n° 33, janvier 1973, Seminaires de Mathématiques Pure, Institut de Mathématiques, Université Catholique de Louvain (multigraphed).

26 February 1998

Outline of Synthetic Differential Geometry

F. William Lawvere

[Initial results in Categorical Dynamics were proved in 1967 and presented in a series of three lectures at Chicago. Since that time a flourishing branch of it called Synthetic Differential Geometry has given rise to four excellent textbooks by Kock, Lavendhomme, Moerdijk & Reyes, and Bell. To help make this subject more widely known and to further encourage its application, I gave some talks in February 1998 in the Buffalo Geometry Seminar. The following outline with 7 appendices was distributed as seminar notes. I have made a few corrections now (Nov. 1998), such as switching some A's and B's. But I have added Appendix 8, concerning the interesting work circulated in October 1998 by Kock & Reyes, in which they verify my main claim and also make several further observations.]

A cartesian closed category is one having exponential or internal hom right adjoint to cartesian product:

There is a natural bijection

$$\frac{X \to Y^A}{A \times X \to Y}$$

between the indicated sets of maps, for all X. It follows that this bijection is mediated by natural transformations

$$X \xrightarrow{\lambda_X} (A \times X)^A$$

$$A \times Y^A \xrightarrow{\varepsilon_Y} Y$$

for each A. It also follows that the exponential is a contravariant functor of the given A; i.e. given a map $A \xrightarrow{\alpha} B$, there is an induced natural map in the opposite direction

$$Y^B \xrightarrow{Y^\alpha} Y^A$$

in the same category. Calling 'points of X' the figures of shape 1 in X, i.e. the maps $1 \to X$ where 1 is the terminal object, it follows in particular that the points of an exponential Y^A indeed do parameterize the maps $A \to Y$, and that ε_Y acts as 'evaluation', etc. (Recall the computer philosophy, according to which a stored program is merely a special sort of data.)

The induced maps may be seen as a special case of the fact that for any three objects there is in the category a 'composition' map

$$\mathrm{Y}^{\mathrm{X}} \times \mathrm{Z}^{\mathrm{Y}} \longrightarrow \mathrm{Z}^{\mathrm{X}}$$

However, cohesiveness and variability of the objects in general comes from the figures of more general shape than punctural. The category X/S, where S is a given object of X, has as objects the objects of X further structured by a given map to S, and as morphisms has commutative triangles over S in X. This is the usual geometric way of dealing with families of spaces parameterized by S, namely the spaces in the family are the fibers of the structural map; the forgetful functor $\mathit{X}/\mathrm{S} \to \mathit{X}$ takes the *total* space of the family. Note that the identity map 1_{S} is the terminal object of X/S so that a 'point' in the latter category is 'the locus of a moving point' in the view of the original X. There is the functor

$$\mathit{X} \xrightarrow{(\ )_{\mathrm{S}}} \mathit{X}/\mathrm{S},$$

assigning to each X the 'constant family' $\mathrm{X} \times \mathrm{S} \xrightarrow{proj} \mathrm{S}$ (all of whose fibers are X); this functor is not full, since even between two constant families of spaces there are usually many non-constant S-parameterized families of maps. Indeed, it is easily seen that there is a bijection

$$\mathit{X}\ (\mathrm{S},\mathrm{Y}^{\mathrm{A}}) = (\mathit{X}/\mathrm{S})\ (\mathrm{A}_{\mathrm{S}}, \mathrm{Y}_{\mathrm{S}})$$

between hom sets, i.e. that the S-figures in Y^{A} are just the maps 'from A to Y', but in X/S.

A locally cartesian closed category X is one in which each of the categories X/A is cartesian closed; since products in X/A are just the fiber products over A in X, that is easily seen to be equivalent to the requirement that for each map $\mathrm{A} \xrightarrow{\alpha} \mathrm{B}$, the pullback functor $\alpha\ : \mathit{X}/\mathrm{B} \to \mathit{X}/\mathrm{A}$ has a right adjoint, often denoted by $\prod_{\mathrm{A}/\mathrm{B}}$ when α is understood.

Spaces of intensive (contravariant) or extensive (covariant) variable quantities on X are seen to partake of the same kind of cohesion/variation as the domain spaces X by using the cartesian closed structure. Thus for a given rig R in $\mathcal{X}$, $R^X = \mathcal{F}(X)$ is another rig *in* $\mathcal{X}$ (i.e. continuous, smooth, bornological, etc.). Using the action of the multiplicative monoid R on R^X. one can carve out from R^{R^X} (using an equalizer) the subobject $\mathrm{Hom}_R(R^X,R)$ $= \mathcal{M}(X)$ which plays the role of 'distributions of compact support' on X. (In the smooth $\mathcal{X}$, the homogeneous maps are usually automatically linear.)

The above method of obtaining the infinite-dimensional spaces of analysis from the finite-dimensional ones of geometry, is in principle well known. Implicit for 300 years and explicit for at least 30 is also (see my 1968 paper on 'diagonal arguments' or Michor and Kriegl's recent AMS book, but also see Fox, Brown, Kan, Steenrod much earlier for the sequentially continuous and simplicial cases) the fact that even these infinite-dimensional spaces W are determined by their categories $\mathcal{X}_{\mathrm{fin}}/W$ of finite-dimensional figures. For example, if S is a 1-dimensional segment, then a figure of shape S in W could also be called a *path*, and in fact many spaces are essentially path-determined. The maps in the category of small figures in W are commutative triangles in $\mathcal{X}$

$$\begin{tikzcd} S' \arrow[rr] \arrow[dr] & & S \arrow[dl] \\ & W & \end{tikzcd}$$

and determine all significant 'incidence relations' among figures. The adequacy of small figures in W means that a map W --------> R is determined by the functor it induces from the category of small figures in W to the category of small figures in R, for any space R. The mathematically reasonable condition on any proposed determination of 'small spaces vs. general spaces' requires this adequacy for all W.

But exponentiation also provides, in another manner, the way to determine the small (finite dimensional) spaces from the infinitesimal ones, the essential goal of differential geometry and continuum physics ! This is done by postulating that among the finite dimensional spaces there are some 'amazingly tiny' ones which in a suitable sense determine everything. (Although papers in algebraic geometry refer to these as 'zero-dimensional', I prefer to reserve that term for the much more restricted class of nearly discrete objects, and consider that the a.t.o.m.s (=amazingly tiny object models) have an 'infinitesimal dimension' in a sense which can be made precise.) All the cited books do this via a postulated rig object R (for coordinatizing a line) which is assumed to have enough nilpotent elements: Defining $D_n \rightarrow R \overset{(\ )^{n+1}}{\underset{0}{\rightrightarrows}} R$ to be subobject of R such that a figure x in R belongs to D_n iff $x^{n+1} = 0$ (and $D = D_1$ for short), the key axiom is that the map $R^{n+1} \longrightarrow R^{D_n}$ (which parameterizes functions representable by n-th degree polynomials) is an isomorphism. Identifying D_1-figures as tangent vectors and hence $X^{D_1} \longrightarrow X$ as the tangent bundle of any space X, we thus have in particular that the tangent bundle of the line is 2-dimensional, relative to R. Since

$$(Y^A)^{D_1} = (Y^{D_1})^A$$

(we bypass Banach manifolds, etc. etc.) to arrive at the tangent bundle of any function space, e.g. those where Y = R, A = 5 or $A = R^5$.

Since the tangent bundle is thus a *representable* functor, a vector field $X \rightarrow X^{D_1}$ on X can equivalently be viewed as an infinitesimal action $D_1 \times X \rightarrow X$ (by the basic transformation of cartesian closure); such always automatically induces a derivation (in the Leibniz sense) of the function algebra R^X. Note that there is automatically a *space* Vect(X) of all vector fields on X, with its own intrinsic cohesion and variation, but that there is also the (cartesian closed) category Vect(*X*) in which an object is a space equipped with a given vector field and in which a morphism is an infinitesimally equivariant map. Restriction along the inclusion $D_1 \rightarrow R$ induces a functor

$$\text{Flows}\,(\mathcal{X}) \longrightarrow \text{Vect}\,(\mathcal{X})$$

whose adjoints help to organize thinking about actually solving the ODE implicit in a given vector field. (Here a flow means an action of the additive monoid R).

Actually, D_1 is a *coordinatized* bit of time, $D_1 \approx T_1$, where the only structure T_1 has is a unique point $1 \xrightarrow{0} T_1$. The axiom above implies that ${T_1}^{T_1}$ contains R as a canonical retract and that the multiplication in R comes from the canonical monoid structure on ${T_1}^{T_1}$ that any space of endomorphisms has. Each D_n thus consists of those retardations whose n+1 iterate is the point in the instant. Thus one can actually reconstruct R and its algebraic structure from the purely geometrical data of a cartesian closed category with a given pointed object T_1 .

For any pointed object T, T^T is canonically bipointed for there are the names of both the identity map and the constant map. These points are both in the object R, now defined to be the part of T^T which consists of 0-preserving endomaps. An important axiom is that the bipointed object R be a connected object. Now we see that the finite dimensional spaces R^n are all retracts of function spaces of 'infinitesimal' objects.

But in what sense are these a.t.o.m.s T really tiny? There are several answers, some based on the intrinsic Heyting logic of subobjects, etc. A strong condition (and amazingly, all this is actually concretely realizable) is that the tangent bundle functor have a further right adjoint:

$$\frac{X \longrightarrow Y^{1/T}}{X^T \longrightarrow Y}$$

This permits representation of differential forms as merely 'functions' $X \longrightarrow R^{1/D}$ valued in a bigger fixed rig. It also permits to show that the category of *2nd order* ODE's is also locally cartesian closed (even a topos), as is briefly seen as follows:

Given a map $A \xrightarrow{\alpha} B$, one can consider the type of structure which for any X would consist of a given map

$$X^A \longrightarrow X^B$$

which is a section of α_X . This may be thought of as a "B-tuple of A-ary operations' on X subject to one defining identical equation. In case $A = 1$, such an $X \longrightarrow X^B$ is equivalent to a pointed action $B \times X \longrightarrow X$ and it is well-known that such actions constitute a new topos over which the original one is 'essential'. In case $B = D_1$, this topos consists of all first order ODE's. Even with $A = 1$ the property that B is an a.t.o.m. was important in Kock's proof of a theorem of Sophus Lie concerning the flow lines of an ODE.

If A is an a.t.o.m., such a section of α_X is equivalent to a map $X \longrightarrow X^{B/A}$ satisfying the one equation, which implies that the forgetful functor $\mathcal{X}^{\alpha} \xrightarrow{f^*} \mathcal{X}$ (from the category of prolongation structures being considered, back to the topos of spaces) thus has not only the 'free' left adjoint $f_!$ expected algebraically, but also 'cofree' right adjoint f_* which in turn implies in particular that $\mathcal{X}^{\alpha}$ is a topos.

The higher-order 'monoids' implicit in the above are not familar because in the category of abstract sets there are no a.t.o.m.s, except $A = 1$. The case of particular interest for physical dynamics is that where α is the inclusion$T_1 \longrightarrow T_2$ into a second-order instant (coordinatized by the inclusion $D_1 \longrightarrow D_2$ into the space of r in R for which $r^3 = 0$). A structure in $\mathcal{X}^{\alpha}$ is exactly a smooth prolongation of tangent vectors in a space X to 2nd order jets in X, i.e. a second order ODE or dynamical law on X. With the obvious definition of morphism, these therefore form a topos, by the above argument. Of course if X itself is a function space E^B, then among the ingredients for a prolongation operator are differential operators considered as maps; in other words, these ODE's include PDE's as a special case.

Actual motions which follow a given α-dynamical law in X are to be considered as morphisms $T \longrightarrow X$ in $\mathcal{X}^{\alpha}$ where T is equipped also with an α-prolongation law modeling time; for example, if T is coordinatized by R (usually not itself an a.t.o.m. if there is non-trivial Grothendieck topology in the picture) there is the obvious prolongation structure given by $R^D \xrightarrow{I} R^{D_2}$ for which

$$\tau(x,v)(t) = x + vt \qquad \text{for all } t^3 = 0$$

i.e. 'zero acceleration'.

APPENDIX 1

Figures as structure: Graphs

A simple example of the role of figures in deriving the internal structure of objects is provided by the ubiquity of (reflexive directed multi) graphs. If $1 \rightrightarrows I$ is any choice of two distinct points in any chosen object of *any* category $\mathcal{X}$, there is an induced functor, from $\mathcal{X}$, to the category of such graphs: each object X of $\mathcal{X}$ has not only the set $\mathcal{X}(1,X)$ of points, but the set $\mathcal{X}(I, X)$ of 'directed edges'; composition with the chosen points gives the needed 'source and target' structure. Any map $X \longrightarrow Y$ in the category $\mathcal{X}$ induces a map of these graphs which preserves this source and target structure, merely because of the associativity of composition. In case $\mathcal{X}$ is a topos and I is adequate, we can conclude that $\mathcal{X}$ *is* the category of graphs if moreover I has no endomorphisms other than the obvious three.

APPENDIX 2

Nilpotent Calculus

As engineers have implicitly known for centuries, the use of nilpotent quantities is very effective in reducing differential calculus to high school algebra. For example, we can prove that the length C(r) of the boundary of a disk is $2\pi r$ if we define π to be the area of a unit disk, as follows: By homogeneity the area of a disk of radius r is $A(r) = \pi r^2$ and the area of a perturbed disk of radius $r + h$ is $A(r+h) = \pi(r + h)^2$. Thus the area of a thin strip around the boundary circle is

$$A(r + h) - A(r) = \pi(2rh + h^2) = 2\pi rh$$

where in the last step we interpret 'thin' to mean $h^2 = 0$. But the same area is also equal to C(r) h, for all such h, and since there are enough of these nilpotents to permit cancelling them from universally-quantified equations, the result follows.

Similarly, one can prove for example that the electrical attraction of an infinitesimal dipole is inversely proportional to the *cube* of the distance from it.

APPENDIX 3

Higher (and lower) connectivities

Basic topological intuition can be applied synthetically in terms of the contrast between a category $\mathcal{X}$ of spaces with some cohesion and variability and a category of sets $\mathcal{S}$ with none (or qualitatively less). This applies not only to smooth spaces, but also to continuous or combinatorial ones. Namely, the points functor $\mathcal{X}(1, -)$ usually has a right adjoint (the inclusion of codiscrete spaces) and two successive left adjoints (the inclusion of discrete spaces and the functor 'set of components' which represents the attempts to map a given space to discrete spaces). An object is 'connected' iff its set of components is 1, i.e. iff its only maps to discrete spaces are constant. The study of the set of components of the space X^S of S-figures implies the homotopy and homology of X. (S ranging over reference spaces such as spheres and balls).

Indeed, the needed set-theory S is best *derived from* the geometry X by defining discrete spaces to be those C for which $C \longrightarrow C^S$ is an isomorphism for a few selected figure forms S which are considered to be connected and which at least have no nontrivial coproduct decompositions.

APPENDIX 4

Lie and the Discovery of a.t.o.m.s

Indeed, my 'fractional-exponent' definition of the a.t.o.m. property above was based on a key discovery of my student Anders Kock which I was able to correlate with work of my student Marta Bunge who had characterized presheaf toposes. Kock's discovery was about what was needed to prove a theorem of Sophus Lie concerning the space of flow lines of an ODE

$$D_1 \times X \rightrightarrows X \longrightarrow X/D_1$$

In computing such coequalizers, the fact that D_1 is internally projective is crucial; coupling that with the fact that D_1 is internally connected (i.e. that $(\ )^{D_1}$ preserves sums), one sees that $(\ )^{D_1}$ preserves colimits; then the special adjoint functor theorem (of one of my teachers, Peter Freyd) implies the existence of the fractional exponents.

APPENDIX 5

Time Speed-ups and Coordinates

The sense in which the non-commutative monoid T^T of time speed-up (for the case $T = T_1$ of a first-order microspace with one point) is 'just slightly bigger' than its commutative submonoid R of 0-preserving elements, can be seen quite clearly in terms of a choice $T \approx D_1$ of unit of time, where D_1 is the subspace of the coordinatized line R whose figures are those figures t of R for which $t^2 = 0$ with respect to the multiplication of R. For then the elements of T^T , in terms of the endomappings which they parameterize, are easily seen to be pairs $<a_0, a_1>$ in R for which

$$(a_0)^2 = 0, \text{ and } a_0\, a_1 = 0$$

Then, as usual, the elements a_i can be figures of any shape S, but we are using the multiplication of R ; the (S-family of) endomorphisms of T which the pairs parameterize are described by the formula

$$t \;¾\!\!\rightarrow\; a_0 + a_1 t$$

and the composition rule is the restriction to D of the substitution of one affine-linear transformation into another, which could be called 'the high-school monoid'. The inclusion R ¾→ T^T is the part where $a_0 = 0$, and a_1 arbitrary, but D itself is a subspace (not a submonoid) in a perpendicular direction where $a_1 = 0$. It is then easy to see that all the invertible elements of the monoid T^T (for $T = T_1$) belong to the commutative group of invertible elements of R. Disjoint from that, one can calculate, for example, the parts of T^T whose elements f satisfy $f \circ f = 0$, or $f \circ f \circ f = 0$.

Any tangent bundle X^T obviously has a natural right action of T^T , not only of the latter's submonoid R.

APPENDIX 6

Strong Adjoints versus General Exponents

Two endofunctors F and U of a cartesian closed category are strongly adjoint iff there is a natural isomorphism

$$Y^{FX} = (UY)^X$$

of function spaces, for all spaces X,Y in the category. Taking points of both sides of the above isomorphism we obtain a natural bijection

$$\frac{FX \to Y}{X \to UY}$$

of map-sets, the usual notion of adjointness. But the converse does not hold; there may not exist any way of making some given adjoint pair strong. For example, $A \times (\ )$ is strongly adjoint to $(\ )^A$ for any fixed space A, which is one of the laws of exponential algebra which is true in any cartesian closed category. But if T is an a.t.o.m., the adjointness of $(\ )^T$ to $(\ )^{1/T}$ is usually not strong, unless $T = 1$. We can measure 'how strong is it?' by considering the subcategory $\mathcal{S}_T$ (certainly including 1) of those test-figures C which perceive it so; that turns out to mean just that the canonical 'inclusion of constant figures (or zero tangent vectors)'

$$C \xrightarrow{\sim} C^T$$

is an isomorphism. This is a special case of the idea of defining a subcategory of 'discrete' spaces in terms of a whole category of spaces (as mentioned in Appendix 3), which might be called the 'microdiscrete' case; it turns out that the inclusion $\mathcal{S}_T \to \mathcal{X}$ has a right adjoint, which may be considered as one version of 'the space of points' of an arbitrary space, for indeed the composite of these two adjoints is a kind of '0-skeleton' endofunctor of $\mathcal{X}$, with $\mathrm{sk}_0(X) \to X$. Given such an adjoint pair, we can define

$$\mathcal{X}(A,Y) = \mathrm{sk}_0(Y^A)$$

and define a intermediate notion of $\mathcal{S}$-strong adjointness.

Actually, a strongly adjoint pair of endofunctors is determined by a single object $F1$, since for all Y

$$UY = Y^{(F1)}$$

as is seen by substituting $X = 1$ in the definition. But for a given $\mathcal{S} \rightleftarrows \mathcal{X}$, the $\mathcal{S}$-strong adjoint endofunctors in general form a larger category of 'exponents'. We consider these exponents as a category by considering as morphisms the natural transformations betwen the left-adjoints, but the exponents act as 'exponents' (also notationally) via the right adjoints; this acting is thus a right action.

$$E' \to E \text{ implies } Y^E \to Y^{E'}$$

$$(Y^{E1})^{E2} = Y^{E1 \cdot E2}$$

for exponents E where the multiplication of exponents means their composition considered as functors. Thus for a cartesian closed category $\mathcal{X}$,

X itself is canonically embedded in its category of exponents, and indeed

$$A_1 \cdot A_2 = A_1 \times A_2$$

for those special exponents which come from X. For any given a.t.o.m. T, the fractional exponent B/T is well-defined, for any object B; but for $T \neq 1$ the multiplication of these is no longer commutative. Still more general examples can be obtained by taking *colimits* of known exponents (which of course actually involves taking the opposite *limits* of the corresponding right-adjoints in the pairs). For example, the sum of two exponents can be defined via their right actions as

$$Y^{E1+E2} = Y^{E1} \times Y^{E2}$$

for all Y, which obviously will agree with the coproduct of spaces $A_1 + A_2$ for those 'integral' exponents coming from X.

Call an exponent $E=\langle E,F\rangle$ *strict* if F itself has a further left adjoint $E_!$; then F1=1 so that no nontrivial integral exponent is strict.

APPENDIX 7

Galilean Monoids

With respect to a given bifunctor 'multiplication' in a category, a *monad* is defined as in Eilenberg-Moore to be an object E together with a given internal unit and a given 'internal multiplication' map

$$1 \longrightarrow E, \qquad E \cdot E \longrightarrow E$$

(satisfying the associative and 2-sided unit laws), where 1 is the object acting as unit for the functorial multiplication. In case the multiplication functor is merely the cartesian product in a category X which has such, monads are usually called *monoids*. The category of actions in X of a given monoid in X is typically another cartesian closed category. In case the category in which we consider the monads is the category of exponents of a given category X, it seems reasonable to call such monads 'Galilean monoids'. They are a real generalization of the monoids in X, yet far more special than the usual monads in the category of all endofunctors. In particular, the category X^{E} of actions on 'configuration spaces' for a Galilean monoid E has *both* left and right adjoints (free and cofree) to the same forgetful functor $X \longleftarrow X^{E}$, as was proved by Eilenberg and Moore in 1965. A *strictly -generated* Galilean monoid in X is one generated by a pointed strict exponent $1 \longrightarrow E$. If the generating process in that case involves a filtered colimit in a topos X then the category X^{E} of actions of E is not only a topos (in particular cartesian-closed), but the functor "underlying configuration space" will be a "local geometric morphism" from X^{E} to X, because the left adjoint aspect will be lex. A distinct system of adjoints is "equilibrium configurations", X^{E} to X, with trivial action and cotrivial action as left and right adjoints.

While 1st order ODE's in X and their equivariant maps constitute a topos because they are just the actions of an 'ordinary' monoid infinitesimally generated by $T_1 \approx D_1$ together with the constraint that the point of the instant act as the identity, the 2nd order ODE's in X form a topos because they are the actions of a strictly−generated Galilean 'monoid' infinitesimally generated by the inclusion $T_1 \longrightarrow T_2$ (coordinatizable by $D_1 \quad D_2$) together with the section constraint.

APPENDIX 8

(Additional comments added Nov. 4, 1998)

Recently Kock and Reyes verified my main claim in this outline, namely that second-order ODE's (and many similar prolongation structures) in a given topos constitute another topos receiving an essential morphism from the first, provided certain fractional exponents exist. Rather than using filtered colimits of fractional exponents as in the outline, their proof uses general properties of Top/S such as the existence of fibered products. They also emphasize that the particular "α-prolongation law modeling time" given as a simple example at the end of the outline, namely a one-dimensional interval with zero acceleration, does not represent all lawful motions. Indeed, second order time, the object T representing as an abstract general the concrete generality of all lawful motions in all objects of the topos X^{α}, is a richer "Algebra of Time" than just one-dimensional, as I pointed out in my lectures of that title at the Hamilton Sesquicentennial (Dublin 1993) and at La Sapienza (Rome 1995). Even richer is Galilean time, the object serving the analogous role in the category of dynamical systems, which are structures involving a pair of second-order ODE's on the same configuration space, one of the pair being homogeneous to serve as the 'inertial' zero of specific force. Precise descriptions of those representing objects are needed.

A further observation is suggested by my June 1998 talk 'Why functionals need analyzing' to the Canadian Mathematical Society, where I pointed out that solution-operators for boundary-value problems are also α-prolongation structures. Of course, the domain of α (i.e. the 'boundary' of the codomain) is in such cases not usually an a.t.o.m.; however, any topos is a subtopos of another one in which any given object becomes an a.t.o.m., so that the instrument whose development has been taken up by Kock & Reyes may ultimately also shed light on those problems.

Toposes of Laws of Motion

F. William Lawvere

Transcript from Video, Montreal September 27, 1997

Individuals do not set the course of events; it is the social force. Thirty-five or forty years ago it caused us to congregate in centers like Columbia University or Berkeley, or Chicago, or Montreal, or Sydney, or Zurich because we heard that the pursuit of knowledge was going on there. It was a time when people in many places had come to realize that category theory had a role to play in the pursuit of mathematical knowledge. That is basically why we know each other and why many of us are more or less the same age. But it's also important to point out that we are still here and still finding striking new results in spite of all the pessimistic things we heard, even 35 or 40 years ago, that there was no future in abstract generalities. We continue to be surprised to find striking new and powerful general results as well as to find very interesting particular examples.

We have had to fight against the myth of the mainstream which says, for example, that there are cycles during which at one time everybody is working on general concepts, and at another time anybody of consequence is doing only particular examples, whereas in fact serious mathematicians have always been doing both.

1. Infinitesimally Generated Toposes

In fact, it is the relation between the General and the Particular about which I wish to speak. I read somewhere recently that the basic program of infinitesimal calculus, continuum mechanics, and differential geometry is that all the world can be reconstructed from the infinitely small.
One may think this is not possible, but nonetheless it's certainly a program that has been very fruitful over the last 300 years. I think we are now finally in a position to actually make more explicit what that program amounts to. As you know 30 years ago I made certain proposals in Chicago and then again

15 years ago in Buffalo. There has since been a lot of work on what came to be called synthetic differential geometry. At least 20 people in the world have made important advances in synthetic differential geometry; indeed several of these people are here. And there are also very encouraging developments about the simplification of functional analysis. So I think that on the basis of these developments we can focus on this question of making very explicit how continuum physics etc. can be built up mathematically from very simple ingredients.

To say that a topos can be built up from an object T will mean here that every object is a direct limit of finite inverse limits of exponentials of T. By exponentials of T we mean T^T, T^{T^2} etc. and of course the inverse limits involve equalizers of maps between finite products of these. Such equalizers may be considered as varieties, and in particular the equalizers of maps between finite products of T itself are intended to be infinitesimal varieties. There are actually many interesting useful toposes which are built up in that way from an object T which in some of several senses is infinitely small. Of course T is not just a single point; but it may *have* only a single point, or more generally the set of components functor may agree with the functor represented by 1 on T and its products and sums. One of these senses is that it is a space whose algebra of functions is linearly finite-dimensional; of course that presupposes that we have some linear algebra in the topos, in particular a base rig. But actually it turns out that the base rig itself can be constructed from T.

I'm going to assume T to be a pointed object

$$1 \longrightarrow T$$

This arrow itself is a kind of contradiction expressing that an instant of time involves a point and yet is more than a point. A crucial role is played by the internal endomorphism monoid T^T of T. Also very important is the submonoid of that consisting (in the internal sense) of those endomorphisms which preserve the point.

I am actually going to define R to be that part. If this works we can consider

$$R^X$$

as the space of the simplest kind of intensively variable quantities. We can also consider the R-homogeneous part of the space of functionals

$$\mathrm{Hom}_R(R^X, R)$$

as representing the simplest kind of extensively variable quantities over the domain X; typically this means something like the space of distributions of compact support in X. The basic spaces which are needed for functional analysis and theories of physical fields are thus in some sense available in any topos with a suitable object T. It would be nice if we could prove that R is commutative, but I don't know how to do that from more basic assumptions. You might ask "couldn't T^T itself be commutative?" But there is a very general fact about cartesian closed categories: If an object T has a commutative internal endomorphism monoid, then T itself is a subobject of 1. Intuitively, T^T always includes constants, $T \longrightarrow T^T$ and if constants commute, they are equal. Thus although T itself may be very small, we must have that T^T is a little bit bigger than R. The idea is that a real quantity λ is just a temporal speed-up or retardation

$$1 \longrightarrow T^{C\lambda}$$

As we will see, although R is in a sense the more familiar, the bigger monoid T^T and its actions also play very important roles. Again, there is a general fact about cartesian closed categories. For any monoid M in such a category we can consider also the category of all internal right actions of M. There are, of course, the co-free actions $(\quad)^M$, but we can also consider the action on $(\quad)^T$ where T is the space of constants of M; this functor will be right adjoint to the fixpoint functor if T has a point. In case M = T^T, this right adjoint is actually a full inclusion. We want to think of X^T as the tangent bundle of X with the evaluation at the point of T as the bundle projection. (This idea is already described by Gabriel in SGA3, for example.) The fullness is in contrast with the situation obtained, if we consider that the tangent bundles are equipped only with the R-action, in which case maps between them are essentially contact transformations, not necessarily induced by differentiating (i.e. exponentiating) maps between the configuration spaces. The space can be recovered from its tangent space as the zero section, but even the maps between the spaces can be recovered if we take into account the action of this slightly larger monoid T^T.

As we know, there are many examples of such categories: algebraic geometry, smooth geometry, analytic geometry (real or complex), and many variations on those; actually, in my Chicago lectures I pointed out that there are many potentially interesting intermediate examples of such toposes, for example obtained by adjoining the single operation

$$\exp(-(1/(\ )))$$

to the ordinary theory of real polynomials, so that we can obtain the typical partitions of unity of smooth geometry and yet work in a concrete "algebraic extension" context. (But these are algebraic extensions of systems of quantities of various arieties, so appropriately modeled as a category, rather than just as a differential ring.) All these examples have something in common, and part of the program was to figure out what that "something" is, while at the same time providing a language powerful enough to make all significant distinctions between them. Part of what they have in common is that they are all defined over a simple base topos, the classifying topos for a pointed instant T acted on by a pointed monoid M and satisfying some rather remarkable special axioms. In this base topos M probably is T^T since there is nothing else between, although without additional hypotheses that isomorphism will not persist into arbitrary toposes defined over this base. In the standard examples, exponentiation by T (the tangent bundle concept) is actually preserved up to isomorphism by the inverse of the classifying morphism although this is not a general topos-theoretic fact. Those examples differ mainly in the higher types, that is, in the precise determination of the maps whose domain is the finite-dimensional space T^T. Anyway, since R is a basic definable sub-object of M, we see why the usual examples are infinitesimally generated in the sense of this lecture; smoothness of morphisms between infinite-dimensional spaces has been successfully tested via smooth maps from finite-dimensional varieties for 300 years. In the standard examples the functor to this infinitesimal base topos has additional adjoints so that the latter is actually even an essential sub-topos; this implies that there is a comonad on the big topos which deserves the name of infinitesimal skeleton.

In all the standard examples the object T is isomorphic to the spectrum of the algebra of dual numbers. This implies some rather remarkable axioms that the pair M,T may be required to satisfy. For example, T has a fixed point operator

$$T^T \longrightarrow T$$

assigning to each endomap a fixed point of it; in this case this operator is nothing but the bundle projection (evaluation at zero) and can be interpreted as a map with domain M. This is a consequence of a more general remarkable axiom, namely if β is in M and if λ is in the zero-preserving part R of M, then

$$\beta\lambda\beta = \beta^2\lambda$$

Of course, this striking commutation relation is interpreted as a commutative square whose domain is $M \times R$ in our infinitesimal base topos.

The speedups/retardations λ of the temporal instant $1 \longrightarrow T$ should actually form a rig R. The multiplication is just composition of speedups, but the addition is also intrinsic, somewhat as in Mac Lane's 1950 analysis of linear categories. More precisely, the requirement is that the extensive-quantity functor

$$\mathrm{Hom}_R(R^X, R)$$

is additive in the sense that it takes finite coproducts of spaces X to cartesian products, the needed projections coming ultimately from the point 0 of T. Of course, it suffices to assume this for the case $X = 2$. It is the expectation, that in the smooth world R-homogeneous maps are automatically linear, that underlies this axiom.

Many people have thought about related questions. For example, Peter Freyd had many unpublished ideas, and David Yetter's thesis develops some of our suggestions quite far. The part of R consisting of elements of square zero may be called D as has become customary in synthetic differential geometry. An isomorphism between T and D should amount to the same thing as a choice of a unit of time.

There is another striking property which seems to be frequently correlated with being very small. In order to settle once and for all the various terminological differences, perhaps we can use

a. t. o.m.

as an abbreviation for “amazingly tiny object model”. Whatever we call it, the property is that of the exceptional existence of an additional right adjoint. Since I first wrote about this in 1980, it occurred to me that a suggestive name for this adjoint is fractional exponentiation. Briefly, certain very special objects T may be not only exponentiable, but also fractionally so in the sense that there is an adjointness where the new functor is denoted as the fractional exponent $1/T$.

$$\frac{X \longrightarrow Y^{1/T}}{X^T \longrightarrow Y}$$

Of course, $(\ )^T$ itself is defined by another adjointness (lambda conversion)

$$\frac{W \longrightarrow X^T}{T \times W \longrightarrow X}$$

More generally, if A is any object (not necessarily an a.t.o.m.) which is exponentiable, then we even have fractional exponents A/T. These fraction symbols compose as right adjoint right operators. When the denominators are trivial, this composition is just represented by cartesian product of the numerators. When the denominators are general a.t.o.m.’s the composition is not commutative, but nonetheless can be reduced to the simple fractional form. That is, it follows from the assumed adjointnesses that (right actions)

$$(1/T)B = B^T/T$$

These fractional exponents will play a crucial role in what follows.

2. Galilean 'monoids' for 2nd order ODE's in toposes

Nowadays many mathematicians study abstract objects that are called dynamical systems. Dynamical systems conceptually are intended as what one might call an analysis of Becoming. Already with Aristotle it became customary to analyze Becoming into two aspects, Time and States, with the Time somehow acting on the States. There are many variants on this model, but it is the one we still have. The action of Time on the States is the particular law of motion. More precisely, a given model of Time (a discrete monoid, a continuous group, etc. etc.) serves as an Abstract General which is accompanied by the Concrete General which is the category of all dynamical systems, i.e. systems of states, acted on by that model of Time.

But in a sense much of the current work on dynamical systems is within a framework that still hasn't caught up with Galileo. Galileo made a big advance on the basic idea. In his book "Dialogues concerning two new sciences" written toward the end of his life, Galileo put forward the dynamical refinement of the Time/States analysis which involves the following features.

(1) States are states of Becoming. This again admits many variants: the States may involve velocities, or memories, or destinies, but in any case they themselves should be more structured than just points which abstract static Being particularized as configurations. [In Birkhoff's 1919 Palermo paper which gave rise to the theory of fiber bundles, states are fibered over configurations.]

(2) The particular law of actual motion is accompanied by another law which is not the actual law, but which "would be if there were no forces", as Newton put it. This accompanying law is called inertial or geodesic or spray. The latter merely means that the law is homogenous with respect to the monoid R of time−speedups. Thus the Abstract General itself is more detailed and refined than just a group. (It is of course not excluded that the actual law may be itself homogenous in some particular cases, but the accompanying inertial law always is, it seems.) Although the notion of affine connection is given a much more complicated explanation in most text books, in fact it just expresses this homogeneity idea: if we speed up by a factor λ, then move ahead inertially in time for duration t, we arrive at the

same state as if we had proceeded without speedup for a duration λt and then sped up.

(3) Typically, all the laws constitute an infinite-dimensional affine space which is not a vector space, but the specification of the inertial law provides an origin in this space. Thus we can define the specific force to be the difference between an actual law and the inertial law, and the forces can be added vectorially.

There could be no science or technology without something like feature (3). The actual motion of a piece of chalk thrown into the air is influenced by Jupiter's third moon and any number of other things. But the most important thing is gravity, or the most important thing is wind resistance, or wind resistance and gravity, etc., i.e. we can make an understandable theory of a law of motion which depends only on a few forces, and to the extent that other forces really are negligible that theory will lead to workable technological design. It is the specification of the inertial law as a zero in the affine space of all laws which permits the vectorial addition of individual simplified laws (or rather of their corresponding specific forces). For example, a viable law s of Becoming might be (without mentioning forces explicitly) an alternating sum

$$s = s_1 - s_0 + s_2$$

(so that the coefficients add up to 1 as required for an affine combination), wherein the inertial law s_0 is homogeneous with respect to all of R, whereas the others (without themselves being linear in the usual sense), might enjoy some restricted homogeneities. For example, s_1 might be a purely reactive law homogeneous with respect to those lambdas in R which are involutions, while s_2 might be a purely dissipative law homogeneous for the lambdas in R which are idempotent; the first is often expressed by saying that a purely reactive force (no friction present) enjoys reversibility in time, while the second expresses roughly that pure friction or viscosity is inoperative when the velocity happens to be zero. Each of these three laws s_1, s_0, or s_2 could be considered as a simple model in its own right, but the alternating sum will often be more accurate. Expressed in terms of the specific force laws this combination is

$$\begin{aligned} s_i - s_0 &= f_i \qquad i = 1, 2 \\ f_1 + f_2 &= f \\ s - s_0 &= f \end{aligned}$$

What, more precisely, do we mean by a 'law'? and how could the laws possibly form a topos as promised in my title? First, note that the usual 'dynamical systems' involving for example the smooth actions of a monoid, if properly construed, will surely form a topos with all the virtues that that entails such as internal logic, good exactness, function space of 'dynamical systems', etc. Likewise, the infinitesimal version of such systems i.e. vector fields or first-order ODE's, will also form a topos as I pointed out in my Chicago lecture. But what about actual dynamical systems in the spirit of Galileo, for example, second-order ODE's? [Of course, the symplectic or Hamiltonian systems that are also much studied do address this question of states of Becoming versus locations of Being, but in a special way which it may not be possible to construe as a topos; in any case, most systems arising in engineering are not conservative.]

To specify an Abstract General whose corresponding Concrete General will consist of state spaces, each equipped with its own genuinely dynamical law, I propose the following. Consider any given map

$$T \longrightarrow A$$

in a topos of spaces, subject only to the restriction that its domain should be an a.t.o.m. (the codomain A need not be an a.t.o.m., although it often will be). If I have any space X I can consider the restriction

$$X^T \longleftarrow X^A$$

along the map induced by my given map on the map spaces. The kind of structure that I want to consider is that of a given section s to this restriction map

$$X^T \xrightarrow[s]{} X^A$$

This section will serve as a law of Becoming in the sense that given a map from T to X, considered as a state of instantaneous Becoming, the law will provide a definite extension to A considered as a distinctly longer instant.

$$
\begin{tikzcd}
T \arrow[r] \arrow[dr, "x"'] & A \arrow[d, "s(x)"] \\
 & X
\end{tikzcd}
$$

The standard example has T as a first-order infinitesimal instant and A as a second-order instant. In that case the choice of a unit of time would identify A with the part D_2 of R consisting of elements of cube zero in T^T. An actual motion following a law s would be a map (in the dynamical topos) whose domain is a relatively small object idealizing the state space of a clock, i.e. an interval of time equipped with its own (often homogeneous) law which the map must preserve.

Theorem: For any given map $T \longrightarrow A$ in a topos with natural number object, where T is an a.t.o.m., the category of all pairs X, s as above, with the obvious motion of morphism, is a topos lex-comonadic over the given topos. In fact, the resulting 'surjective' geometric morphism is essential.

Proof: By the basic adjointness such a section is equivalent to a map

$$X \longrightarrow X^{A/T}$$

which, when 'evaluated at' the given $T \longrightarrow A$ reduces to the identity on X. The co-pointed endofunctor (pullback of the fractional exponentiation) has a left adjoint. Therefore, iterating it and passing to the sequential limit yields a lex-comonad which even has a left adjoint monad. Since the comonad is left exact, the coalgebra/algebras for this pair constitute a new topos, but they are equivalent to the laws of motion s under discussion. The essentiality is thus a special case of the Eilenberg-Moore theorem about adjoint monads.

Thus we see that there is a Galilean generalization of the notion of monoid. Recall that any monoid in a 'cartesian' closed category is equivalent to a pair consisting of an adjoint monad with its right adjoint comonad. But the converse of that statement is true only if the adjointnesses are internal. In our case the right adjointness of the comonad is defined only over some lower topos (as was discussed in my 1981 Cambridge lecture and investigated in Yetter's thesis). This is, however, still a very special kind of lex comonad since it is generated by this fractional exponent.

To sum up, the actions of such a Galilean 'monoid' thus constitute a topos of laws of motion in the Galilean dynamical sense. For example, if A consists of second-order infinitesimals, all the usual smooth dynamical systems, including the infinite-dimensional ones, (elasticity, fluid mechanics, and Maxwellian electro-dynamics) are included as special objects.

3. Infinitesimals bodies too?

Galileo's second new science, as interpreted by Noll, concerns the particularity of the ways in which constitutive relations of actual materials give rise to laws of motion s on the configuration space X *of a body*, when the body is subjected to arbitrary external conditions. While there was no time at the AMS lecture to elaborate on that, I did discuss it in my 1992 lecture in the engineering faculty at Pisa and in my talk at the 1993 Nollfest. In addition to providing a flexible general conceptual setting for considerations of materials science, the methods involving infinitesimal objects in toposes seem to also offer a definite particular model for the kind of surfaces studied by the Cosserat brothers and for the pseudorigid or zero-dimensional bodies studied by Muncaster and Cohen.

For Perugia Studies in Mathematics (Proceedings of the May 1997 Meeting in Perugia)
(October 1998 correction)

Volterra's functionals and covariant cohesion of space

F. William Lawvere

Abstract:

Volterra's principle of passage from finiteness to infinity is far less limited than a linearized construal of it might suggest; I outline in Section III a nonlinear version of the principle with the help of category theory. As necessary background I review in Section II some of the mathematical developments of the period 1887-1913 in order to clarify some more recent advances and controversies which I discuss in Section I.

I

The immediate impetus to this historical exploration came from two articles by Gaetano Fichera. The resulting line of study needs to be deepened and considered in more detail, but it already supports certain conclusions concerning the precise methodological direction of global analysis. These methodological conclusions can perhaps be tested on the ample material provided by the important new book by Kriegl and Michor, just published by the AMS [17]. On that basis we can hopefully begin a serious reply to the challenge of 'I difficili rapporti fra l'analisi funzionale e la fisica matematica'. Some of those difficulties are outlined in the article [8] (with that title) by Fichera.

The other article by Fichera, 'Vito Volterra and the birth of functional analysis' [9], is basically a re-affirmation of the role of the Volterra school in the development of modern analysis, in response to Dieudonné's [6] :

> We must finally mention the first attempt at 'Functional Analysis' of the young Volterra in 1887, to which, under the influence of Hadamard, has been attributed an exaggerated historical importance.

More specifically, in connection with Volterra's notion of the derivative of a functional, Dieudonné further states:

> With our experience of 50 years of functional analysis we cannot help feeling that without even the barest notion of general topology, these ad hoc definitions were decidedly premature. [ibid]

(One might ask whether the pioneers of the calculus of variations were also ‘premature’.)

To lay the ground for his response to Dieudonné, Fichera proposes that

> the true historian must make the effort to shed himself of today’s way of thinking and of all the experience he has acquired ... discarding all the superstructures which have arisen during the passage of the years... [9]

My own defense of Volterra will be perhaps a bit less modest than Fichera’s. Like Fichera, I believe that the possibility for an eminent French mathematician to make such statements derives partly from a conceptual identification by mainstream mathematics:

Functional analysis = the study of topological vector spaces.

But I want to depart from the above definition of ‘true historian’ and rather re-visit past developments in light of today’s problems in order to explain my conclusion that this narrow conceptual identification is one of the reasons for the ‘difficili rapporti’, and is one of the ideas which must be refined to improve those relations.

The authors of a recent very useful biography of Hadamard [20], while acknowledging Fichera’s arguments, nonetheless describe as ‘naive’ the theory of analytic functionals studied by Fantappiè, Pellegrino, Haefeli, Succi, Teichmüller, Silva, Volterra, Zorn and others [23]. More specifically, they say that

> The naivety of Fantappiè’s approach lies in the fact that he, as Hadamard before him, avoided the topology and assumed analyticity in the sense of his general theory, instead of using continuity.

That these mathematicians were not in fact naive about the role of topology is adequately demonstrated by the fact that a portion of their work was devoted to showing that analytic functionals are automatically continuous with respect to some notion of neighborhood. But more importantly, as I will try to explain in section III, the ‘general theory ‘ is more powerful than commonly supposed.

The idea of the preponderance of a neighborhood structure in infinite dimensional spaces grew with the enormous work on partial differential equations which has been accomplished in this century, in particular the work revolving around the notion of 'a well-posed problem in the sense of Hadamard' [20]. That important guiding notion is as follows: In the solution of certain problems of elasticity, electromagnetism, etc., the solution itself may change only slightly if the boundary data and region of definition are themselves varied sufficiently slightly. Making this guiding idea precise apparently requires specifying a notion of neighborhood in the domain space of the solution operator. But the difficulty is that these notions of neighborhood are not unique. Indeed general topology is so general that one could trivially achieve 'well-posedness' by simply defining the neighborhoods in the domain space to make the principle true (assuming the specification in the other space is less problematic; but note that a solution operator is typically a section of another operator, in the opposite direction, which one may hope is continuous also.) Of course, the well-posedness principle is not so tautological; one intends that the choice of topology (or neighborhood notion) is natural with respect to a category of problems at hand.

Grothendieck remarked to me in 1981 that already around 1950 he had had great difficulty in persuading analysts that there can be more than one natural topology for the same family of linear spaces. This remark in effect presupposes that there is some more basic cohesive structure which can be the 'same' for various notions of neighborhood. Perhaps this situation could be schematically represented as

$$\begin{tikzcd} & \mathrm{LBT} \arrow[dd] \\ C \arrow[ur, dashed] \arrow[dr, "F"'] & \\ & \mathrm{LB} \end{tikzcd}$$

where the family F of basically-cohesive linear spaces is functorially parameterized by a category C of problems, and we consider liftings across the forgetful functor from basically-cohesive spaces which are moreover compatibly equipped with a given topology.

But what could such a more basic notion of cohesiveness amount to? A very strong candidate is bornology. In their 'Convenient Setting for Global Analysis' [17], Kriegl and Michor state clearly that

> the locally convex topology on a convenient vector space can vary in some range - only the system of bounded sets must remain the same.

Their choice of morphisms for the purposes of infinite dimensional differential calculus was based on experience referred to, for example, in the fundamental work [25] on the spaces of differential forms :

> Nous allons définir dans chacun de ces espaces, une notion *d'ensemble borné*, qui sera utile pour analyser la condition de continuité qui intervient dans la définition des courants. (de Rham 1953)

and in the important step [26] in the development of general differentiation:

> Je me suis persuadé que, pour cette généralisation, c'est la notion d'ensemble borné, plutôt que celle de voisinage (ou de sémi-norme), qui doit jouer un rôle essentiel. (Silva 1960)

Waelbroeck, who also participated with Silva in the 1960 Louvain meeting, published two articles in 1967 (in the first volume of the Journal of Functional Analysis) clearly establishing the correctness of these views of de Rham and Silva concerning the intimate connection of bornology with smoothness.

But there is another notion of basic structure more directly relevant to the approximation procedures involved in 'solving' partial differential equations by computer: Several different topologies can give rise to the same notion of *convergent sequence*. This is directly involved in another traditional construction in functional analysis which does not fit into the monolithic 'topological vector space' mode; for example Carathéodory, in his 1935 treatise on 'The Calculus of Variations' finds appropriate to give, as the first topic on page one of chapter one, the definition of 'continuous convergence' for a sequence f_k of functions defined on a domain A in n-dimensional space:

> For every possible sequence P_1, P_2... of points which lie in A and which converge to P_0, the limit $\lim_k f_k(P_k)$ always exists and represents a definite number. (my translation) [4]

This sort of definition of a cohesive structure on a function space is in fact forced by the universal property of cartesian closure, when the 'universe' is a category of spaces whose cohesion is determined by convergence of sequences. This fact is explicit in Johnstone [16] but also in Fox 1945 [10], though the latter did not explicitly use the term 'cartesian closed category' (nor the term 'λ-calculus' which his Princeton colleague Church introduced at about the same time to signify an important symbolic aspect of cartesian closure.)

Like the idea of convergent sequences, bornology is applicable in both linear and nonlinear settings (indeed bornology in its own sequential version has a very simple topos incarnation which could be considered parallel to Johnstone's sequential-convergence topos). However, it is only in tandem with linearity that bornology suffices for some part of global analysis (via the notion of Mackey convergence (1945)). Yet the needed applications of global analysis to calculus of variations or continuum physics are usually nonlinear. Another unspoken presupposition of mainstream mathematics seems to be

nonlinear is a generalization of linear and hence more difficult.

But there are important ways in which a nonlinear category can be simpler than the linear category of vector space objects in it. These ways imply, as I will attempt to explain, that Volterra's Principle of the passage from the finite to the infinite can be freed from the limitations described by Fichera, i.e., that the Principle does in fact have the potential to deal with such phenomena as non-closed linear subspaces or total continuity. [9]

Nonlinear categories of C^∞ and of analytic spaces and maps are central to global analysis. The key concept which yields analytic maps in the sense of the Volterra school, namely that a good map is one that takes good paths into good paths, was applied successfully to define also a category of smooth maps by Hadamard [13], Boman [1], Lawvere-Schanuel-Zame [18], and Frölicher [11], and is central to [17] . In the latter, Kriegl and Michor state at the outset that they will define the smooth maps to be those which take smooth paths to smooth paths and that everything follows from that definition.

II

What are the main turn-of-the-century events which gave rise both to the current rich mathematical development, as well as to these historical controversies? In this section I attempt to give a very succinct summary.

After 200 years of both pure development and extensive application of calculus, Betti in 1887 presented to the Accademia dei Lincei a series of three notes by Volterra under the title 'Sopra le funzioni che dipendono da altre funzioni' (Opere pp 294-314). In that same year Volterra also treated 'le funzioni dipendenti da linee' and then in 1889 wrote [27]

mais les points ne sont pas les seuls éléments géometriques.....

He emphasizes that also curves and surfaces (and we might add, tangent vectors) in X are elements of a space X, and just as one considers functions of points one must also consider functions of these species of elements as 'generalized' functions on X. Nor was Volterra driven solely by esthetic considerations (as Hadamard in his 1943 *Psychology of Mathematical Invention* inexplicably asserts), for he explicitly had in mind dependencies in continuum physics, such as the dependence of interior temperature on the boundary distribution of temperature of a body and similarly of the interior displacement of a flexible surface on the boundary distribution of displacements. In these papers Volterra made explicit some basic problems, concepts, and kinds of results which still occupy researchers in global geometric analysis and its application to continuum physics; among the key properties of functionals which he established [22] was the local existence theorem which is now referred to as the Poincaré lemma on the exactness of the de Rham complex of sheaves.

In 1897 Hadamard sent from Bordeaux a note [14] which was read by Picard at the Zurich International Congress. This note proposed the explicit consideration of sets of functions as mathematical spaces in their own right. (Bourbaki [2] considers this as the moment in history at which the usefulness of set theory began to be accepted by the mathematical community; because of the nature of the problems to which Hadamard referred, we can see from today's perspective that sheaf theory, at least as much as the theory of abstract sets, was being called for.) Pincherle [24] rose at the Congress to point out that the sort of theory sought by Hadamard, involving cohesive variation within function space, was already well under development in Italy, not only in the cited works of Volterra, but also in works of Ascoli, Arzela, and others. By 1903 Hadamard had already mastered much of that theory and developed it further, in particular coining the term 'functional'; in 1910 he published, with the help of Frechet, his beautiful lectures [15] on the calculus of variations, including a detailed exposition of the warmly praised theory of Volterra. Among the many later developments, Volterra became widely known for his pioneering work on Boltzmann's hereditary elasticity, as well as for his treatment of the fluctuation of the fish population in the Adriatic, while Hadamard became famous, not only for his theory of 'well-posed' problems, but also for his analysis of Huygen's Principle concerning wave equations in spaces of an odd number of dimensions. In the subsequent period, the important concept of neighborhood in function space somehow came to be considered as all-important; I cannot fully account historically for that rigidification of ideas, though I hope someone will be able to do so.

III

In this last section I will try to summarize synthetically the mathematical situation as it appears in light of these and previous studies.

The core disciplines such as geometric measure theory and differential geometry (i. e. the domains of the main problems which are treated via partial differential equations) gave rise, in the work of Volterra and others of his epoch, to the great auxiliary subjects of algebraic topology and functional analysis. The typical infinite-dimensional spaces produced for study by those disciplines are nuclear or at least have all bounded sets precompact and hence are essentially disjoint from the Banach and Hilbert categories which are necessarily introduced in the course of numerical and other analyses of these spaces. Similarly, the smooth manifolds (which serve as domains of variation for these spaces) are essentially disjoint from the simplicial and combinatorial categories which are necessarily introduced in calculating their algebraic-topological qualities. An important foundational goal is the specification of a flexible category (or system of categories) which can serve as a setting for both these objective and necessary subjective aspects, and to account for the relation between them.

A category of locally convex topological vector spaces is often considered as the linear side of such a setting, with manifolds modeled on open subsets of these as the nonlinear side. However, this vast generality makes room for a whole menagerie of counterexample spaces which are much more complicated than either the nuclear space or Hilbert space aspects; moreover, the reliance on the contravariant structure consisting of neighborhood systems, in accounting for the needed cohesion, makes almost impossible the achievement of those intuitively simple function-space transformations which are expressed by the exponential law of cartesian-closed categories.

I have been using the vague term ‘cohesion’ (in the way that physicists used to employ the term ‘continuity’) to avoid prejudice in favor of one or the other determination of it. Certainly we are considering that ‘C^{∞}-structure’ is a central kind of determination, and that ‘topology’ in the now-standardized sense is an important derived determination. In general, we might have occasion to treat, as a species of cohesion, almost any *extensive* category [5] for the non-linear side, and for any given rig object R in it, the category of all internal R-modules as a corresponding linear side; a linear subcategory of 'complete' R-modules can often be defined in terms of the necessary higher-order structure discussed below. But how are significant non-combinatorial examples, of extensive categories with cartesian products, to be constructed?

A paradigmatic example has been the consideration of sets equipped with neighborhood structures, with cohesion-preserving morphisms f considered to be those which *contravariantly* respect the neighborhood structure: "No matter how small a neighborhood V of f(x) is demanded, a neighborhood U of x can be found which is so small that $U \quad f^{-1}V$ ", i.e., that f^{-1} preserves openess. The trouble with this determination, as a basic setting, is that, as was pointed out by many authors, it does not support the supple higher-order structure which it was designed to model. Many proposed remedies (such as those of Kelley, Spanier, Brown, and Steenrod), although perhaps still couched in the language of neighborhoods, in effect replaced the structure-and-morphism specification by a *covariant* one such as "for any compact figure C in the domain of f , fC is a compact figure in the codomain".

Both bornology and convergence, by contrast, are inherently covariant and hence lead immediately to cartesian-closed categories in a manner quite painless compared with the restricting remedies which had to be imposed on contravariant structures to achieve a similar end.

In general, if in a category a subcategory of special 'forms of elements' is distinguished, then a general 'space' X in the category determines a geometric structure consisting of elements $A \longrightarrow X$ and incidence relations given by $A' \longrightarrow A \longrightarrow X$ for $A' \longrightarrow A$ in the subcategory; sometimes this covariantly-associated structure determines the spaces, in the sense that any abstract mapping $X \cdot\cdot\cdot\cdot\cdot\to Y$ of all these elements comes from an actual space morphism $X \longrightarrow Y$ if only it is compatible with the incidence relations. Dually, one could consider a subcategory of objects R as basic quantity types, and associate (in a contravariant manner) to each general space X an algebraic structure consisting of functions $X \longrightarrow R$ with algebraic operations given by $X \longrightarrow R \longrightarrow R'$. Even when the algebraic structure does not determine the general space, it may provide useful notions of approximation; it is often possible to arrange that adequate ‘forms of elements’ A can be chosen so that the R-algebra structure suffices at least *for them*. Given any rig object R, with its multiplicatively-invertible part U, we define a subspace of a general space X to be ‘R-open’ if it is of the form $\varphi^{-1}U$ for some function $\varphi\colon X \longrightarrow R$ in the category; then every morphism f in the category is automatically R-continuous! Note that all sober objects in the usual category of topological spaces are entirely determined by this ‘algebraic’ structure, with the choice R = the two-point Sierpinski space.

One could dispute the necessity for a cartesian-closed base category of spaces, if one were content to deal only with spaces of extensive quantities [19]. For example, on the category top of all topological spaces, there is the functor C_+ which assigns to every X the *abstract additive monoid* $C_+(X)$ of all nonnegative continuous real functions on X; this functor has an adjoint which composes with it to yield a linear monad M_+ on top, with $\delta\colon X \longrightarrow M_+(X)$ a continuous map and $M_+(X+Y) = M_+(X) \times M_+(Y)$. A key concept of functional analysis, namely that of a continuous path of extensive quantities, is thus approached via maps $I \longrightarrow M_+(X)$ from the interval. I do not know exactly what are the Eilenberg-Moore spaces for this monad; they are in some sense ‘complete’ linear spaces. Of course, serious problems of analysis begin when one considers not-necessarily-positive quantities; but the abelian-group reflections of the M_+-spaces should provide an approach to that. This example was mentioned as an analogy in my 1966 Oberwolfach lecture in which I proposed the study of *S*-valued measures in the category Top of *S*-toposes; that category shares with top the feature of not being closed in general, but rather of admitting only ‘locally compact’ objects as exponents. But among the many striking results of the recent work by Bunge and Funk [3] is that, as in top, the Top-measures do enrich to become the points of new toposes in Top, even though the corresponding intensive quantities in general do not.

Both intensive quantities as well as extensive quantities (varying over spaces) need to be representable by spaces in order to fully exploit the possibilities of functional analysis in approaching the content of continuum physics. This leads inevitably to the basic requirement that the category of spaces be cartesian closed. Recall that this means that for any space A, there is for any space B an exponential space B^A characterized by the adjointness ‘conversion’ property that for any space T, the maps from T to B^A correspond naturally to the maps from A × T to B. If T ranges over an adequate subcategory of ‘element forms’, then the conversion property uniquely determines the various elements and incidence relations in B^A and hence determines B^A itself. If R is another space, a *functional* is then a map

B^A ¾→ R

Actually, we can distinguish two important kinds of functionals-functor. For fixed A and R, the maps X^A ¾→ R determine a contravariant algebra of ‘extended functions’ on spaces X, which become Volterra’s ‘funzioni di linee’ when A is one-dimensional. On the other hand, for fixed B (for example R) and variable exponent, the homogenous maps R^X ¾→ R constitute a covariant functor of spaces X which represents the ‘double-dual’ approach to extensive quantities (as integration processes) associated with the name of F. Riesz. Of course, in any cartesian-closed category exponentiation can be iterated, so that the *spaces* of intensive or extensive quantities have a uniquely determined notion of variable element internal to themselves. As I pointed out in 1967, the Fermat-Study-Kähler approach to tangents and cotangents works also well in a C^∞ category, which is a natural enlargement [7] of the basic path-determined one: For a suitable infinitesimal element-form A, X^A is precisely the tangent bundle of X and the homogeneous ‘funzioni di linee infinitesimali' $X^A \longrightarrow R$ are the differential forms. Combining these ideas, one represents de Rham's currents as maps of the kind $R^{\underline{X}} \longrightarrow R$ where $\underline{X} = X^A$ for infinitesimal A.

The fact that basic smooth categories are 'path-determined' seems a strong form of a principle of passage from the finite to the infinite of the kind possibly envisaged by Volterra. It means in particular that the whole smooth category is 'generated' in the sense that every object, including the infinite-dimensional spaces, is a direct limit of finite-dimensional spaces (among which are the element-forms such as a line); such a thing is surely false for any *linear* functional-analytic category, which is perhaps why Volterra's principle seems limited when confined to such a category. Using the (nonlinear) functions and curves, it is reasonable to define notions of 'R-closed part' Y of an infinite dimensional space X by requiring

(1) that Y be the equalizer of some pair of maps from X to R, or

(2) that the inclusion consider R as an 'injective object'; or

(3) that for all paths $R \longrightarrow X$ the inverse-image of Y be closed in R in the sense of (1) or (2).

(The second of these possibilities is in the spirit of the notion of R-opens mentioned above, but turns Tietze's theorem into a principle, as is implicit in the usual approach to algebraic geometry.)

It would seem that sufficient knowledge of functional analysis *and* category theory has been achieved in 100 (respectively 50) years to permit the formulation of a concrete basis (for global geometric analysis) which is at least approachable without first achieving the erudition of expertise in topological vector spaces. A rough outline is as follows. On the monoid of C^∞ endomaps on a one dimensional space R, there is the topos E_1 of sheaves; this is a locally-cartesian-closed (and locally extensive) category with excellent exactness which contains the Frölicher category F of [12] as a reflective subcategory; the reflector preserves finite products but not equalizers, so that F is not so exact and not locally-cartesian-closed although, crucially, it is globally cartesian closed. The virtue of F is (by definition) that every space in it is (like the line itself) equally well described in terms of line elements and their ‘incidence’ (=reparameterization of paths) or by the algebra of functions with finite-dimensional values. (Perhaps a reasonable parameterized version of this Frölicher duality axiom can be found which will yield a subcategory F(X) of the topos E_1/X with a reflector preserving finite products, thus after all achieving a reasonable modified version of local cartesian closedness.) Since by Boman the algebraic theory of C^∞ algebras is a full subcategory of E_1 and since the smooth toposes E_∞ considered in Synthetic Differential Geometry [21] may be construed to consist of (dual) sheaves on finitely-generated C^∞-algebras, there is a natural inclusion $E_1 \rightarrowtail E_\infty$ of toposes which in a suitable sense generates E_∞ . The virtue of E_∞ is that it contains infinitesimal element forms A which even admit ‘fractional exponentiation’ $(\ )^{1/A}$ so that jets and currents become firmly representable. Strikingly, the exponential of infinitesimals A^A has the one-dimensional space R of homotheties (i.e. time-speed-ups or motion-retardations) as a retract, so that in a precise sense the cartesian-closed category E_∞ of global analysis is literally infinitesimally-generated! The internal R-homogeneous maps from R^X to R constitute a linear monad M whose Eilenberg-Moore objects are very ‘complete’ linear spaces; but all spaces of the intensive kind R^X or the extensive kind M(X) are automatically that complete. The natural appearance of M(X) makes it (and its ramifications) a central concept of extensive quantity perhaps replacing the idea of ‘distributions of compact support’ with which it agrees at least on finite-dimensional manifolds.

Within the linear categories so arising it is surely possible to discern ideals of the nature of compact operators or nuclear operators and indeed to analyze the concrete non-linear origin of such ideals in the ideals of strong inclusions (among the subspace inclusions within a domain space X). That structures existing in finite dimensions tend to persist in infinite dimensions need not be construed (as Fichera apparently did) to mean that statements (such as ‘all operators are nuclear’) should be expected to persist. That is, structure which is trivial may become non-trivial (for example the ring R^X has nonzero 2nd order linear differential operators if $X = R^n$, but not if $X = n$); but all structure is still describable in terms of its effects on finite parameterizations.

We can even attempt to describe a principle of passage from finite to infinite compatible with the additional century of experience. While we clearly need several categories and transformations between them (for example a simplicial topos and a real analytic topos mediated by a sequential-convergence topos) yet within each of the basic categories it is reasonable to expect that the following sort of construction is uniquely determined. Given two spaces A and B we can form B^A, and then the part P of B^A defined by an equation between two maps from B^A to C; then it makes unique sense to speak of a variation of a function on P, with the variation within the category. Stronger still, every space P is in fact determined by its elements (with incidence) of a few representable finite-dimensional forms (such as points, curves, surfaces, tangent vectors, 2 jets). This principle was reasonable 300 years ago, 200 years ago, and 100 years ago, and now, in much more explicit form, seems again reasonable and realizable, despite a century of counterexamples concerning contravariant cohesion, during much of which such a principle seemed ‘naive’.

B i b l i o g r a p h y

[1] Boman, J. Differentiability of a function and of its compositions with functions of one variable, *Math. Scand.* **20**, 429-268, 1967.

[2] Bourbaki, N. *Eléments d’histoire des mathématiques,* Hermann, Paris, 44, 1960.

[3] Bunge, M., Funk, J. Spreads and the symmetric topos, *Journal of Pure and Applied Algebra* **113,** 1-38, 1996.

[4] Carathéodory, C. *Variationsrechnung und Differentialgleichungen erster Ordnung*, B.G. Teubner, Leipzig und Berlin, 1935.

[5] Carboni, A., Lack,S., Walters, R. Introduction to extensive and distributive categories, *Journal of Pure and Applied Algebra* **84**, 145-158, 1993.

[6] Dieudonné, J. *History of functional analysis,* Mathematics studies **49**, North-Holland, Amsterdam, 1981.

[7] Dubuc, E. J. C^∞-schemes, *American J. Math.* **102**, 683-690, 1981.

[8] Fichera, G. I difficili rapporti fra l'analisi funzionale e la fisica matematica, *Rendiconti dell'accademia nazionale dei Lincei,* **9** (1), 161-170, 1990.

[9] Fichera, G. Vito Volterra and the birth of functional analysis, *Development of Mathematics 1900 - 1950*, Pier, J.-P. ed., Birkhäuser, Basel, 1994.

[10] Fox, R. On topologies for function spaces, *Bull. Amer. Math. Soc.* **51,** 429-432, 1945.

[11] Frölicher, A. Applications lisses entre espaces et variétés de Fréchet, *C.R. Acad. Sci.* Paris **293**, 125-127, 1981.

[12] Frölicher, A., Kriegl, A., *Linear Spaces and Differentiation Theory,* J. Wiley, Kriegl A. Chichester, 1988.

[13] Hadamard, J. La notion de différentielle dans l'enseignement, *Script Univ. Jerusalem* **1**, 1923, reprinted in *Enseign. Sci* **35,** 136-137 (1931).

[14] Hadamard, J. Sur certaines applications possibles de la théorie des ensembles, *Verhandlungen des ersten internationalen Mathematiker Kongresses,* B.G. Teubner, Leipzig, 201-202, 1898.

[15] Hadamard, J. *Leçons sur le calcul des variations*, Hermann, Paris 1910.

[16] Johnstone, P. On a topological topos, *Proc. London Math. Soc.* **38** 237-271, 1979.

[17] Kriegl, A., Michor, P. W.*The Convenient Setting of Global Analysis,* American Mathematical Society **73**, 1997.

[18] Lawvere, F. W., Schanuel, S., Zame, W. On C^∞-infinity function spaces. (unpublished) 1981. summarized in [12].

[19] Lawvere, F. W. Categories of Space and of Quantity, *The Space of Mathematics. Philosophical, Epistemological and Historical Explorations.* DeGruyter, Berlin (1992) 14-30.

[20] Maz'ya, V., Shaposhnikova, T. *Jacques Hadamard, a Universal Mathematician*. History of Mathematics, American Mathematical Society/London Mathematical Society **14,** 1998.

[21] Moerdijk, I., Reyes G. *Models for Smooth Infinitesimal Analysis*, Springer-Verlag, New York, 1991.

[22] Narashiman R. *Analysis on real and complex manifolds*, 3rd ed. North-Holland, Amsterdam, viii-ix, 1985.

[23] Pellegrino, F. La théorie des fonctionnelles analytiques et ses applications, part 4 of Levy P., *Analyse Fonctionnelle*, 2nd ed., Gauthier-Villars, Paris 1951.

[24] Pincherle, S. Remarque relative à la communication de M. Hadamard, *Verhandlungen des ersten internationalen Mathematiker Kongresses*, B.G. Teubner, Leipzig, 203, 1898.

[25] de Rham, G. *Variétés différentiables*, Hermann, Paris, 1953.

[26] Silva, J. S. Les espaces à bornés et la notion de fonction différentiable, *Colloque sur L'Analyse Fonctionnelle*, Louvain 1960, Centre Belge de Recherches Mathématiques, 57-62, 1961.

[27] Volterra, V. Sur une généralisation de la théorie des fonctions d'une variable imaginaire, *Acta Mathematica* **12,** 233-286, 1889

Mathematics Department
SUNY at Buffalo
106 Diefendorf Hall
14214 Buffalo, N.Y. USA
e-mail: wlawvere@acsu.buffalo.edu

June 1, 1998